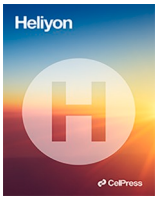
Heliyon 10 (2024) e36913

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

The influence of form on motion signal processing in the ventral intraparietal area of macaque monkeys

Lingqi Kong [a,1], Fu Zeng [a,**,1], Yingying Zhang [a], Li Li [b,c], Aihua Chen [a,c,*]

[a] Key Laboratory of Brain Functional Genomics (Ministry of Education), East China Normal University, Shanghai, 200062, China
[b] Faculty of Arts and Science, New York University Shanghai, Shanghai, 200122, China
[c] New York University-East China Normal University Joint Research Institute of Brain and Cognitive Science, New York University Shanghai, Shanghai, 200062, China

ARTICLE INFO

Keywords:
Optic flow
Glass pattern
Heading perception
Integration
Ventral intraparietal (VIP)

ABSTRACT

The visual system relies on both motion and form signals to perceive the direction of self-motion, yet the coordination mechanisms between these two elements in this process remain elusive. In the current study, we employed heading perception as a model to delve into the interaction characteristics between form and motion signals. We recorded the responses of neurons in the ventral intraparietal area (VIP), an area with strong heading selectivity, to motion-only, form-only, and combined stimuli of simulated self-motion. Intriguingly, VIP neurons responded to form-only cues defined by Glass patterns, although they exhibited no tuning selectivity. In combined condition, introducing a small offset between form and motion cues significantly enhanced neuronal sensitivity to motion cues. However, with a larger offset, the enhancement effect on sensitivity became comparatively smaller. Moreover, we observed that the influence of form cues on neuronal response to motion cues is more pronounced in the later stage (1–2 s) of stimulation, with a relatively smaller effect in the early stage (0–1 s). This suggests a dynamic interaction between motion and form cues over time for heading perception. In summary, our study uncovered that in area VIP, form information plays a role in constructing accurate self-motion perception. This adds valuable insights into the complex dynamics of how the brain integrates motion and form cues for the perception of one's own movements.

1. Introduction

The interplay between motion and form information is a fascinating aspect of visual perception, wherein the two often intricately influence each other. Take the example of a stationary cheetah in the grass - its spots may initially make it challenging to perceive its presence. However, once it starts moving, we can easily discern the overall shape of the cheetah against the background. This showcases the importance of motion cues to catch our attention and separate figure from background. Conversely, we can also perceive implied motion through stationary cartoons or paintings that incorporate velocity lines, also known as motion streaks [1,2]. This

* Corresponding author. Key Laboratory of Brain Functional Genomics (Ministry of Education), East China Normal University, Shanghai, 200062, China.
** Corresponding author.
E-mail addresses: fzeng@bio.ecnu.edu.cn (F. Zeng), ahchen@brain.ecnu.edu.cn (A. Chen).
[1] These authors contributed equally.

https://doi.org/10.1016/j.heliyon.2024.e36913
Received 7 May 2024; Received in revised form 26 July 2024; Accepted 23 August 2024
Available online 28 August 2024

implies that motion information can be extracted from form cues, demonstrating the dynamic relationship between the two. Despite the fascinating interaction between motion and form perception in the visual system, the underlying neural mechanisms responsible for these processes remain unknown [3]. Exploring these mechanisms could provide valuable insights into how our brains seamlessly integrate different visual cues for the perception of the environment.

Heading perception, which involves the visual system's ability to determine the direction of self-motion in the world, serves as a valuable experimental paradigm for exploring the interaction between motion and form cues. This intricate process entails integrating motion cues with the perception of the visual scene's structure [4–8]. Humans, in particular, heavily rely on optic flow, the visual motion of the environmental image generated during self-motion, to perceive self-motion [9]. In natural settings, both motion and form information coexist within the optical flow, providing a rich array of cues for heading perception. Previous research has highlighted that humans can enhance their ability to estimate heading perception by leveraging a combination of motion and form signals [10], indicating a synergistic relationship between the two types of information.

Several non-human primate brain regions have been reported to be in charge of heading perception from optic flow, e.g., the dorsal medial superior temporal area (MSTd, [11–13]), the ventral intraparietal area (VIP, [14–16]), superior temporal polysensory area (STP, [17,18]), area V6 [19–21] and 7a [22], and. Among these regions, neurons in area VIP show robust responses and distinct heading direction selectivity to optic flow stimuli, indicating a keen sensitivity to motion cues [15,16,23]. Interestingly, VIP also receive some inputs from area V4, a region known for its selectivity to shape patterns in the ventral pathway [24], suggesting that VIP may respond not only to motion cues but also to form cues. These points towards a potential role of area VIP in integrating both motion

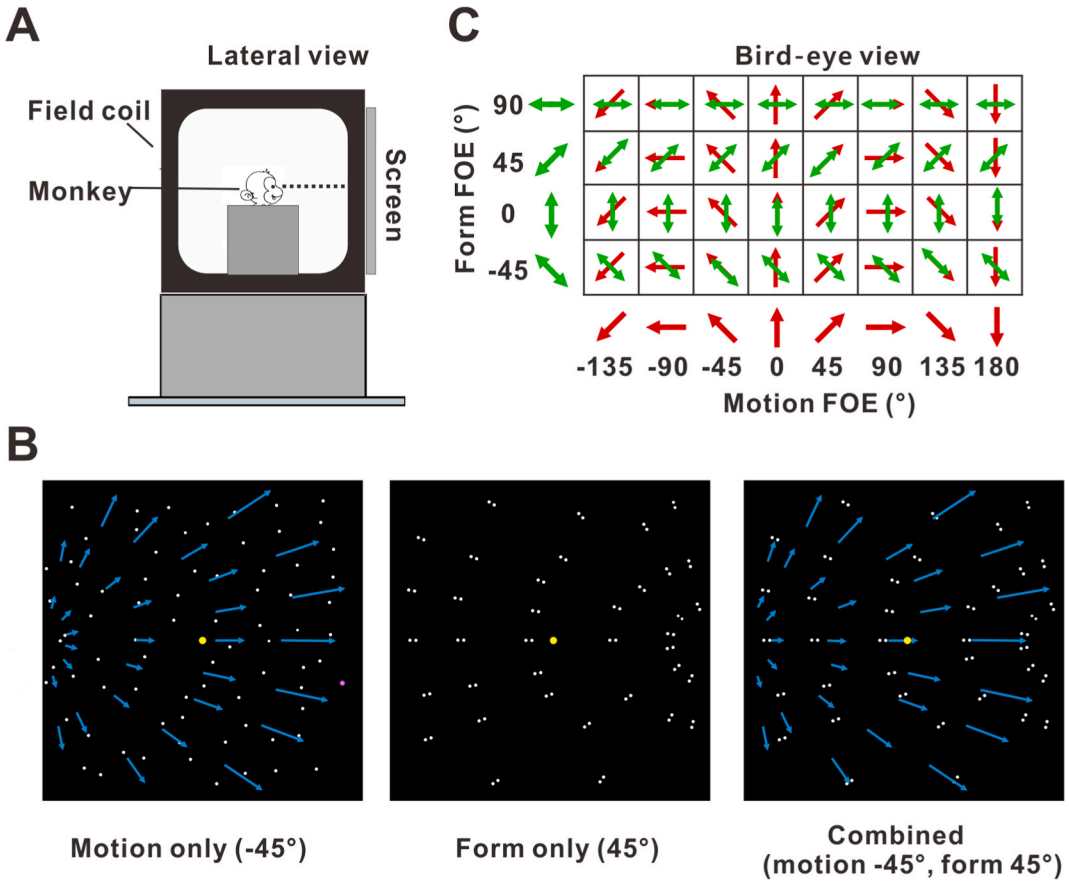


Fig. 1. Equipment setup and stimuli. (A) The lateral view of the experimental apparatus, consists of a screen and field coil. Stimuli were generated on the screen placed ~30 cm in front of the monkey. **(B)** The Optic flow (motion only) condition featured random dots moving away from the motion FOE represented by a radial pattern. Blue arrows indicated the velocity vectors of the dots, depicted as −45° (referred specifically to heading angle). The Static glass pattern (form only) condition consisted of white dot pairs oriented towards the form FOE, depicted as 45°. In the combined condition, white dot pairs oriented towards the form FOE are moving away from the motion FOE in a radial pattern. Blue arrows indicated the velocity vectors of the centroids of the dot pairs, with the motion's FOE depicted as −45° and the form's FOE as 45° in the illustration. Monkeys were required to fixate on a central target (yellow dot) presented on the screen during the stimulus. **(C)** The schematic paradigm included a Form only condition with four form focus of expansion (FOEs, −45°, 0°, 45°, and 90°) in the horizontal plane (bird-eye view). Arrows represented the direction of self-motion. For the motion-only condition, eight motion FOEs (−180°–180°) were evenly spaced every 45° in the horizontal plane. All four form FOEs were paired with all eight motion FOEs, resulting in a total of 32 combined stimuli. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

and form information for heading perception. However, relatively little is known about the form response properties of VIP neurons and how they interact with motion signals for heading perception.

In this study, we examined the form response of VIP by employing Glass pattern stimuli. These stimuli, comprising dot pairs, offer a means to manipulate form information [10,25]. The incorporation of form-defined focus of expansion (FOE) and motion-defined FOE introduces complexity to the stimuli, enabling exploration of neural responses under diverse conditions. Glass patterns, arranged radially around a form-defined FOE, can indicate various heading directions. Concurrently, optic flow moving away from a motion-defined FOE simulates linear translation, capturing the motion-related facet of heading perception. By recording the spatio-temporal response of VIP neurons under conditions of motion-only, form-only, and both presented simultaneously (stimuli composed of randomly distributed dot pairs oriented towards a form-defined FOE and moving away from a motion-defined FOE), our goal is to unravel how VIP neurons respond to form signals. Crucially, we aim to understand the interaction rule between form and motion signals. This approach provides insight on how area VIP integrates these different cues, shedding light on its role in contributing to the overall perception of self-motion. Such findings are essential for advancing our understanding of the neural basis of heading perception.

2. Methods

2.1. Subjects and surgery

Two male rhesus monkeys (Macaca mulatta), specifically monkeys G and N, with weights ranging from 8 to 9 kg, were involved in the experiment. Initially, the monkeys underwent training to familiarize themselves with a custom primate chair and gradually adapt to the laboratory environment. Subsequently, the monkeys underwent chronic implantation of a head-restraint cap and a sclera coil to measure eye movements. Once fully recovered, the monkeys were trained to execute the designated experimental tasks.

2.2. Equipment setup and stimuli

During the experiments, the monkeys were head-fixed and seated in a primate chair. The chair was inside a magnetic field coil frame (Crist Instrument Co., Inc., Hagerstown, MD, USA) mounted on the platform for measuring eye movement with the sclera coil technique (for details, see Ref. [26]).

Visual stimuli were presented on a large computer screen (Philips BDL4225E, Royal Philips, Amsterdam, Netherlands), attached to the field coil frame (Fig. 1A). The display (58 cm × 58 cm) was viewed from a distance of 30 cm, thus subtending a visual angle of 88° × 88°. The sides of the coil frame were covered with a black enclosure, so the monkey could only see the visual stimuli on the screen. The display had a pixel resolution of 1920 × 1080 and was updated at 60 Hz. Visual stimuli were programmed in OpenGL to simulate translation through a 3D cloud of white dots or white dot pairs. The background of the display screen is gray (75 % luminance contrast).

The visual stimulus was either motion-only, form-only, or combined (motion and form stimuli). In the motion-only condition, 2000 white dots were randomly placed in a virtual 3D space of 400 cm wide, 400 cm tall, and 50 cm deep, centered on the central fixation point on the screen. The dots had a diameter of 0.3° and moved outward in a motion-defined focus of expansion (FOE) with a speed of 3.0 m/s. The motion FOE had eight directions evenly spaced every 45° in the horizontal plane between −180° and 180° (Fig. 1B, negative values represented left and positive values represented right). In the form-only condition, 1000 white dot pairs with 0.25° centroid-to-centroid separation were presented on the screen. All dot pairs were oriented toward a location defined as form FOE forming a radial Glass pattern. Because the two form FOEs with a 180° difference were the same, resulting in only four directions evenly spaced every 45° in the horizontal plane between −45° and 90° (Figs. 1B and 0° represented straight ahead). In the combined condition, 1000 white dot pairs were oriented towards a center on the screen. For each frame, the orientation of all dot pairs remained constant but their positions changed (moving outward along a center on the screen). As a result, the screen presented two independent FOEs: the form FOE defined by the orientation of the dot pairs and the motion FOE defined by the centroid of dot pairs moved outward (Fig. 1C). The form and motion cues could be incongruent by various degrees.

2.3. Experimental protocol

During experiments, neuronal responses from area VIP were measured under three experimental conditions described above, including motion-only (8 directions), form-only (4 directions) stimuli, and 32 (4 × 8) combined stimuli. These three stimulus conditions together with a blank trial (no dots or dot pars) were interleaved randomly, resulting in 45 (8 motion-only, 4 form-only, 32 combinations, and a “null” condition) trials for each repetition. In the null condition, stimuli were presented without any coherent form or motion cues, serving as a control to assess baseline neural activity. Each stimulus was repeated 5 times, resulting in a total of 225 trials (45 × 5 = 225).

For each trial, the monkey was required to maintain fixation on a yellow central point (0.3° × 0.3°) for 200 ms, within a 3° × 3° window during the 2 s stimulus presentation. At the end of the trial, successfully maintaining fixation will be rewarded with a liquid. If the monkey breaks fixation, the trial will be terminated and discarded.

2.4. Electrophysiological recordings

We recorded extracellularly from single neurons in area VIP using tungsten microelectrodes (Frederick Haer Company; tip diameter 3 μm; impedance, 1–2 MΩ at 1 kHz). The microelectrode was advanced into the cortex through a transdural guide tube, using a hydraulic microdrive (Frederick Haer Company, Bowdoin, ME, USA). Raw neural signals were amplified, band-pass filtered (400–5000 Hz), digitized, and recorded (AlphaOmega Instruments, Nazareth Illit, Israel). Spikes were sorted online, and spike times along with all behavioral events were collected with 1 ms resolution using the Tempo system. If the online sorting was not adequate, offline spike sorting was performed using the software Spike2 V8 (Cambridge Electronic Design Limited, Cambridge, UK).

Area VIP was identified using a combination of magnetic resonance imaging scans, stereotaxic coordinates, white/gray matter transitions, and physiological response properties, as described in detail previously [15,27,28].

2.5. Data analysis

Data analysis was performed with custom scripts in Matlab R2016a (The MathWorks, Natick, MA, USA).

Temporal response analysis: To analyze VIP neural responses to visual stimuli, we first constructed peri-stimulus time histograms (PSTHs) for each stimulus. PSTHs were computed using 25-ms time bins and were smoothed with a 400 ms boxcar filter. We identified the 400-ms time windows at which firing rates reached the maximum values. We then computed the mean firing rates (FRs) for the whole duration of each stimulus presentation (t = 0–2 s) for each stimulus. FRs were then compared with a baseline response calculated as the average firing rate during the 0.5 s period preceding stimulus onset (−1 to −0.5 s). A neuron to be considered to have a significant temporal response to the stimulus showed a significantly different FR response from baseline (Wilcoxon rank test, $p < 0.05$).

Spatial tuning analysis: We used One-way ANOVA to assess the significance of tuning in the motion-only and form-only conditions, then we divided the neurons into four classes: "Both" neurons with significant tuning to both motion-only and form-only conditions ($p_{motion} < 0.05$ & $p_{form} < 0.05$, one-way ANOVA test); "Motion only" neurons with significant tuning only to motion stimuli ($p_{motion} < 0.05$ & $p_{form} > 0.05$, one-way ANOVA test); "Form only" neurons with significant tuning only to form stimuli ($p_{form} > 0.05$ & $p_{motion} < 0.05$, one-way ANOVA test); "Not-tuned" neurons without significant tuning to motion or form stimuli ($p_{motion} > 0.05$ & $p_{form} > 0.05$, one-way ANOVA test). The significance of motion and form tuning (main effects) in the combined responses was assessed with two-way ANOVA. Due to the absence of significant tuning to form-only stimuli, we exclusively calculate motion FOE preferences for neurons with significant tuning to motion-only stimuli using the vector sum of mean FR responses. The vector sum method determined a neuron's preferred direction by converting its firing rates at different heading angles into vectors, which were then summed to obtain a resultant vector [29]. The direction of this resultant vector indicated the neuron's preferred direction.

Congruent Index: To investigate the congruence between neuronal responses to form and motion cues in VIP neurons, we first transformed the 32 combined responses (4 form × 8 motion FOEs) into 8 form × 8 motion FOE matrix, thus form FOEs were ranged from −180° to 180° (responses of two FOEs that differed by 180° were consistent) the same as motion FOEs. We computed the firing rate for various form FOEs (each being the mean of the same form FOE with 8 different motion FOEs), and various motion FOEs (each being the mean of the same motion FOE with 8 different form FOEs). Then we computed the normalized firing rate of various FOEs:

$$NF_i = \frac{F_i - \overline{F}}{\sqrt{\sum_{i=1}^{8} (F_i - \overline{F_i})^2}} \quad \text{(Eq 1)}$$

$$NM_i = \frac{M_i - \overline{M}}{\sqrt{\sum_{i=1}^{8} (M_i - \overline{M})^2}} \quad \text{(Eq 2)}$$

Where NF_i and NM_i represented the normalized firing rate for the i-th form and motion FOE for each neuron, F_i and M_i were the firing rate for the i-th form and motion FOE, $\overline{F}$ and $\overline{M}$ were the mean firing rate averaged over 8 form and motion FOEs).

Circular cross-correlation was calculated using the following equation:

$$c(k) = \sum_{i=1}^{8} \frac{NF_i * NM_{i+k}}{norm(NF) * norm(NM)} \quad \text{(Eq 3)}$$

where $c(k)$ represented the circular cross-correlation value at lag k (where k ranges from 1 to 8), NF_i and NM_{i+k} represented i-th form normalized firing rate and the i-th motion normalized firing rate after k circular shifts, $norm(NF)$ and $norm(NM)$ represented the Euclidean norms of the normalized form and motion firing rates, respectively. Here we defined the Circular Cross-Correlation at $k = 1$ as the Congruent Index.

Wrapped Gaussian fitting: Regarding the combination response of the 8 form × 8 motion FOEs, we defined the response of identical form and motion FOEs as an "aligned" response. We cyclically rotated the "aligned" and "motion only" responses, covering angles ranging from −180 to 180°. Then the "aligned" and "motion only" responses were fitted using the following equation:

$$R(\theta) = A \times \left[e^{\frac{-2\times\left(1-\cos\left(\theta-\theta_{pref}\right)\right)}{\sigma^2}} \right] + R_0 \quad \text{(Eq 4)}$$

where $R(\theta)$ represented the response at θ FOE, A was the amplitude, σ was the half peak width (bandwidth) and θ_{pref} was the preferred FOE of the fitted curve.

We used amplitude modulation index (AMI) and bandwidth modulation index (BMI) to compare the variations between the combined aligned response (adding form) and the motion-only response, which were computed as follows:

$$AMI = \frac{A_{\text{Aligned}} - A_{\text{motion only}}}{A_{\text{Aligned}} + A_{\text{motion only}}} \quad \text{(Eq 5)}$$

$$BMI = \frac{\sigma_{\text{Aligned}} - \sigma_{\text{motion only}}}{\sigma_{\text{Aligned}} + \sigma_{\text{motion only}}} \quad \text{(Eq 6)}$$

where A_{Aligned}, $A_{\text{motion only}}$, σ_{Aligned} and $\sigma_{\text{motion only}}$ were calculated from (Eq (4)).

Linear and nonlinear model of Combined responses: The combined responses were arranged into two-dimensional arrays indexed by the motion and form FOE. These arrays were then visualized using color-contour maps (Fig. 7).

We used both linear and nonlinear models to explore the motion and form interaction in area VIP. For the linear model, combined responses (averaged across 5 repetitions) were fitted by a linear combination of the weighted corresponding motion-only and form-only responses.

$$r_{combined}(\theta,\varphi) = w_{motion} r_{motion}(\theta) + w_{form} r_{form}(\varphi) + C \quad \text{(Eq 7)}$$

where $r_{combined}$ was the predicted response for the combined condition, and r_{motion} and r_{form} were the responses in the motion-only and form-only conditions, respectively. Angles θ and φ represented motion and form defined FOE. w_{motion} and w_{form} were the weights associated with the motion and form responses, respectively, while C is constant. They were chosen to minimize the sum of the squared errors between the predicted and measured combined responses.

The nonlinear model extends the linear model by including additional terms that account for nonlinear interactions between motion and form stimuli as follows.

$$r_{combined}(\theta,\varphi) = w_{motion} r_{motion}(\theta) + w_{form} r_{form}(\varphi) + w_{motion} r^2_{motion}(\theta) + w_{form} r^2_{form}(\varphi) + w_{motion\cdot form} r_{motion}(\theta) r_{form}(\varphi) + C \quad \text{(Eq 8)}$$

where r^2_{motion} and r^2_{form} were the squared responses of motion-only and form-only. $w_{motion\cdot form}$ represented the weight of interaction between motion and form.

For each fit, the R^2 was computed as

$$R^2 = 1 - \frac{SSE}{SST} \quad \text{(Eq 9)}$$

where SSE is the sum squared error between the fit and the data, and SST is the sum of squared differences between the data and the mean of the data. The goodness of fit was assessed for the linear and nonlinear models using Akaike information criterion (AIC) and the Bayesian information criterion (BIC).

Conditional Fisher Information analysis: Fisher Information quantifies the local information about a specific stimulus [29–31]. Following the approach described by Zhao, Wang [26], we calculated the Fisher Information for motion (θ) considering the influence of form (φ). Firstly, we computed the neuron's firing rate conditioned on the offset between motion and form FOE, denoted as $\delta = \theta - \varphi$. Then we computed the Fisher Information conditioning on the offset δ.

$$J(\theta|\delta) = \sum_{i=1}^{N} \frac{f_i'(\theta|\delta)^2}{f_i(\theta|\delta)} \quad \text{(Eq 10)}$$

Considering the direction of motion and form FOE, we classified offsets into four categories $\delta = 0°$, $-45°$, $45°$, $90°$. To assess the impact of incorporating form FOE on the response of neurons to motion FOE, we can calculate the Fisher information for the VIP neural population to motion FOE under different offset conditions.

To calculate conditional Fisher Information (Eq (10)), the tuning curve and its derivative for each neuron were required. For deriving the tuning curve slope, a cubic spline function was employed to interpolate between the data points sampled at a spacing of 45°. The tuning curves were smoothed by convolving with a Gaussian kernel (SD = 10°). The derivative of the tuning curve was then computed as the spatial derivative of the spline fit. Confidence intervals for the population Fisher Information were determined through a bootstrap procedure. This method involved generating random samples of neurons by resampling with replacement from the recorded neuron population. This resampling process was repeated 100 times. Standard errors for both the tuning curve and Fisher Information were calculated for every reference FOE.

3. Results

In our experiments, two monkeys (monkeys G and N) were actively involved. We conducted single-neuron extracellular recordings in areas VIP, specifically focusing on the lower bank and tip of the intraparietal sulcus. This resulted in a total of 133 neurons, with 40 neurons recorded from monkey G and 93 from monkey N. The recordings took place under three conditions within a virtual-reality system (depicted in Fig. 1A): motion-only (optic flow), form-only (static Glass pattern), and combined (Glass pattern integrated with the optic flow). The stimulus pattern used in the experiments is illustrated in Fig. 1B. Each recorded neuron underwent a block of interleaved trials, encompassing 45 distinct stimulus conditions (8 motion-only, 4 form-only, 32 combinations, and a "null" condition). When a combination of form and motion cues was shown, the cues could be incongruent by various degrees, as illustrated in Fig. 1C. In the null condition, stimuli were presented without any form or motion cues, and the monkeys were only required to fixate on a central target. This condition served as a control to assess baseline neural activity. To provide a visual representation, Fig. 1C demonstrated the stimulus set involving motion and form focus of expansion (FOE). Throughout the presentation of stimuli, the monkeys were tasked with maintaining fixation on a central target displayed on the screen. This experimental design allowed us to explore the neural responses in areas VIP under different conditions, shedding light on how single neurons in this region react to motion-only, form-only, and combined stimuli.

3.1. Responses to form-defined FOE: temporal modulation

In Fig. 2, we presented the responses using peristimulus time histograms (PSTHs) of two example neurons under different conditions: motion-only (8 FOEs in the red box), form-only (4 FOEs in the green box), and combined (32 stimuli in the orange box). The PSTHs provided a time-locked representation of neural activity in response to the stimulus onset. Our initial focus was on examining

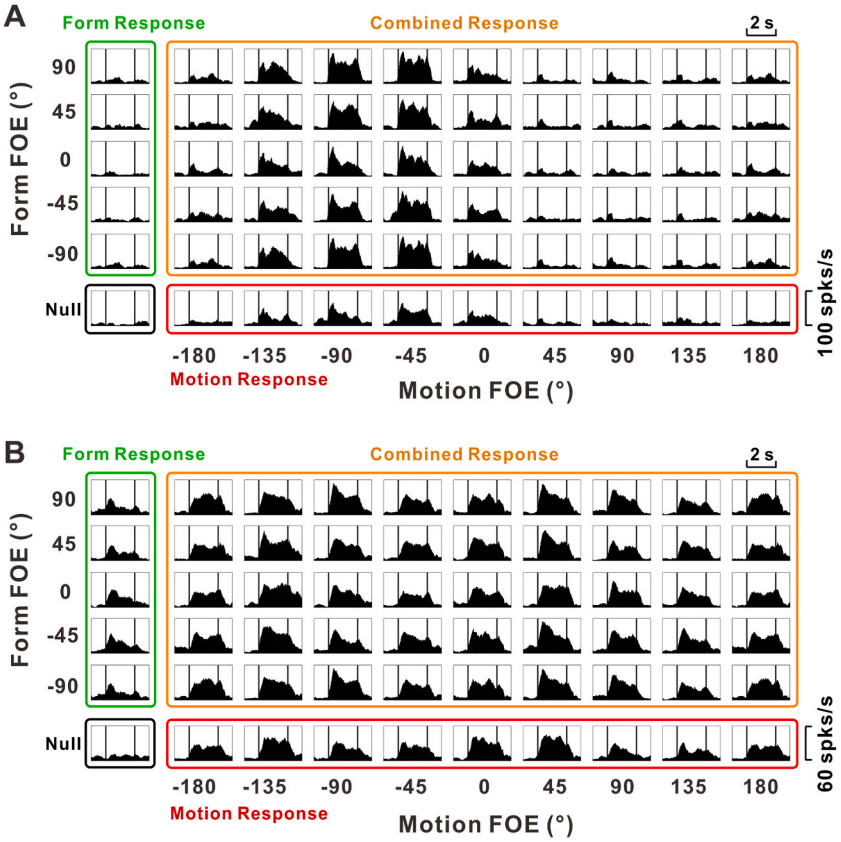


Fig. 2. Responses of two example VIP neurons. Motion FOEs and form FOEs are defined in Fig. 1C. **(A)** Average response PSTHs for a VIP neuron that showed significant temporal modulation to motion-defined FOE, but showed no response to form-defined FOE. The bottom panels in the red box represent the responses to the motion-only condition. The left panels in the green box represent the responses to the form-only condition. The middle part inside the orange box represents the response PSTHs to the different combinations of motion and form FOEs.The panel in the black box represents the response to the null condition. Each PSTH is from 1 s before stimulus onset to 1 s after stimulus offset, with the two black bars in the middle representing stimulus onset and stimulus offset, respectively. **(B)** Average response PSTHs for a VIP neuron showed significant temporal modulation to both motion-defined and form-defined FOE. All the formats are the same as in **(A)**. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

the temporal modulation of neuronal responses to these stimuli. Temporal modulation was considered present when there was a significant difference between the mean firing rate (FR) during the stimulus time and the spontaneous firing. Spontaneous firing was defined as the average FR during the 0.5 s period preceding stimulus onset (−1 to −0.5 s).

In Fig. 2A, the neuron exhibited significant temporal modulation along multiple motion-defined FOE (−135°, −90°, −45°, 0°, $p < 0.05$, Wilcoxon rank-sum test), while showing no significant response to form-defined FOE. Under the combined condition, the example neuron in Fig. 2A exhibited significant responses to 21 stimuli, indicating a specific responsiveness to motion cues and no response to form cues. In Fig. 2B, the neuron demonstrated significant response modulation to all 8 motion-only FOEs (all $p < 0.05$). Notably, it also responded to all form-only FOEs (−45°, 0°, 45°, 90°; $p = 7.9 \times 10^{-3}$, 7.9×10^{-3}, 0.03, and 0.03, Wilcoxon rank-sum test). This suggest that this example neuron was not only responsive to motion cues but also showed a significant response to form cues. These findings provide insights into the distinct neural responses of example neurons under various stimulus conditions, highlighting their sensitivity to motion, form, or a combination of both cues.

In order to quantify the response of neurons to different stimuli, we conducted an analysis by counting the number of FOEs that exhibited temporal modulation for each neuron. The distribution of these counts was illustrated in Fig. 3. For our classification criteria, we defined a neuron as responsive to a specific type of stimulus (motion only, form only, or combined) if it exhibited a significant response to at least one FOE. Observation from this analysis revealed that 26.3 % (35/133) of neurons displayed a significant response to form-only stimuli (Fig. 3A), which was significantly lower ($p = 3.7 \times 10^{-7}$, Z test) than the proportion of neurons showing significant response to motion-only stimuli (57.1 %, or 76/133, Fig. 3B). Notably, 21 % (28/133) of neurons showed significant response modulation to both motion-only and form-only stimuli (refer to Table S1 for detailed information). Under the combined stimulus conditions, the number of neurons displaying significant temporal modulation increased to 102 (Fig. 3C), which was significantly higher than the number responding under single cue conditions (76 + 35–28 = 83, $p = 0.01$, Z test). Additionally, 25 neurons that did not show significant responses under individual cue conditions exhibited significant responses under the combined condition, suggesting that the combined cues can activate neurons that do not respond to single motion or form cues.

Given that the experimental stimuli persisted for 2 s, we were interested in examining the dynamic changes in neural responses over

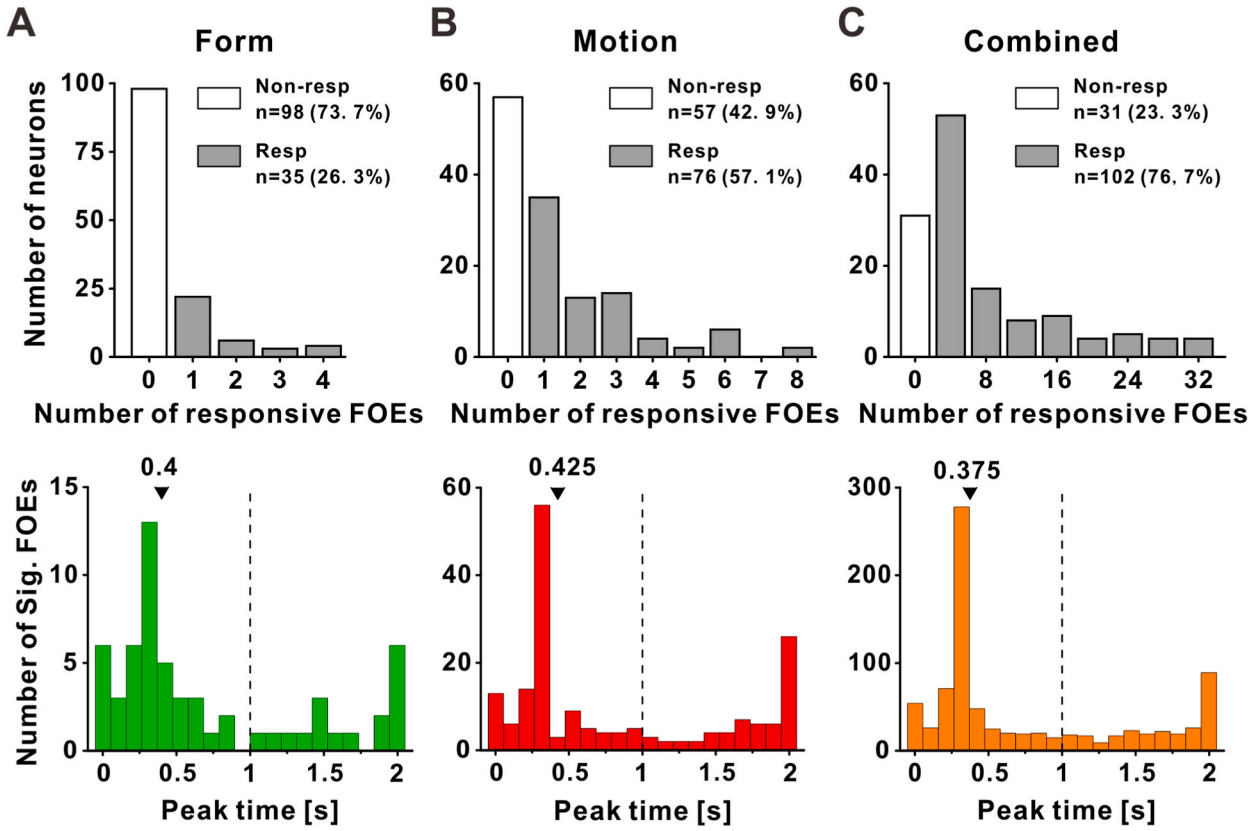


Fig. 3. Summary of temporal response modulation and peak time during different conditions. Distributions of the number of significant response FOEs in form-only **(A)**, motion-only **(B),** and combined condition**(C)**. Gray-filled bars represent neurons' response to stimuli significantly more than one FOE. The white-filled bars represent neurons that do not respond to stimuli. The percentage indicated the proportion of neurons, out of a total of 133, that exhibited or did not exhibit temporal modulation response under specific conditions. Greed, red, and orange bars represent the peak time under form, motion, and combined conditions, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

time. We specifically focus on the distributions of times when firing rates reached their maximum for stimuli with significant response, known as peak time. Analysis of the data revealed that the median time to peak firing rates under motion-only, form-only, and combined stimuli were 0.425 s, 0.4 s, and 0.375 s, respectively (Fig. 3). These values were significantly different from the midpoint (1 s) of the stimulus duration ($p = 6.0 \times 10^{-3}$, $p = 0.03$, and $p = 2.7 \times 10^{-16}$, respectively, Wilcoxon rank-sum test). However, there were no significant differences in peak times between the three conditions (form-only vs. motion-only: $p = 0.41$, form-only vs. combined: $p = 0.71$, motion-only vs. combined: $p = 0.34$, Wilcoxon rank-sum test). These results suggested notable temporal variations in neuronal activity, with neural responses reaching their peak in the early stages of stimulation under all three conditions, even though the stimulus persisted for the entire 2-s duration.

3.2. Responses to form-defined FOE: directional tuning

In analyzing the pattern of temporal responses across different FOEs, our goal was to characterize whether neurons exhibited directional tuning based on the type of stimulus. Taking the example neuron from Fig. 2A, it displayed temporal modulation to 4 different motion-defined FOEs. However, the firing rates in these directions were not uniform, and the maximum response of this neuron was 42.3 spk/s in motion-only conditions when the FOE motion was at −45° ($FOE_{motion} = -45°$). To assess direction tuning, we computed the mean firing rates during 2 s stimulation. This neuron demonstrated significant tuning for motion cues in motion-only conditions ($p_{motion} = 1.3 \times 10^{-11}$, one-way ANOVA). The preferred direction, computed from the population vector sum, was

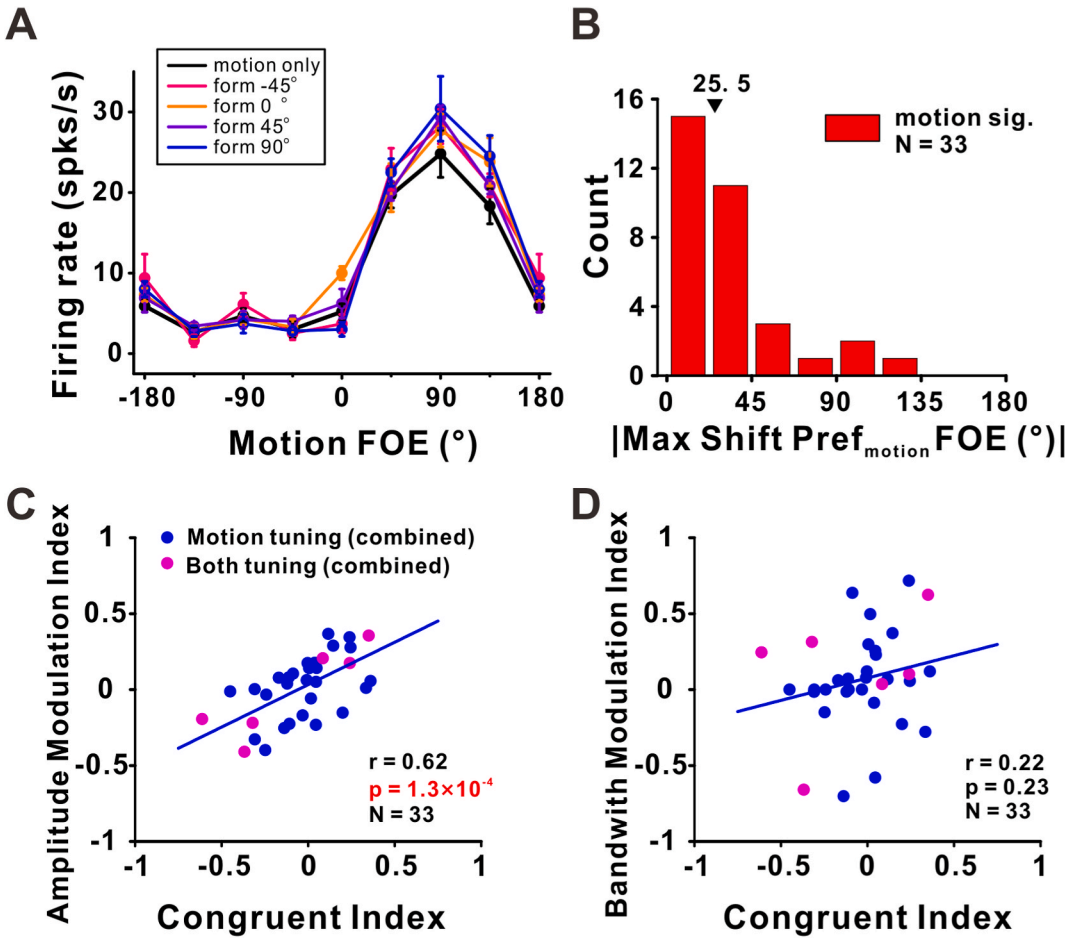


Fig. 4. Tuning properties of VIP neurons. **(A)** The tuning curves of an example neuron. Black lines and symbols are tuning curves during the motion-only condition. Solid lines are tuning curves for motion FOEs with different form FOEs (−45°, 0°, 45° and 90°) indicated by different colors (deep-pink, orange, violet, and blue line, respectively). Error bars represent SEM. **(B)** Distribution of max shift of preferred motion-defined FOE caused by form-defined FOE. Red bars indicate cells with significant tuning for motion-only conditions (N = 33). **(C)** Relationship between amplitude modulation index and congruency index ($r = 0.62$, $p = 1.3 \times 10^{-4}$, $N = 33$). **(D)** Relationship between bandwidth modulation index and congruency index ($r = 0.22$, $p = 0.23$, $N = 33$). The blue dots indicate significant tuning only to motion FOE under the combined condition, the magenta dots indicate significant tuning to both motion and form FOE under the combined condition. The solid lines illustrate the regression lines of the data. r, Pearson's correlation coefficient. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

−66.2°. In contrast, no significant responses were observed for form-defined FOEs, and the neuron showed no significant directional tuning in response to form cues in form-only condition ($p_{form} = 0.38$, one-way ANOVA). However, under combined conditions, the maximum response increased to 64.1 spk/s ($FOE_{motion} = -45°$, $FOE_{form} = 90°$, $p = 0.03$, Wilcoxon rank-sum test). The neuron showed significant tuning to both motion and form cues ($p_{motion} = 8.6 \times 10^{-50}$, $p_{form} = 0.03$, respectively, two-way ANOVA).

Examining Fig. 2B, the neuron did not exhibit tuning to the form-defined FOEs ($p_{form} = 0.58$, one-way ANOVA), despite displaying significant temporal modulation in response to all four form-only cues ($p < 0.05$, Wilcoxon rank-sum test). However, there was significant tuning to motion-defined FOEs ($p_{motion} = 1.3 \times 10^{-4}$, one-way ANOVA) with a maximum response of 35.0 spk/s ($FOE_{motion} = 45°$) under motion-only conditions. In the combined condition, the maximum combined neuronal response increased to 35.4 spikes/s when motion and form cues were aligned ($FOE_{motion} = 45°$, $FOE_{form} = 45°$), which was not significantly different from the maximum response observed under motion-only conditions ($p = 0.94$, Wilcoxon rank-sum test). The neuron showed significant tuning to motion cues ($p_{motion} = 1.7 \times 10^{-9}$, two-way ANOVA), but no significant tuning to form cues ($p_{form} = 0.19$, two-way ANOVA).

Across all neurons, 24.8 % (33/133) of neurons exhibited significant directional tuning to motion-only stimuli, and no neurons were tuned to form-only stimuli (see Table S2 for details). Under the combined condition, the proportion of neurons significantly tuned only to motion cues was 36.1 % (48/133), while the proportion significantly tuned to both form and motion cues was 4.5 % (6/133). Compared to motion-only conditions, the proportion of neurons significantly tuned to motion cues under combined conditions (36.1 % + 4.5 % = 40.6 %) significantly increased ($p = 6.0 \times 10^{-3}$, Z test). This suggested that the introduction of form cues increased the neuron tuning sensitivity to motion cues.

3.3. Relationship between form and motion responses

To investigate whether the influence of form on motion signal processing can be attributed to alterations in motion-defined FOE preferences, we first plotted the motion-tuning curves for stimuli occurring at each motion FOE, then we computed the motion FOE preferences based on the vector sum of these tuning curves. It's important to note that we only selected neurons that were significantly tuned for motion-only cues (N = 33), as there were no neurons tunned to form-only cues.

As shown in Fig. 4A, the example neuron exhibited a preferred FOE to the right (112.1°) in the motion-only condition (represented by the black line). When form cues were added, the preferred motion FOEs were 114.6°, 112.1°, 112.9°, and 115.8° in response to form FOEs at −45°, 0°, 45° and 90° respectively (represented by the deep-pink, orange, violet, and blue line). Consequently, the maximal shift in motion FOE preference between combined responses and motion-only conditions was only 3.7°. Among the 33 neurons with significant motion tuning, the median shift in preferred motion FOE was 25.5° (Fig. 4B), a value significantly less than the resolution limit of our sampling (45°) ($p = 4.4 \times 10^{-3}$, One-sample *t*-test). Therefore, this finding suggests that the form cues did not significantly alter the tuning preference.

Given that the response of neurons to motion-defined FOEs was influenced by different form-defined FOEs, such as in combined conditions, where the maximum response in Fig. 2A corresponded to inconsistent motion and form FOEs ($FOE_{motion} = -90°$, $FOE_{form} = 90°$), and in Fig. 2B, the maximum response corresponded to consistent FOEs ($FOE_{motion} = 45°$, $FOE_{form} = 45°$). We next investigated whether the changes in neuron tuning curves were related to the consistency of neuronal tuning between motion and form.

Since the majority of neuron exhibited tuning only to motion cues and not to form cues when quantifying consistency, we first pool motion responses across different form FOEs and form responses across different motion FOEs under combined conditions. Then we calculated the congruent index (CI) for each neuron (for details, see **Methods**). To assess the influence of form on motion, we compared

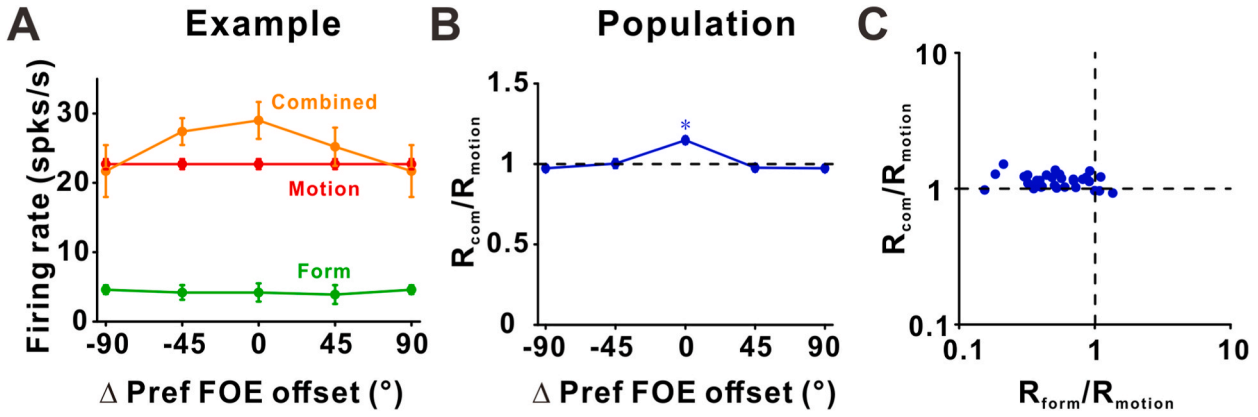


Fig. 5. The neuron response magnitude for motion-defined FOE changed with the addition of form-defined FOE. **(A)** An example neuron with motion (red), form (green), and combined (orange) responses plotted as a function of the offset (Δ°) of the preferred FOE. 0° indicated when both motion and form cues were presented together at the preferred motion and form FOE of the neuron. **(B)** The mean ratio of combined response to motion-only response is plotted as a function of the offset (Δ°). The '*' symbol indicates a significant difference from the value 1. **(C)**The combined ratio (R_{com}/R_{motion}) is plotted as a function of the unimodal ratio (R_{form}/R_{motion}). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

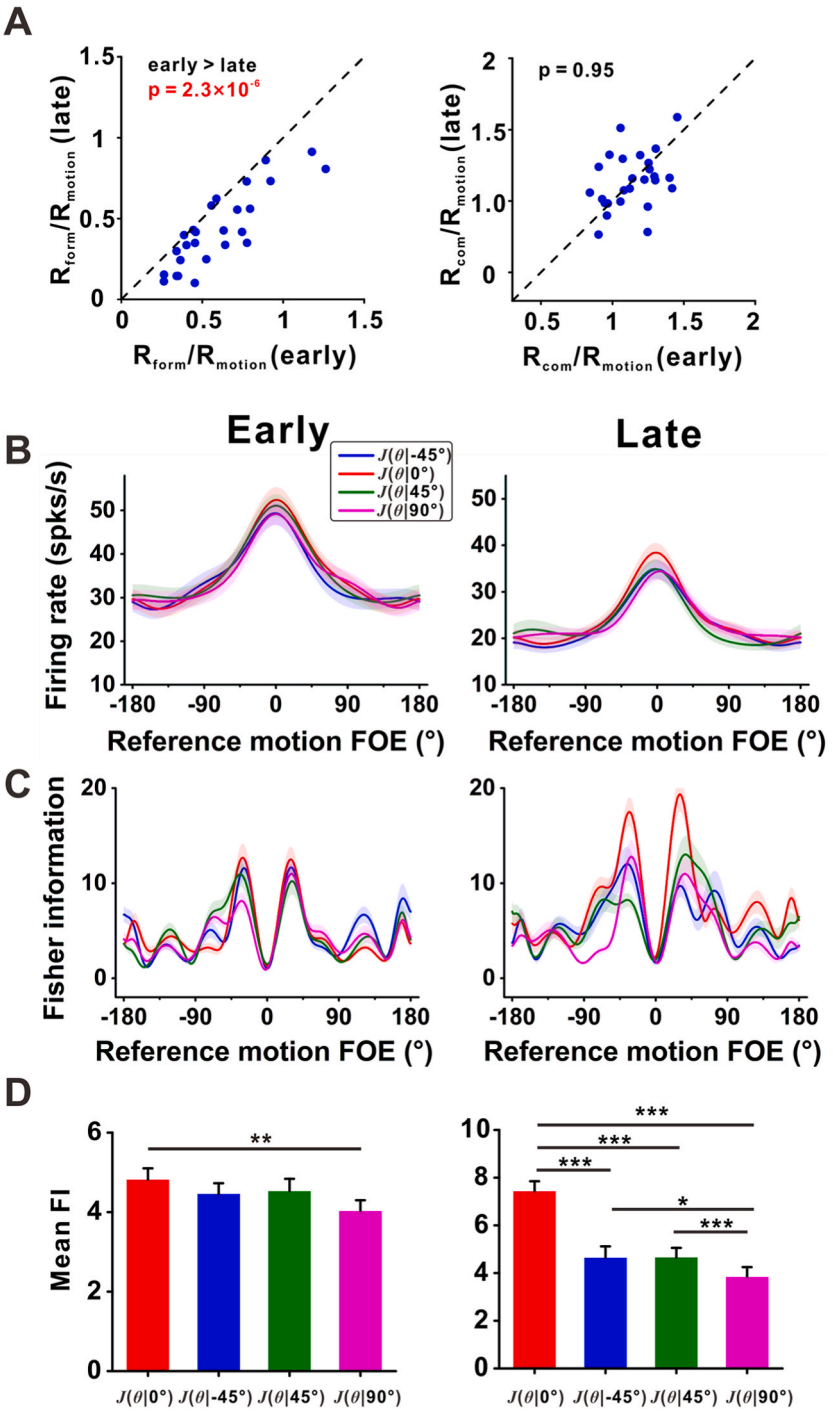


Fig. 6. Comparison of neuronal responses and conditional Fisher Information in the early and late stages. **(A)** The unimodal ratio and combined ratio are depicted as late stage vs. early stage. P values are presented on the corresponding plots (two-tailed paired t-tests). **(B)** The average response to motion-defined FOEs, with different color lines plotted based on varying offsets of motion-defined FOEs and form-defined FOEs. These tuning curves for each neuron were aligned at their preferred motion FOE. Shades indicate standard error. **(C)** Conditional Fisher Information for population neuronal response with different offsets. These curves for each neuron were aligned at their preferred motion FOE. The error bar indicates standard error. **(D)** Average Fisher Information for population neuron. The average is taken over all motion-defined FOEs. The red, blue, green, and magenta curves or bars represented the offset between motion-defined and form-defined FOEs were 0°, −45°, 45°, and 90°, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

the neuronal tuning curve for motion FOEs under aligned combination conditions (where the form FOE and motion FOE were identical) with that under motion-only conditions. Each tuning curve was fitted using a wrapped Gaussian equation, from which we derived the amplitude and bandwidth of the tuning curve. Subsequently, we calculate changes in amplitude and bandwidth to obtain the amplitude modulation index (AMI) and bandwidth modulation index (BMI).

A CI value of 1 indicates that the neuron's preference directions for motion and form are completely aligned, while a CI value of −1 indicates they are completely opposite. As shown in Fig. 4C, there was a significant positive linear correlation between AMI and CI ($r = 0.57$, $p = 5.3 \times 10^{-4}$, $N = 33$, Pearson correlation). This correlation suggested that as the CI value moved away from 0 and closer to either 1 or -1, the influence of form information on the neuron's response magnitude became more pronounced, either enhancing it (when CI approached 1) or reducing it (when CI approached −1). But there was no significant correlation between BMI and CI in Fig. 4D ($r = 0.22$, $p = 0.22$, $N = 33$, Pearson correlation). This suggested that the variability in tuning width was not strongly influenced by the consistency of neuronal tuning between motion and form.

To better visualize the impact of different form-defined FOEs on motion-tuning curves, we aligned the responses across motion-only, form-only, and combined conditions (Fig. 5). 0° indicated when both motion and form cues were presented together at the preferred motion and form FOE of the neuron. Fig. 5A shows the response of a single neuron. When the combined responses (where the motion FOE matches the preferred motion-only FOE) were aligned, they peaked at zero offsets ($\Delta = 0°$) and were larger ($p = 7.9 \times 10^{-3}$, Wilcoxon rank-sum test) than the preferred motion-only responses (red, Fig. 5A). These responses decreased as the absolute value of the offset, $|\Delta|$, increased (orange, Fig. 5A). The form-only responses (green, Fig. 5A) were consistently smaller ($p = 2.3 \times 10^{-3}$, Wilcoxon rank-sum test) than the preferred motion-only responses. This pattern was also observed at the population level, as shown in Fig. 5B. The ratio of the combined response to the motion-only response reached its maximum value at the preferred FOE, significantly exceeding 1 ($p = 6.6 \times 10^{-7}$, One-sample *t*-test), and decreased as the absolute value of the offset increased.

To further investigate whether the enhancement at preferred FOE was related to the relative strength of neuronal responses to motion and form cues, we plotted the measured values of neuronal response enhancement (R_{com}/R_{motion}, extracted from ΔFOE offset at 0°) against the relative strength of two unimodal signal inputs (R_{form}/R_{motion}, responses at preferred FOE). R_{com} represented the response of neurons at the preferred FOE for both motion and form cue, indicating the maximum firing rate in combined condition. R_{motion} represented the maximum response of neurons at the preferred FOE under motion-only condition, while R_{form} represented the maximum response of neurons at the preferred FOE under form-only condition. Most data points were distributed in the top-left quadrant, in which form responses were smaller than motion responses and the maximum combined responses were larger than the maximum motion unimodal responses (Fig. 5C). The correlation between the enhancement and the ratio of form vs. motion responses was not significant ($r = -0.24$, $p = 0.17$, $N = 33$, Pearson correlation), indicating that the enhancement was not strongly influenced by the relative strength of form and motion cues.

3.4. The influence of form on motion is greater in the late stage compared to the early stage

Although the motion cues maintained a constant velocity and the form cue remained static during the whole (2 s) stimulation, the neuronal responses were not constant across the entire stimulation period. From the peak time distribution of neurons responding to stimuli, it can be seen the median time to reach the peak response was around 0.4 s. Additionally, there was a noticeable distribution towards the end of the stimulus period, as depicted in the bottom row of Fig. 3. To delve into variations in neuronal response post-stimulation onset, we categorized the neuronal response into two stages: the early stage (0–1 s) and the late stage (1–2 s).

We did not observe significant differences in the number of neurons with temporal modulation (Fig. S1) and directional tuning (See Table S3 for details) between the early and late stages. However, when calculating the maximum shift of the preferred motion FOE with the addition of form signals, we found that the median offset in the late stage (28.9°) was smaller than in the early stage (40.3°) ($p = 8.8 \times 10^{-3}$, Wilcoxon rank-sum test), as illustrated in Fig. S1. This suggests a more fine-tuned and more stabilized neural representation in response to the combined form and motion stimuli during the later stage.

In our further analysis, we compared the differences in the impact of form signals on motion response, as well as the relative response strength of form vs. motion in the two stages. Across the sample neurons, we observed a significant decrease in the relative strength (R_{form}/R_{motion}) compared to the early stage ($p = 2.3 \times 10^{-4}$, paired t-tests; left panel in Fig. 6A). However, the neuronal response enhancements (R_{com}/R_{motion}) showed no significant change ($p = 0.95$, paired t-tests; right panel in Fig. 6A) in the late stage. This suggested that the influence of form cues on motion cues differed between the early and late stages. In the later stage, as the offset between the form cue and the preferred FOE increased, the combined response in the preferred motion-only FOE significantly decreased, as illustrated in Fig. S2. These findings shed light on the nuanced dynamics of how form signals impact motion response across different temporal stages.

To examine how different form-defined FOEs affect neuronal response to motion cues during the early and late stages, we employed Fisher information decoding (for details, see **Methods**). Initially, we generated average tuning curves for the VIP neural population with respect to motion FOE, conditioned on different offset conditions in both stages. The red, blue, green, and magenta curves were aligned at preferred motion FOE represented by $\delta = 0°$, −45°, 45°, and 90°, respectively (Fig. 6B). Notably, max firing rates occurred at 0° of reference motion FOE, decreasing away from 0°. Subsequently, we plotted the average population Fisher Information at different reference motion FOE (Fig. 6C). The minimum Fisher Information coincided with 0°, where maximum firing rates were observed. The Fisher Information represented the corresponding neuron's derivative of the tuning curve to reference motion FOE. These analyses provided insights into how neural populations respond to motion cues under varying form-defined FOE conditions in both early and late stages.

In our further analysis, we calculated the average population Fisher Information across all motion FOE conditions (Fig. 6D). In the

early stage, the mean Fisher Information under 0° offset (red bar) was significantly greater than under 90° offset (magenta bar, $p = 0.01$). However, there was no significant difference compared to −45° offset (blue bar, $p = 0.20$) or 45° offset (green bar, $p = 0.35$, paired t-tests). This suggests that in the early stage, when the offset between form FOE and motion FOE was less than 90°, the neural population contained similar levels of information about the motion FOE θ, with no significant differences. Contrastingly, in the late stage, the mean Fisher Information under 0° offset was significantly larger than offset of −45° ($p = 2.4 \times 10^{-9}$), 45° ($p = 4.0 \times 10^{-11}$) and 90° ($p = 6.2 \times 10^{-19}$, paired t-tests). Additionally, the mean Fisher information at −45° offset and 45° offset was greater than at 90° offset ($p = 0.04$, $p = 9.9 \times 10^{-4}$, respectively). This suggests that at small offsets, the neural population contained more information about the motion FOE θ, thereby enhancing sensitivity to motion FOE. As the offset increased, this enhancement effect diminished, indicating a reduction in the influence of form on motion.

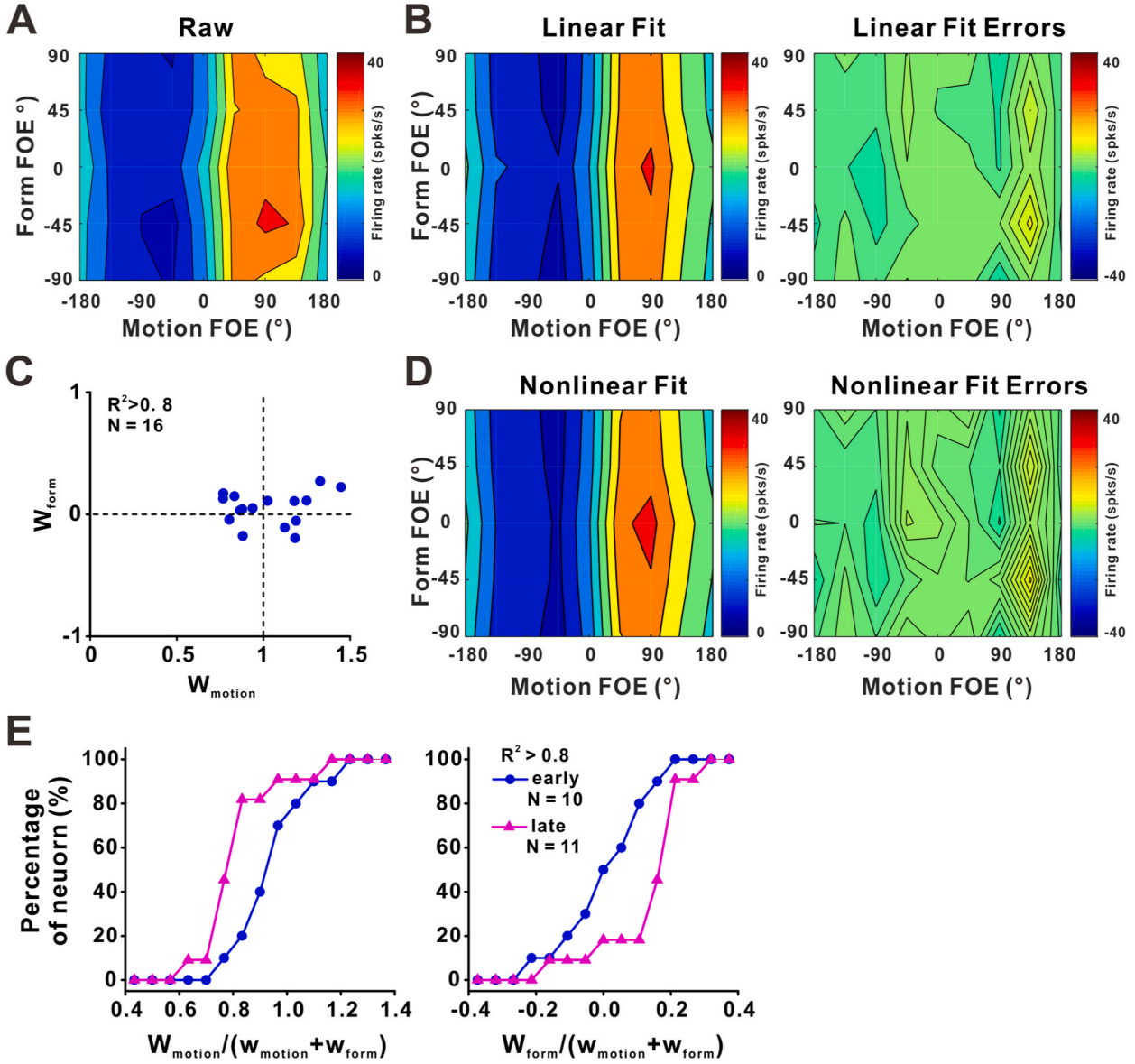


Fig. 7. Example fitting data and population summary of linear and nonlinear models. **(A)** Contour map of raw combined response. **(B)** Left panel: contour map of linear fitting combined response. R^2: 0.946. Right panel: contour map of the errors that raw data minus linear fitting data. **(C)** Scatter plot of motion weight against form weight in VIP. **(D)** Left panel: contour map of nonlinear fitting combined response. R^2: 0.948. Right panel: contour map of the errors that raw data minus nonlinear fitting data. **(E)** The cumulative distribution of the motion contribution (left panel) and form contribution (right panel) for neurons with good linear fitting ($R^2 > 0.8$) in the early stage and late stage.

3.5. Linear models better fit the responses of VIP neurons to motion and form compared to nonlinear models

To better understand the rules of interaction between motion and form cues, we used linear and nonlinear models to fit neuronal responses, respectively. Firstly, we used a linear model to fit the combined response of a single neuron with a linear sum of responses from the motion-only and form-only conditions (for details, see **Methods**). The combined responses (mean firing rates during stimuli) obtained from the PSTHs were visualized as a color contour map with the motion FOEs along the abscissa and the form FOEs along the ordinate (Fig. 7A). For this example, the linear model (left panel in Fig. 7B) predicted the combined response profile very well, with an R-squared (R^2) value of 0.946. The motion and form weights derived from the linear model fit were 1.02 and 0.51, respectively. The weight reflected the contribution or impact of each cue on the overall neural response. The sign of the weights indicated the direction of influence. Based on the goodness of fit, we summarized neurons with R^2 greater than 0.8 (Fig. 7C), the linear model within VIP revealed that the form weights were quite small (mean ± SEM: 0.05 ± 0.03, N = 16), while the motion weights were nearly close to 1 (mean ± SEM: 1.03 ± 0.05, N = 16). The motion weights were significantly higher than that of the form weights ($p = 4.1 \times 10^{-11}$, paired t-tests), indicating a much stronger weighting of motion signals compared to form signals in VIP.

We further used a nonlinear model to fit the combined response, and the result showed that it provided a more accurate prediction for the combined response as well with an R^2 value of 0.948 (left panel in Fig. 7D). The R^2 values of the nonlinear model were higher than that of the linear model. However, for the example neuron in Fig. 7, both the Akaike information criterion and Bayesian Information Criterion values of the linear model (AIC: 213.7, BIC: 219.1) were lower than those of the nonlinear model (AIC: 218.6, BIC: 229.4), indicating that the linear model was superior in this case, and the population results were consistent with it. The use of linear models allowed for quick and accurate processing of visual information related to self-motion.

Next, we applied a linear model to fit the neuronal responses in both early and late stages. For neurons with good fitting (i.e., R^2 values greater than 0.8), we observed interesting patterns. In the late stage, the motion weights (mean ± SEM: 0.99 ± 0.03, N = 11) were significantly smaller (P = 0.022, Wilcoxon rank-sum test) compared to the early stage (mean ± SEM: 1.12 ± 0.04, N = 10), while the form weights (mean ± SEM: 0.22 ± 0.05) were significantly larger (P = 0.045, Wilcoxon rank-sum test) than those in the early stage (mean ± SEM: 0.05 ± 0.05) (Fig. 7E). The weight changes suggested a shift in the relative contributions of motion and form information to the neural responses. The reduction in motion weights in the late stage may indicate a decreased reliance on motion information, whereas the increase in form weights suggests an amplified influence of form cues on motion perception. This insight provides valuable information about how the neural processing dynamics change between the early and late stages of stimulation.

4. Discussion

In the current study, we conducted recordings of VIP neurons responding to stimuli involving motion-only, form-only, and combined conditions with varying conflicting offsets. Our findings revealed that, under motion-only conditions, neurons exhibited a robust tuning response to optic flow, consistent with previous research. Additionally, while neurons did not show significant direction selectivity to form-only stimuli, some neurons still displayed temporal modulation. Introducing form signals into combined stimuli resulted in increased neuronal responses, indicating an interaction between motion and form signal in VIP. Linear model analysis demonstrated that the combined response of individual neurons could be well captured, with motion weight significantly greater than form weight. Further analysis using conditional Fisher information revealed the difference in the influence of form on motion cues during the stimulus time course. In the late stage of stimulation, as the offset between motion and form FOE increased, the information content of neurons regarding motion significantly decreased. This suggested that with small offsets, the form signal tended to enhance the perceptual sensitivity of neurons to motion cues, particularly during the late stage of stimulus. However, this difference was less apparent in the early stage.

4.1. VIP neurons respond to form signals

The growing body of physiological studies challenges the notion of independent processing in the dorsal and ventral pathways, indicating a close connection between the processing of visual motion and form cues [32,33]. Early visual areas such as V1, V2, and V3 have been shown to respond to changes in FOE positions defined by motion or form cues, highlighting their involvement in processing motion and form information [34–40]. Moreover, studies using magnetoencephalography (MEG), such as Liu, Wang [41] have explored the recurrent brain connections involved in global form processing in dynamic Glass Patterns. Their findings indicated an association between the perceptual integration of form cues and consistent responses in the visual dorsal pathway.

VIP, as a higher-level multisensory area in the dorsal pathway [42], exhibits clear visual direction selectivity [15,28] and receives inputs primarily from MT and MST [43,44], along with partial input from V4 [45]. The experimental findings presented here contribute to this understanding, suggesting that VIP neurons not only respond to motion cues but also to form cues. Furthermore, there is an observed interaction between these two cues, with the addition of form cues enhancing the response of neurons to the motion-preferred FOE.

4.2. The influence of form on motion perception in VIP

The observed directionally tuned responses of VIP neurons to optic flow contribute significantly to heading perception. Under motion-only conditions, optic flow provides directional information, defining the FOE for motion. On the other hand, under form-only conditions, the FOE defined by the form also provides implied motion information. Fisher information analysis sheds light on the gain

modulation of form on motion, showcasing variability based on the offset between the two FOEs. Specifically, when the FOEs are consistent (i.e., the offset is 0), neurons exhibit maximum responses under combined stimuli, indicating a greater gain modulation. This implies that neurons integrate both motion and form cues, enhancing sensitivity to heading perception and incorporating more motion information. As the offset increases (i.e., the offset is ±45, 90), this enhancement effect gradually diminishes, suggesting that the perceptual influence of form on motion is smaller when there is a larger offset between the two FOEs. This implies that the differences or consistency between form and motion cues can affect the brain's processing and integration of these two signals to perceive the motion direction.

Notably, neurons exhibit varied responses to stimuli over the 2 s duration, with peak times predominantly occurring around 0.4 s at the stimulus onset. The division of response into early (0–1s) and late (1–2s) stages reveals that the introduction of form signals has different effects on neuron responses in these periods. In the late stage, with the offset values between the two cues increased, there is a notable decrease in the motion information contained in neurons. This indicates heightened sensitivity to form information in the late stage, particularly when the form FOEs align with motion FOEs aiding in heading perception. The enhancement effect of form on motion direction sensitivity significantly weakened with increasing offsets, presenting a distinct stepwise pattern. In contrast, this stepwise weakening was not evident in the early stage, indicating that the inclusion of different forms did not significantly alter neuronal sensitivity to motion in the early stage.

The observed decrease in neural discharge in the late stage compared to the early stage may be attributed to stimulus repetition, a phenomenon known to lead to lower neural activity while enhancing performance (i.e., priming; Grill-Spector, Henson [46]). However, the presence of global form information in visual stimuli, as seen in dynamic Glass Patterns [47], enhances sensitivity in discerning the motion between frames. This suggests an accumulation of global form signals between frames.

Furthermore, we utilized a linear model to fit neuronal responses under unimodal and combined stimuli conditions. We observed that some neurons integrate form and motion stimuli in a linear additive manner. Specifically, we found that in the early stage of stimulation, the weight of motion is relatively large compared to form, indicating a substantial influence of motion cues. However, as stimulation progresses into the later stage, we observed an increase in the weight of form cues relative to the early stage, accompanied by a decrease in the weight of motion cues.

In addition to the changes in linear regression weights, we performed Gaussian fitting on the firing activity of neurons during the early and late stages to calculate the impact of form inclusion on the amplitude and width of motion tuning at different stages. The results showed a significant positive correlation between the change in tuning amplitude (AMI) and the congruence index in the late stage ($r = 0.47$, $p = 0.01$, $N = 33$, Pearson correlation), while there was no significant correlation in the early stage ($r = 0.06$, $p = 0.76$, $N = 26$, Pearson correlation) (Fig. S2). This indicates that the enhancement of neuronal firing activity by form primarily occurs in the late stage. Based on these findings, we hypothesize that VIP neurons integrate form and motion information in a dynamic manner, with the relative influence of form cues increasing over time. This suggests a temporal evolution in the integration process, wherein form cues become increasingly prominent in shaping neural responses during the later stages of stimulation.

4.3. Limitations of the study

Our study revealing the gain modulation of VIP neurons in perceiving motion signals by form signals provides valuable insights. However, to further elucidate the correlation between neural responses and perceptual behavior, future experiments could consider introducing active tasks that require monkeys to engage in activities such as heading judgment. This shift towards active tasks may provide a more comprehensive understanding of how VIP neuron activity translates into observable behavior.

Moreover, within the visual pathway, other brain regions exhibit directionally tuned responses to optic flow, such as MST [11–13] and STP (Superior Temporal Polysensory area) [17]. These regions are situated in the dorsal part of the superior temporal sulcus, with STP anterior to MSTd. Both regions have relatively large receptive fields, with STP neurons occasionally having larger receptive fields that sometimes extend to the opposite side [48]. While MST and STP are primarily located in the visual dorsal pathway, it's worth noting that the anterior part of STP also receives projections from the ventral pathway [49]. Whether there is also an interaction between visual motion cues and form cues in MST and STP during self-motion perception remains unclear. Exploring these interactions in these additional brain regions could provide a more comprehensive understanding of how various components of the visual pathway contribute to the integration of motion and form cues in the context of self-motion perception.

5. Conclusion

In summary, our study sheds light on the intricate mechanisms underlying the interaction between motion and form information in VIP, offering fresh insights into the role of higher-level brain areas in visual integration. The findings suggest that the incorporation of form information plays a crucial role in VIP, enabling the brain to refine the perception of self-motion. This nuanced understanding contributes to the broader comprehension of how visual processing in higher-level brain regions involves the integration of multiple cues to construct a comprehensive perceptual experience.

Data availability statement

The raw data for all figures, and the main codes involved in data analysis have been deposited at Mendeley and are publicly available as of the date of publication. The link is https://data.mendeley.com/drafts/5w6pr9ps3w.

Ethical statement

All animal surgeries and experimental procedures received approval from the Institutional Animal Care and Use Committee at East China Normal University (IACUC protocol number: Mo20200101).

CRediT authorship contribution statement

Lingqi Kong: Writing – original draft, Investigation, Formal analysis, Data curation. **Fu Zeng:** Writing – review & editing, Writing – original draft, Visualization, Project administration, Investigation, Formal analysis, Data curation. **Yingying Zhang:** Writing – original draft, Funding acquisition, Data curation. **Li Li:** Writing – review & editing, Writing – original draft, Conceptualization. **Aihua Chen:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Software, Resources, Project administration, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by grants from the “STI2030-major projects” (No.2021ZD0202600), the National Basic Research Program of China (No. 32171034), Shanghai Municipal Science and Technology Major Project (No. 2021SHZDZX) to A.C, and National Natural Science Foundation of China (No. 31800866) to Y.Z. We thank Minhu Chen for outstanding computer programming.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e36913.

References

[1] D.C. Burr, J. Ross, Direct evidence that "speedlines" influence motion mechanisms, J. Neurosci. 22 (19) (2002) 8661–8664.
[2] W.S. Geisler, Motion streaks provide a spatial code for motion direction, Nature 400 (6739) (1999) 65–69.
[3] R. Donato, A. Pavan, G. Campana, Investigating the interaction between form and motion processing: a review of basic research and clinical evidence, Front. Psychol. 11 (2020) 566848.
[4] D.C. Bradley, et al., Mechanisms of heading perception in primate visual cortex, Science 273 (5281) (1996) 1544–1547.
[5] B. Krekelberg, et al., Neural correlates of implied motion, Nature 424 (6949) (2003) 674–677.
[6] J. Ross, The perceived direction and speed of global motion in Glass pattern sequences, Vis. Res. 44 (5) (2004) 441–448.
[7] G. Mather, et al., Psychophysical evidence for interactions between visual motion and form processing at the level of motion integrating receptive fields, Neuropsychologia 50 (1) (2012) 153–159.
[8] A. Pavan, R.B. Marotti, G. Mather, Motion-form interactions beyond the motion integration level: evidence for interactions between orientation and optic flow signals, J. Vis. 13 (6) (2013) 16.
[9] J.J. Gibson, The Perception of the Visual World, 1950.
[10] D.C. Niehorster, J.C. Cheng, L. Li, Optimal combination of form and motion cues in human heading perception, J. Vis. 10 (11) (2010) 20.
[11] K.H. Britten, Mechanisms of self-motion perception, Annu. Rev. Neurosci. 31 (2008) 389–410.
[12] Y. Gu, D.E. Angelaki, G.C. DeAngelis, Neural correlates of multisensory cue integration in macaque MSTd, Nat. Neurosci. 11 (10) (2008) 1201–1210.
[13] Y. Gu, et al., Visual and nonvisual contributions to three-dimensional heading selectivity in the medial superior temporal area, J. Neurosci. 26 (1) (2006) 73–85.
[14] T. Zhang, K.H. Britten, The responses of VIP neurons are sufficiently sensitive to support heading judgments, J. Neurophysiol. 103 (4) (2010) 1865–1873.
[15] A. Chen, G.C. DeAngelis, D.E. Angelaki, Representation of vestibular and visual cues to self-motion in ventral intraparietal cortex, J. Neurosci. 31 (33) (2011) 12036–12052.
[16] A. Chen, G.C. DeAngelis, D.E. Angelaki, Functional specializations of the ventral intraparietal area for multisensory heading discrimination, J. Neurosci. 33 (8) (2013) 3567–3581.
[17] K.C. Anderson, R.M. Siegel, Optic flow selectivity in the anterior superior temporal polysensory area, STPa, of the behaving monkey, J. Neurosci. 19 (7) (1999) 2681–2692.
[18] B. Zhao, Y. Zhang, A. Chen, Encoding of vestibular and optic flow cues to self-motion in the posterior superior temporal polysensory area, J. Physiol. 599 (16) (2021) 3937–3954.
[19] R.H. Fan, et al., Heading tuning in macaque area V6, J. Neurosci. 35 (50) (2015) 16303–16314.
[20] V. Cardin, L. Hemsworth, A.T. Smith, Adaptation to heading direction dissociates the roles of human MST and V6 in the processing of optic flow, J. Neurophysiol. 108 (3) (2012) 794–801.
[21] V. Cardin, A.T. Smith, Sensitivity of human visual cortical area V6 to stereoscopic depth gradients associated with self-motion, J. Neurophysiol. 106 (3) (2011) 1240–1249.
[22] R.M. Siegel, H.L. Read, Analysis of optic flow in the monkey parietal area 7a, Cerebr. Cortex 7 (4) (1997) 327–346.
[23] T. Zhang, H.W. Heuer, K.H. Britten, Parietal area VIP neuronal responses to heading stimuli are encoded in head-centered coordinates, Neuron 42 (6) (2004) 993–1001.
[24] D. Ostwald, et al., Neural coding of global form in the human visual cortex, J. Neurophysiol. 99 (5) (2008) 2456–2469.
[25] J. Ross, D.R. Badcock, A. Hayes, Coherent global motion in the absence of coherent velocity signals, Curr. Biol. 10 (11) (2000) 679–682.
[26] B. Zhao, et al., The computational rules of cross-modality suppression in the visual posterior sylvian area, iScience 26 (6) (2023) 106973.
[27] F. Zeng, A. Zaidel, A. Chen, Contrary neuronal recalibration in different multisensory cortical areas, Elife 12 (2023).

[28] A. Chen, G.C. DeAngelis, D.E. Angelaki, A comparison of vestibular spatiotemporal tuning in macaque parietoinsular vestibular cortex, ventral intraparietal area, and medial superior temporal area, J. Neurosci. 31 (8) (2011) 3082–3094.
[29] Y. Gu, et al., Decoding of MSTd population activity accounts for variations in the precision of heading perception, Neuron 66 (4) (2010) 596–609.
[30] A.S. Ecker, et al., The effect of noise correlations in populations of diversely tuned neurons, J. Neurosci. 31 (40) (2011) 14272–14283.
[31] R. Moreno-Bote, et al., Information-limiting correlations, Nat. Neurosci. 17 (10) (2014) 1410–1417.
[32] J.H. Maunsell, W.T. Newsome, Visual processing in monkey extrastriate cortex, Annu. Rev. Neurosci. 10 (1987) 363–401.
[33] L. Ungerleider, T. Pasternak, Ventral and dorsal cortical processing streams, The visual neurosciences 1 (34) (2004) 541–562.
[34] S.G. Kuai, et al., Integration of motion and form cues for the perception of self-motion in the human brain, J. Neurosci. 40 (5) (2020) 1120–1132.
[35] D.H. Hubel, T.N. Wiesel, Receptive fields and functional architecture of monkey striate cortex, J. Physiol. 195 (1) (1968) 215–243.
[36] J.B. Levitt, D.C. Kiper, J.A. Movshon, Receptive fields and functional architecture of macaque V2, J. Neurophysiol. 71 (6) (1994) 2517–2542.
[37] K.R. Gegenfurtner, D.C. Kiper, J.B. Levitt, Functional properties of neurons in macaque area V3, J. Neurophysiol. 77 (4) (1997) 1906–1923.
[38] J. Hu, et al., Visual motion processing in macaque V2, Cell Rep. 25 (1) (2018) 157–167 e5.
[39] A. Pavan, et al., The neural basis of form and form-motion integration from static and dynamic translational Glass patterns: a rTMS investigation, Neuroimage 157 (2017) 555–560.
[40] A. Pavan, et al., The interaction between orientation and motion signals in moving oriented Glass patterns, Vis. Neurosci. 34 (2017) E010.
[41] L. Liu, et al., Perceptual integration rapidly activates dorsal visual pathway to guide local processing in early visual areas, PLoS Biol. 15 (11) (2017) e2003646.
[42] D.J. Felleman, D.C. Van Essen, Distributed hierarchical processing in the primate cerebral cortex, Cerebr. Cortex 1 (1) (1991) 1–47.
[43] J.H. Maunsell, D.C. van Essen, The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey, J. Neurosci. 3 (12) (1983) 2563–2586.
[44] L.G. Ungerleider, R. Desimone, Cortical connections of visual area MT in the macaque, J. Comp. Neurol. 248 (2) (1986) 190–222.
[45] L.G. Ungerleider, et al., Cortical connections of area V4 in the macaque, Cerebr. Cortex 18 (3) (2008) 477–499.
[46] K. Grill-Spector, R. Henson, A. Martin, Repetition and the brain: neural models of stimulus-specific effects, Trends Cognit. Sci. 10 (1) (2006) 14–23.
[47] J.F. Nankoo, et al., Temporal summation of global form signals in dynamic Glass patterns, Vis. Res. 107 (2015) 30–35.
[48] C. Bruce, R. Desimone, C.G. Gross, Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque, J. Neurophysiol. 46 (2) (1981) 369–384.
[49] M.W. Oram, D.I. Perrett, Integration of form and motion in the anterior superior temporal polysensory area (STPa) of the macaque monkey, J. Neurophysiol. 76 (1) (1996) 109–129.