Check for updates

A Dynamic Efficient Sensory Encoding Approach to Adaptive Tuning in Neural Models of Optic Flow Processing

Scott T. Steinmetz[1]*, Oliver W. Layton[2], Nathaniel V. Powell[1] and Brett R. Fajen[1]

[1] Cognitive Science Department, Rensselaer Polytechnic Institute, Troy, NY, United States, [2] Computer Science Department, Colby College, Waterville, ME, United States

This paper introduces a self-tuning mechanism for capturing rapid adaptation to changing visual stimuli by a population of neurons. Building upon the principles of efficient sensory encoding, we show how neural tuning curve parameters can be continually updated to optimally encode a time-varying distribution of recently detected stimulus values. We implemented this mechanism in a neural model that produces human-like estimates of self-motion direction (i.e., heading) based on optic flow. The parameters of speed-sensitive units were dynamically tuned in accordance with efficient sensory encoding such that the network remained sensitive as the distribution of optic flow speeds varied. In two simulation experiments, we found that model performance with dynamic tuning yielded more accurate, shorter latency heading estimates compared to the model with static tuning. We conclude that dynamic efficient sensory encoding offers a plausible approach for capturing adaptation to varying visual environments in biological visual systems and neural models alike.

Keywords: efficient coding, adaptation, neural modeling, optic flow, visual processing

INTRODUCTION

Biological visual systems are remarkably adaptive to a wide range of visual environments and rapidly varying conditions, enabling animals to maintain perceptual contact with their surroundings regardless of whether they are familiar or novel, static or fluctuating. Although biological agents often encounter dramatic shifts in stimulus values as they move from one context to another, they rarely experience meaningful degradation in perception or performance. Understanding the principles and mechanisms that allow for such robustness is critical for both advancing our theoretical understanding of the visual system and informing the development of computational neural models of biological vision. The aim of this study is to explore and evaluate the idea that adaptation can be modeled as continuous changes in the mapping of neuronal inputs to outputs. We tie this mapping to changes in the distribution of a stimulus variable in accordance with the principles of efficient sensory encoding (Ganguli and Simoncelli, 2014), extending an

approach that was originally envisioned as a "set-it-and-forget-it" method for static parameter tuning to scenarios that demand rapid adaptation in real-time.

Visual adaptation encompasses a broad set of phenomena that extend across multiple timescales (Wainwright, 1999): the visual system is broadly shaped by its environment via evolutionary processes (Attneave, 1954; Barlow, 1961), it is attuned to local natural scene statistics through experience (Simoncelli and Olshausen, 2001; Simoncelli, 2003), and it rapidly adapts to current stimuli (Werblin, 1974; Webster, 2015). Adaptation has been studied extensively at both the perceptual and neural levels. The perceptual consequences of such adaptation include aftereffects, which are perceptual vestiges of adaptation to recently presented stimuli, as well as decreases in discrimination thresholds for recently experienced stimuli. At the neural level, adaptation is reflected in shifts in the firing rates of cells in response to prolonged exposure to a constant stimulus. Such effects have been observed across a variety of stimulus properties, including motion (Addams, 1834; Bartlett et al., 2018), light intensity, and orientation (Carandini et al., 1998; Müller et al., 1999; Dragoi et al., 2000; Gutnisky and Dragoi, 2008), and in both human and non-human animals (Shapley and Enroth-Cugell, 1984; Howard et al., 1987; Laughlin, 1989).

Adaptation in the visual system can be understood as a consequence of optimal information transmission, which captures how neural connections adjust to more efficiently encode relevant sensory variables (Laughlin, 1989; Barlow, 1990; Wainwright, 1999; Brenner et al., 2000). One highly influential theory that ties these concepts together is the efficient coding hypothesis (Attneave, 1954; Barlow, 1961), which is rooted in the idea that the information available to early sensory neurons is highly redundant and that such neurons distill from the flood of information the most relevant perceptual properties (Olshausen and Field, 1997). As the agent moves and its surroundings change, the statistics of naturally occurring stimuli undergo variations, which in turn require neural systems to adapt in order to maintain efficiency. Indeed, a static mapping from input to output is necessarily less efficient at extracting and encoding a signal than a dynamic mapping that adapts to new conditions.

Such adaptation to changing stimuli has long been predicted to have perceptual consequences, as sensitivity to more relevant information increases at the cost of lower sensitivity to less relevant information (Laughlin, 1989). Neurophysiological studies have found that post-adaption changes in neural activity that reflect coding efficiency are correlated with perceptual adaptation, for example in macaques viewing variously oriented gratings (Gutnisky and Dragoi, 2008). This has also been borne out in behavioral studies where humans have exhibited improved discrimination to recently encountered optic flow and degraded discrimination to older flow patterns (Durgin and Gigone, 2007). It is in this sense that some forms of adaptation can be understood as a consequence of coding efficiency.

The Present Study

The goal of this study is to explore a principled mechanism by which neural models may automatically regulate sensitivity when processing sensory information under a wide range of conditions. Our solution extends efficient sensory encoding (Ganguli and Simoncelli, 2014) which is itself built upon the efficient coding hypothesis (Attneave, 1954; Barlow, 1961). Efficient sensory encoding offers a principled approach to defining the parameters of tuning curves for N neurons such that they optimally encode a given distribution of stimulus values. In Ganguli and Simoncelli's formulation, this distribution was based on the values of stimulus variables found in the agent's environment. Their approach provides a mathematically precise means of codifying and optimizing tuning curve selection such that the neural population produces activity and discrimination thresholds similar to those seen in previous studies. However, it also assumes a static distribution of stimulus variables, which yields a static neural tuning.

We introduce dynamic efficient sensory encoding (DESE) as a mechanism for adaptive neural tuning to rapidly shifting distributions of stimulus variables. DESE extends the approach to tuning curve selection introduced by Ganguli and Simoncelli (2014) to define attunement based on recently detected sensory information, dynamically adapting with changes in the distribution of stimulus values. This provides a mathematical basis for continually and automatically adjusting the tuning curves of early visual neurons to optimize the encoding of target variables from stimuli the agent recently encountered, instead of the statistical average for the environment. Such patterns of adaptation of early sensory neurons have been observed in the H1 neuron in flies (Brenner et al., 2000). DESE could be especially useful to modelers who build computational neural models of vision and aspire to efficiently simulate the behavior of a large number of units with finite computing resources. Dynamic attunement also has the advantage of allowing for adaptation to the current sensory environment without modeler intervention, which is important both for avoiding overfitting and for practical applications where manual parameter selection is both unprincipled and time consuming.

We explore this solution in the context of an existing neural model, which provides us with a concrete example and allows for performance comparison of DESE against previous static approaches. Specifically, we discuss the dynamic tuning of speed-sensitive cells in the middle temporal (MT) area of the competitive dynamics (CD) model of optic flow processing in the primate dorsal stream (Layton and Fajen, 2016a,c). In the next few sections, we briefly introduce the reader to the topic of heading perception, summarize the CD model, and explain how we implemented dynamic efficient sensory encoding within the CD model. We then report the results of a set of simulation experiments that explore the improvement in heading estimation when dynamically attuning to the stimulus distribution.

Optic Flow and the Perception of Heading

Humans routinely navigate complex and dynamic environments without colliding with the other inhabitants and structures. By most accounts, the ability to successfully avoid obstacles and reach goals relies on sensitivity to one's direction of self-motion or heading (Li and Cheng, 2011). Although multiple sensory

modalities contribute to the perception of self-motion (Greenlee et al., 2016; Cullen, 2019), the ability to perceive where one is headed relative to potential goals and obstacles is predominantly driven by vision. In particular, heading perception is based on the structured patterns of optical motion that are induced by self-motion and known as optic flow (Gibson, 1950; Warren et al., 2001; Browning and Raudies, 2015). When the agent is traveling along a linear path through a static environment, the global optical motion radiates from a single point called the focus of expansion (FoE) that specifies the direction of self-motion. Humans can perceive their heading direction from optic flow with an accuracy of 1–2° of horizontal visual angle (Warren et al., 1988). Neurons that respond to global optic flow patterns and exhibit heading tuning have been found in various regions of the primate visual system, including MSTd (Duffy and Wurtz, 1991; Gu et al., 2006) and VIP (Britten, 2008; Chen et al., 2011).

Competitive Dynamics Model

Research on the neural mechanisms involved in the processing of optic flow in the primate visual system has informed the development of computational neural models of heading perception (Lappe and Rauschecker, 1994; Beintema and Van den Berg, 1998; Royden, 2002; Browning et al., 2009; Perrone, 2012; Beyeler et al., 2016), including the CD model (Layton and Fajen, 2016b) used in the present study. The CD model comprises stages that correspond to areas along the dorsal stream of the primate brain (e.g., V1, MT, MST; see **Figure 1**). Activity in model area V1 encodes optic flow estimated from video stimuli. Model area MT consists of units tuned to specific ranges of optic flow speeds and directions. At each region of the visual field, there is an identical population of MT units (a "macrocolumn") with different speed and directional sensitivities but the same receptive field location. The activity of the units in a macrocolumn encodes the optic flow at that location. Model area MSTd consists of "template cells" sensitive to large radially expanding patterns of optic flow associated with forward translation toward the center of the pattern, the focus of expansion (FoE). At each position in a 64 × 64 grid that spanned the visual field, there is an MSTd radial cell whose FoE is located at that position in the grid. Together, these units possessed sensitivity to forward heading directions throughout the visual field with a resolution of ~1.40° of visual angle (assuming a 90° × 90° field of view). We defined the heading estimate of the CD model as the heading preference of the maximally active MSTd cell. Previous work has shown this decoding strategy is sufficient for capturing human heading judgements (Royden, 2002; Browning et al., 2009; Layton et al., 2012).

Estimates of heading direction generated by the model are consistent with those of humans across a variety of scenarios, including in the presence of locally and globally discrepant optic flow resulting from factors such as independently moving objects and blowing snow, respectively (Layton and Fajen, 2016b,c, 2017; Steinmetz et al., 2019). Such robustness is largely due to mechanisms such as spatial pooling and recurrent feedback that allow the model to make use of the temporally evolving global

flow field rather than relying on an instantaneous snapshot of optic flow as prior models did (Layton and Fajen, 2016b).
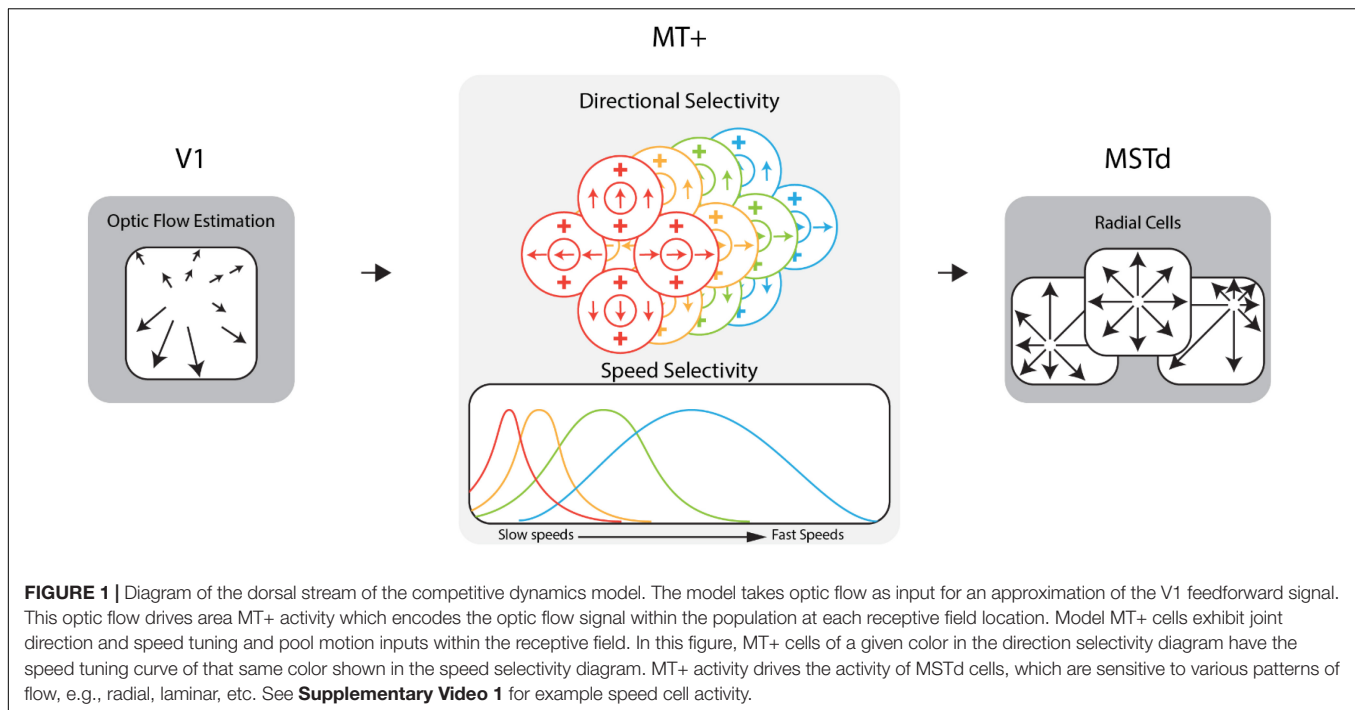
Speed Selectivity in the Competitive Dynamics Model

In the CD model, as in the primate visual system, the local optical motion in any small region of the visual field is encoded in terms of the activity of MT neurons. Each individual unit exhibits tuning to a range of speeds and directions of motion. The model population consists of units tuned to all combinations of 24 preferred directions and seven preferred speeds. To study the effects of adaptive tuning, we used DESE to automatically regulate the parameters of MT speed tuning curves. We kept each neuron's direction tuning curve constant to focus on the effect of DESE within larger CD model system from a single set of tuning curves.

Our primary goal was to compare and evaluate this mechanism with other methods for characterizing MT speed tuning. Taking a naïve approach, we defined each model MT cell to have a Gaussian-shaped tuning curve with two governing parameters: the mean (μ) which corresponds to the optic flow speed that maximally excites the unit, and the standard deviation (σ) which defines the range of speeds to which the unit responds. As such, the response of any given MT cell depends on how closely the optic flow speed matches the cell's preferred speed, dropping to zero for speeds above and below the cell's range of sensitivity. While the tuning curves satisfied a Gaussian function for simplicity in the present study, the CD model could accommodate others, such as log-Gaussian or gamma distributions (Nover et al., 2005). Because with DESE the tuning curve parameters adaptively fit the sensory information over time, we would expect minimal potential benefit in selecting one of these other tuning curve functions, although this could be systematically tested in a future study.

The optical speed of an object or surface in the world is a function of the relative distance, direction, and motion between the agent and object (Longuet-Higgins and Prazdny, 1980). As such, during naturalistic self-motion, changes in the motion of the agent and the structure of the environment cause the distribution of optic flow speeds that one encounters to vary across a wide range. Although no single MT cell in the CD model is sensitive to the entire range of optic flow speeds, variations in the neuronal sensitivities allow for the population as a whole to be responsive to a wider range. Ideally, the number of unique speeds to which MT neurons are tuned would be sufficiently large to fully sample with fine-grained precision the entire distribution of optic flow speeds that will be encountered in the world. Simply adding more units, however, introduces problems for both biological vision systems and neural models, which must be efficient in the use of resources. In the CD model, for example, one model MT macrocolumn consists of 168 units (24 preferred directions, seven preferred speeds). Increasing the number of macrocolumns that sample the visual field comes at a substantial computational cost, as does adding another preferred speed to each macrocolumn.

The model's overall sensitivity to motion depends not only on the number of unique speed preferences, but also on how well

FIGURE 1 | Diagram of the dorsal stream of the competitive dynamics model. The model takes optic flow as input for an approximation of the V1 feedforward signal. This optic flow drives area MT+ activity which encodes the optic flow signal within the population at each receptive field location. Model MT+ cells exhibit joint direction and speed tuning and pool motion inputs within the receptive field. In this figure, MT+ cells of a given color in the direction selectivity diagram have the speed tuning curve of that same color shown in the speed selectivity diagram. MT+ activity drives the activity of MSTd cells, which are sensitive to various patterns of flow, e.g., radial, laminar, etc. See **Supplementary Video 1** for example speed cell activity.

speed tuning curve parameters match to the input distribution of optic flow speeds. Because each neuron is only sensitive to motion within a limited range (see **Supplementary Video 1**), selecting model tuning curve parameters that do not capture the range of speeds present in the optic flow input would result in regions of poor or absent sensitivity. This is a critical point because a poorly tuned model will haphazardly use only a portion of the available optic flow, potentially leading to sub-optimal performance. Conversely, when the tuning curves are distributed along the stimulus space according to the likelihood of the stimuli, the neurons encode the signal more efficiently (Laughlin, 1981; Ganguli and Simoncelli, 2014).

A naïve hypothesis might hold that tuning curves should uniformly tile the range of stimuli seen, as depicted in **Figure 2A**. However, speeds often do not arise with uniform probability in the optic flow (**Figure 2B**), and a uniform tiling results in equal precision between frequently observed stimuli and infrequently observed stimuli. A more efficient use of resources would be to recruit more neurons for signaling common stimulus values so that such stimuli can be encoded more precisely, and fewer neurons to signaling less common stimulus values (**Figure 2D**).

Efficient Sensory Encoding

Efficient sensory encoding (Ganguli and Simoncelli, 2014) offers a principled approach rooted in the efficient coding hypothesis for calculating the peak locations and widths of the optimal tuning curve for a given stimulus. More precisely, sensory variables are encoded to optimize the Fischer information across the neural population given a stimulus probability distribution. The derived tuning curves proportionately sample the stimulus probability distribution, resulting in narrow and densely packed tuning curves where the probability of that

stimulus is high (e.g., where instances of that optic flow speed are very common) and broad, sparse tuning curves where the probability of that stimulus is low. This results in each speed cell capturing an approximately equal probability mass of the sensory distribution.

The following procedure describes how the distribution of the stimulus values in the environment was used to calculate the parameters (μ and σ) of the Gaussian tuning curves. The density function, d(s), for a given stimulus magnitude, s, with N neurons in the speed cell population is defined as

$$d(s) = N \ p(s) \tag{1}$$

where p(s) is the probability distribution of the stimuli (see **Figure 2B** for an example). Note that multiplying by N causes the integral of d(s) to equal N. The peak location of the nth speed cell's tuning curve (μ_n) is determined by evaluating where the cumulative distribution function (D) of the density function equals n:

$$\mu_n = s_n \text{ where } D(s_n) = n \tag{2}$$

This process is illustrated in **Figures 2C,D** for $N = 7$ (the number of unique speed preferences in the model used in the present study). The full width at half maximum (FWHM) of the nth tuning curve is equal to the inverse of the density function evaluated at s_n:

$$FWHM_n = \frac{1}{d(s_n)} \tag{3}$$

The FWHM is then used to calculate σ_n as follows (for gaussian tuning curves):

$$\sigma_n = \frac{FWHM_n}{2.355} = \frac{1}{2.355 \cdot d(s_n)} \tag{4}$$
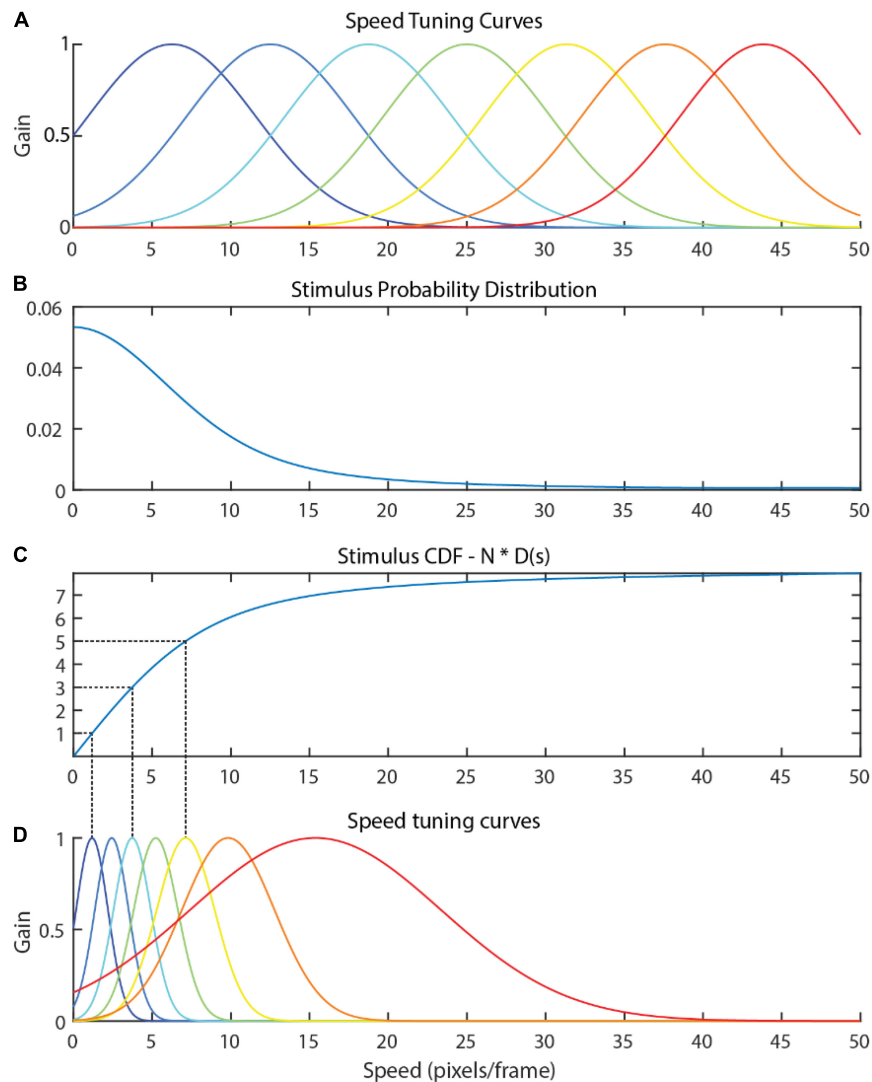
FIGURE 2 | (A) A naïve attunement in which speed tuning curves uniformly tile the stimulus range 0–50. Naturalistic flow is rarely uniform, and the distribution of optic flow speeds present in the input optic flow within the video dataset used in Experiments 1 and 2 of the present study **(B)** is clearly non-uniform. The uniform tiling strategy leaves much of the encoding capacity of the uniform tuning wasted on infrequently observed speeds. The tuning curves derived via efficient sensory encoding **(D)** allocate neural sensitivities based on the frequency that each speed arises in the stimulus, calculated via the **(C)** CDF of the probability distribution multiplied by the number of encoding neurons, N. In this way, encoding capacity (and computational cost) is optimally allocated for encoding this distribution.

For an in-depth derivation of these equations, see Ganguli and Simoncelli (2014).

Dynamic Efficient Sensory Encoding of Optic Flow

Efficient sensory encoding defines the ideal attunement for a given stimulus probability distribution but implicitly assumes that this attunement is stationary, which can be inefficient for many dynamic environments. The statistics of optic flow speeds vary not only across the visual field but also over time as the agent moves relative to objects and surfaces in the world (Calow and Lappe, 2007). Self-motion above a flat ground surface produces a smooth gradient of speeds proportional to the relative depth. By

contrast, self-motion toward a flat wall produces in narrow range of optic flow speeds. Movement through realistic environments often produces a combination of constant speeds and gradients in the optic flow field. This is depicted in **Supplementary Video 2**, which shows the optic flow generated by self-motion through a simulated environment (**Supplementary Video 2A**) alongside a histogram that shows the distribution of optic flow speeds pooled across the entire image (**Supplementary Video 2B**). As the agent moves, changes in the direction and relative depth of surfaces as well as in self-motion speed and direction result in dramatic variations in the shape of the distribution over time.

In this study, we explore how the attunement process described in the previous section can be implemented on rapid timescales (<1 s). The model records and stores the distribution

of optic flow magnitudes across the entire visual field over the past n frames. This defines the "rolling time horizon" of the dynamic attunement, within which all flow is considered equally. The stored distribution of optic flow is then fed into the efficient sensory encoding equations defined in the previous section, and the speed cell tuning curves are adjusted according to the resulting values (**Supplementary Video 2C**).

MATERIALS AND METHODS

We conducted two simulation experiments to systematically test the potential benefits of dynamic efficient sensory encoding and to determine the effects of different time horizons. In Experiment 1, we compared dynamic tuning based on the evolving distribution of optic flow speeds observed over the last 10 frames to two static tunings, which served as control conditions. The first static tuning was based on a uniform distribution of optic flow speeds with a range of 0–5.3° per frame. This range was chosen to span the most commonly observed optic flow reasonably (i.e., the body of the aggregate probability distribution which contained 96.37% of the observed optic flow). The second static tuning was based on the distribution of optic flow speeds aggregated across all frames of the entire test set of videos (described in the next paragraph). We refer to these three conditions as the Dynamic, Static—Uniform, and Static—Aggregate conditions, respectively. Experiment 2 investigated how performance with dynamic tuning depends on the number of previous frames across which speeds were aggregated to generate the optic flow distribution (i.e., the time horizon). Specifically, we compared the time horizon used in Experiment 1 (10 frames) with shorter (one frame) and longer time horizons (30 frames).

We tested each tuning method on a set of 60 videos depicting self-motion through a detailed virtual environment. Examples of the videos can be seen in **Supplementary Video 3**. The videos were generated using Microsoft AirSim in the Unreal game engine. We selected two distinct environments to provide a broad sample of natural and human-made scene structures: an outdoor residential neighborhood (**Figure 3A**) and the inside of a warehouse (**Figure 3B**). The former included houses, parked cars, trees, and telephone poles with powerlines situated in residential zones and parks, various ground-surface textures (e.g., pavement, concrete, grass), and a blue sky with distant clouds. Most vantage points exhibited naturalistic depth variation across the visual field, from nearby structures to distant trees at the horizon. The warehouse environment was an enclosed space comprising corridors lined by shelves with boxes and other items. In general, surfaces were located at closer distances to the camera and depth varied over a smaller range compared to the neighborhood environment.

At the start of each video, the camera was positioned in the simulated environment at a randomly selected height between 1 and 5 m. The camera then traveled along five connected linear segments each lasting 30 frames. The heading direction for each segment was determined by selecting a pixel within the camera image of the scene at random toward which to move (see

Figure 3C). Camera speed was fixed within each segment and varied randomly across segments between 1 and 20 m/s. Between segments, the camera rotated to the new heading directions and accelerated or decelerated to the new speed over three frames. As a result, each video lasted 162 frames (or equivalently, 5.4 s at 30 fps). The field of view was 90° H × 90° V. Videos were manually reviewed before inclusion in the dataset to ensure that there were no collisions with objects or the ground.

For both experiments, the resolution of the videos was 512 × 512 pixels and optic flow was estimated using the Farneback method (Farnebäck, 2003). This method examines the sequence of images at several different resolutions, using salient motion at one resolution as a starting point for tracking motion in another. Unlike sparse optic flow estimation procedures like (Lucas and Kanade, 1981), Farneback estimation produces dense optic flow estimates. This estimation contains some noise, particularly in areas of low visual contrast like the virtual sky, meaning that the model must be robust to small errors. Model MT macrocolumns processed this signal in a 256 × 256 non-overlapping grid akin to an image resolution. The activity of each MT neuron was the result of convolving a gaussian kernel over the speed cell activity. This kernel had a sigma of 1.0, resulting in each model neuron having a receptive field spanning a circle ˜9 pixels across, ˜3.15° in diameter. In the present study, area MT had seven neurons with unique speed preferences at each region of the visual field; that is, each MT unit could take on one of seven unique parameter value pairings (μ_i, σ_i, i = 1, ..., 7). Model area MSTd consisted of a 64 × 64 grid of units tuned to headings evenly spaced across the visual field, with each neuron possessing a receptive field that encompassed the entire image. We found that model performance was generally robust with this grid size and did not benefit from higher resolutions. The spatial resolution of each area balances model runtime against precision of the heading estimate, however, MSTd has a greater effect, since each unit therein involves a dot product between the radial motion template and a large number of feedforward signals from MT. Thus, increasing MSTd resolution rapidly increases the number of calculations necessary. These parameters were selected after preliminary pilot testing based on what produced the optimal heading estimate. It is worth noting that these values deviate from those observed in neurophysiological studies of primates, where receptive fields vary in size according to eccentricity and MT neurons have receptive fields 4–25° across (Felleman and Kaas, 1984).

RESULTS

Experiment 1: Comparing Dynamic and Static Tuning

Figure 4A shows the mean absolute heading error per frame averaged across videos for the Dynamic, Static—Uniform, and Static—Aggregate conditions. Absolute heading error was measured by calculating the angle in the 2D image plane between the model estimate and the ground-truth heading direction on each frame (see **Supplementary Video 4**). The four sudden transitions in the heading error time series shown in **Figure 4A**

FIGURE 3 | Example screenshots from the **(A)** Neighborhood and **(B)** Warehouse environments. For video examples, **see Supplementary Video 3**. The process of translational segment selection is shown in **(C)** with heading indicated by a black square. After an initial heading direction is randomly selected, the camera translates in that direction for 30 frames. Subsequently, a second random heading is selected, and three interim headings are linearly chosen between the first and second heading. After the three-frame transition, the camera translates toward the second heading for 30 frames. This process is then repeated for segments 3, 4, and 5 in each video.

are a consequence of the abrupt shifts in heading direction between segments. Segments 2 through 5 were preceded by three transition frames, during which heading was rapidly changing from the previous direction to the new one. Heading error surged during the transition period, then gradually dropped due to temporal dynamics within the CD model. Specifically, the activity of each MT+ and MSTd unit in the CD model depends not only on feedforward input at that instant but also on the activation of that unit and its neighbors on the previous time step. As such, the model's heading estimate on each frame is not based entirely on a snapshot of the optic flow field at that instant. Rather, heading estimates depend on the temporal evolution of the flow field, enabling smooth, non-instantaneous changes in estimates from the previous to current heading even in the face of abrupt changes in optic flow input (Layton and Fajen, 2016b). To be clear, model activity was not time-averaged over any temporal window, but rather activity was dependent on neural dynamics in the recent time history. The first segment of each video is unique in that there is no existing neural activity from previous stimuli for activity due to the new heading to overwhelm, so the heading estimate quickly stabilizes. For this reason, the first segment was excluded from the following analyses.

We compared the Dynamic, Static—Uniform, and Static—Aggregate tunings on two measures of heading estimation, both of which were extracted from the time series of absolute heading

error. The first measure was the converged heading error, which was based on the heading error on the final frame of segments 2–5. Dynamic attunement was more accurate than both static conditions resulting in a mean converged heading error of 4.24° [95% CI (3.83, 4.66), $n = 240$] which was significantly lower compared to both the Static—Uniform [$M = 9.17°$, 95% CI (7.96, 10.39), $n = 240$] and Static—Aggregate [$M = 5.60°$, 95% CI (4.88, 6.32), $n = 240$] conditions (**Figure 4B**). A one-way repeated-measures ANOVA with degrees of freedom corrected for violation of the sphericity assumption (Greenhouse-Geisser $\varepsilon = 0.77$) revealed a main effect of Tuning Condition [$F_{(1.54, 367.01)} = 42.70$, $p < 0.001$] and all pairwise comparisons with Bonferroni corrections were significant ($p < 0.001$). Heading error was also more consistent across segments and trials in the Dynamic condition. The difference between the three conditions was even more pronounced earlier in segment (i.e., on frame 15; see squares in **Figure 4B**).

Although the accuracy of model heading estimates relative to human estimates is not directly relevant to the focus of this study, it is worth noting that with dynamic attunement, model estimates are only slightly less accurate than those of human subjects under similar self-motion conditions. During linear translation through a static environment, human heading judgments are accurate to within 1–2° (Warren et al., 1988; Foulkes et al., 2013), which is ∼2–3° more accurate than model estimates. However,
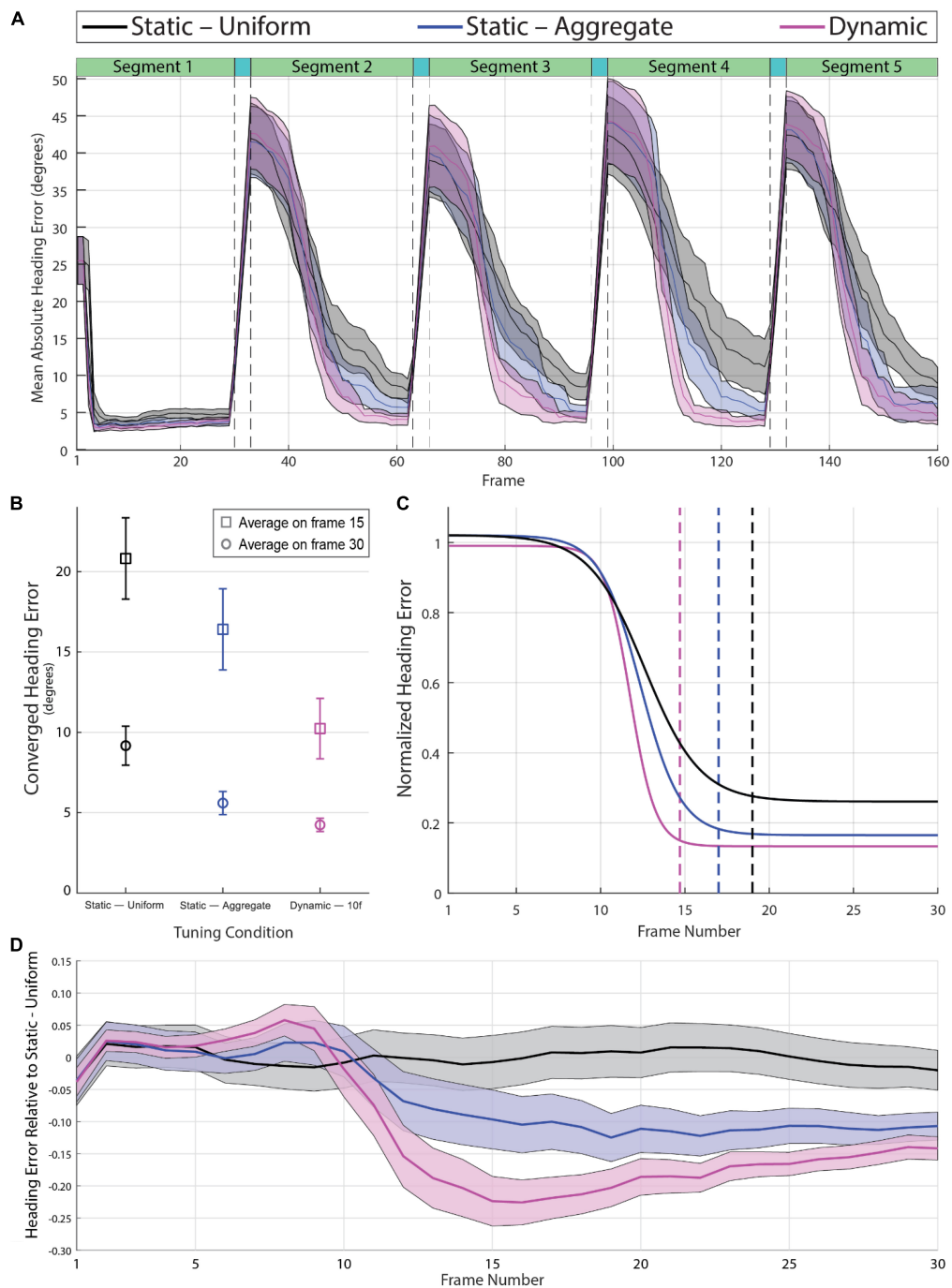
FIGURE 4 | (A) Mean absolute heading error across all 60 videos in the Static-Uniform (black), Static-Aggregate (blue), and Dynamic (magenta) conditions. The shaded regions represent the 95% confidence interval of the mean heading error in each frame. The dashed vertical lines indicate the beginning and end of the three-frame transition to a new heading prior to the 2nd through 5th segments. **(B)** Mean heading error at frames 15 and 30 over segments 2–5 ($N = 240$) with 95% confidence intervals. **(C)** Best-fitting logistic curves for the heading error in segments 2–5 of each condition ($N = 240$), normalized to the magnitude of the heading shift for that segment; i.e., heading error starts around one regardless of whether the initial heading differed from the previous heading by 1 or 80°. The vertical dashed lines indicate the frame at which each curve reaches 98% of its range, which provides a measure of each curve's convergence onto its final heading estimate. Note that the best-fitting curve for the dynamic attunement condition converges more quickly to a lower and therefore more accurate heading error than the static conditions. **(D)** Mean heading error per frame of segments 2–5 ($N = 240$) relative to the best-fitting curve of Static-Uniform condition. See **Supplementary Video 4** for example heading activity.
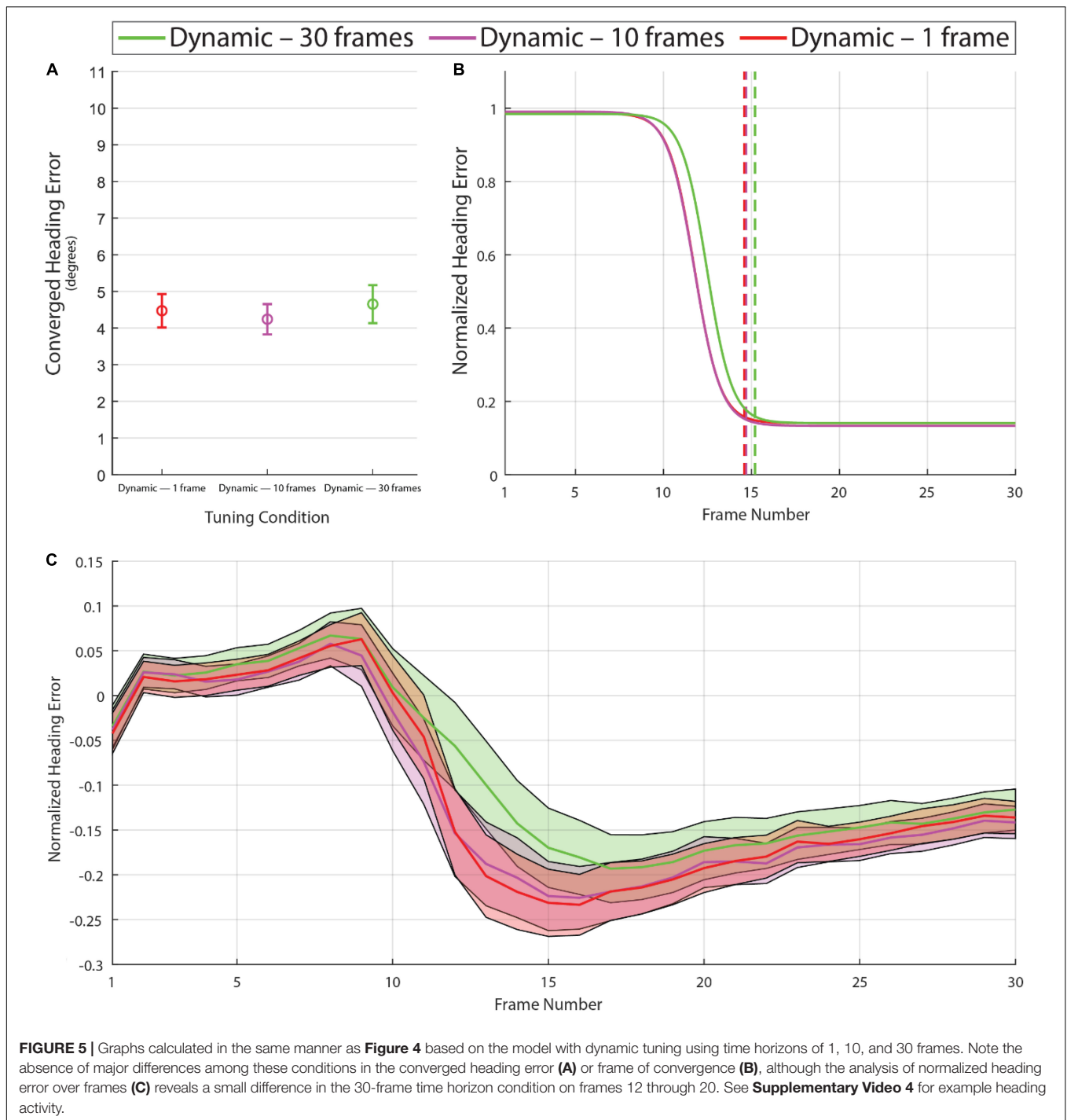
FIGURE 5 | Graphs calculated in the same manner as **Figure 4** based on the model with dynamic tuning using time horizons of 1, 10, and 30 frames. Note the absence of major differences among these conditions in the converged heading error **(A)** or frame of convergence **(B)**, although the analysis of normalized heading error over frames **(C)** reveals a small difference in the 30-frame time horizon condition on frames 12 through 20. See **Supplementary Video 4** for example heading activity.

in studies of human heading perception, the range of simulated self-motion directions is typically restricted to central heading along the horizontal (azimuthal) dimension. This contrasts with the conditions used in the present study, in which self-motion direction varied in both azimuth and elevation and the primary measure of heading accuracy was the angular difference between the estimated and actual heading in two dimensions. When heading error is decomposed into its two components, we find that with dynamic tuning, mean converged heading error at frame 30 was 2.44° [95% CI (2.13, 2.76)] along the azimuth and 2.99° [95% CI (2.64, 3.34)] along the elevation axis, which is within 1° of human-level performance. Furthermore, Farneback optic flow estimation is imperfect, introducing noise into the model's estimate. For human observers, adding noise to the vector directions in the optic flow field resulted in degraded heading accuracy (Warren et al., 1991).

The second measure of performance that we considered was latency, which we defined as the number of frames it took the model to converge to a stable heading estimate after each transition. We estimated latency by first normalizing the heading error time series for each individual segment (excluding the 1st segment) of each video such that the heading error on the first frame was 1. This allowed us to parse out the variance due to differences in the magnitude of the heading shift between segments, which varied widely from a few degrees of visual angle up to about 80 degrees. We then fit a logistic curve to each time series, then averaged the curve parameters to determine the average logistic curve for each condition. We defined the heading convergence as the number of frames that was needed for the best-fitting curve to drop by 98% of the difference between the upper and lower bounds of the logistic function. We found that heading error decreased in fewer frames with Dynamic tuning (14.7 frames) compared to both Static—Aggregate (17.0 frames), and Static – Uniform (19.1 frames) (see **Figure 4C**).

To better visualize the differences between the three tuning conditions, we calculated for each segment the difference between the actual heading error in each tuning condition and the best-fitting curve in the Static—Uniform condition. That is, we subtracted the curve that best fit the heading error time series (segments 2 through 5 of **Figure 4A**) in the Static—Uniform condition from the actual heading error time series on each trial across conditions. The curves in **Figure 4D** show the mean and 95% CI of this difference for each frame. This way of representing the data highlights how the improvement in the heading estimate due to dynamic tuning evolves over time while controlling for the size of the heading shift and other sources of variance. Heading error is similar across tuning conditions for the first 8–10 frames but then begins to decrease more rapidly in the Dynamic and Static—Aggregate conditions. By Frame 15, the mean heading estimate in the Dynamic condition is more accurate than the heading estimate in the Static—Uniform condition by about 22.4% of the heading shift magnitude, which is considerably better than the improvement with Static—Aggregate tuning (∼9.7% on Frame 15). Over frames, the difference between the three conditions shrinks, suggesting that the primary benefit of Dynamic tuning is that it allows the model to reach an accurate heading estimate in fewer frames.

Experiment 2: Comparing Dynamic Time Horizons

Experiment 1 demonstrated that with dynamic tuning based on the distribution of optic flow speeds over the past 10 frames, the model converges toward a more accurate heading estimate in less time. In Experiment 2, we examined whether the benefits of dynamic tuning depend on the number of previous frames used to build the optic flow distribution. Specifically, we compared model performance with 10 frames against performance with one frame and 30 frames. As shown in **Figure 5**, the manipulation of time horizon with the range that was tested had negligible effects on converged heading error [$F_{(1.72, 407.73)} = 1.80$, $p = 0.16$, with degrees of freedom corrected for violation of sphericity (Greenhouse-Geisser $\varepsilon = 0.86$)]. Some small differences were

observed in the latency of the heading estimate with a 30-frame time horizon (**Figures 5B,C**). On some frames in the single frame time horizon condition, the derived speed tuning curves collapsed to identical values as the observed flow was mostly one value, producing erratic activity in this limit case. These outlier issues aside, the model performed similarly across the three time horizon conditions that were tested.

DISCUSSION

Incorporating dynamic efficient sensory encoding into the competitive dynamics model improved the accuracy and responsiveness of the heading estimate, bringing the average error closer to human levels of performance on the realistic video stimuli. Despite these gains, one might wonder how plausible such a mechanism is in the human visual system, and how downstream areas of the visual system might handle tuning that continually fluctuates based on recently detected stimuli. Here we address both of these issues in turn.

Biological Plausibility

Although many forms of visual adaptation have been observed throughout the early visual system, there remains the question of whether speed cells in MT adapt to recently detected optic flow as well as whether that adaptation occurs on a timescale similar to DESE as implemented here. Let us first consider evidence that the tuning curves of many MT neurons are dynamic rather than static. Dynamic shifts in the directional selectivity of MT neurons have been demonstrated to solve the aperture problem: tuning shifts from the direction of motion perpendicular to a contour in their receptive field to encoding the true direction of motion over a 60 ms period of exposure (Pack and Born, 2001). There is also evidence that surround modulation of MT neurons is stimulus-dependent, with some surrounds becoming either excitatory or inhibitory depending on the stimulus (Huang et al., 2007). Furthermore, the encoding of direction of motion of multiple transparently moving stimuli shifts over time from an averaged response to encoding separate directions (Xiao and Huang, 2015). At the behavioral level, humans have exhibited improved discrimination of recently seen optic flow speeds produced by walking and standing still (Durgin and Gigone, 2007). Taken together, these findings demonstrate that adaptation to recently presented optic flow and other visual stimuli occurs at least to some degree in MT.

There is also evidence that adaptation in MT occurs over periods of time that correspond to the dynamic conditions tested in the present study. The highest performing condition from both experiments, the dynamic 10-frame time-horizon condition, simulates adaptation on a timescale that is dependent on the framerate of the input video. Assuming video at 30 frames per second, a 10-frame time horizon equates to 333 ms. Adaptation has been observed in the activity of macaque V1 orientation neurons after a 400 ms period (Gutnisky and Dragoi, 2008). Human adaptation to recently detected optic flow was produced after exposures to motion patterns for 1.25–1.5 s (Durgin and Gigone, 2007), comparable to the dynamic 30

frame time-horizon condition. These findings lend support to the existence of adaptive mechanisms that operate on timescales like those simulated in the present study.

Nevertheless, important questions remain about the biological implementation of DESE. Most notably, in the present study, MT tuning curve parameters were defined by the cumulative distribution function (CDF) of recently detected optic flow. This was intended to be an abstraction. We are not suggesting that MT has direct access to the distribution of optic flow or even that the distribution is explicitly represented in the visual system. Our goal in this study was merely to explore the population-level effects of adaptation seen in neurophysiological studies, which could prove useful to modelers focused on the effects of such adaptation on downstream areas. An important goal for future work is to develop mechanistic models of this process that capture how MT tuning curves adapt to recently detected optic flow.

Even if DESE does not simulate the exact, low-level mechanisms of adaptation in biological organisms, we contend that it captures the effects of adaptation in a functional sense. Adaptation serves to improve perceptual discrimination via efficient encoding of the stimulus distribution, which is precisely what DESE achieves. In the same way that the rate model of the neuron does not specify or attempt to capture the dynamics of neural activity at the lowest level (Brette, 2015), dynamic application of efficient sensory encoding provides a means of capturing the consequences of adaptive mechanisms on the feedforward signal without simulating the exact mechanism of adaptation. For example, there is some evidence that MT neurons in fact do not adapt their sensitivities to recently detected stimuli, but rather that later brain areas adjust their weighting of various MT neurons as they become more or less relevant to the current stimulus (Liu and Pack, 2017). If this were the case, DESE could be thought of as an approach to simulating in fewer computations the behavior of many statically tuned neurons. From this perspective, the tuning curves calculated under DESE are the subset of neurons from the much larger population that have the most relevant activity (at that point in time) for the property being encoded. These tuning curve selections are optimal with respect to the number of neurons being simulated, assuring the most efficient encoding given computational constraints or, alternatively, capturing the properties of the feedforward signal of a large population of neurons with a small number of simulated ones. Whether the mechanisms behind perceptual adaptation lie in MT or beyond, the resulting effects can be captured using dynamic efficient sensory encoding as an abstraction of adaptation.

Downstream Interpretability

One intriguing implication of dynamic tuning in MT is that the signal sent to the downstream model area (i.e., MSTd) reflects not only changes to the input to MT but also by shifts in the tuning curves. As such, a particular response from a given speed cell may result from motion at one speed at one point in time and a different speed at a later point in time. This raises the question of how the downstream area makes use of that speed cell's activity to generate an accurate heading estimate when that activity could have resulted from many different inputs.

We hasten to point out that the answer does not require giving the downstream area (MSTd) access to the adapted tuning parameters of MT cells. As argued by Hosoya et al. (2005), there is no need to communicate the current state of adaptation in one area to downstream areas. The brain does not construct a perfectly veridical reproduction of light on the retina but instead detects information that is useful and relevant to behavior. In the present study, the relevant perceptual variable is heading direction and the relevant visual information is carried in the directions rather than the magnitudes of optic flow vectors. For purely translational self-motion, heading is specified by the position of the FoE in the global radial flow pattern, which is invariant over changes in the speed of individual motion vectors. Indeed, human heading estimates during translational movement are unaffected by the addition of noise to flow vector magnitudes but significantly impaired by noise added to vector directions (Warren et al., 1991). This is adaptive because variations in flow vector speed could result from variations in depth or self-motion speed, even if self-motion direction is constant. In contrast, the direction of motion vectors depends on self-motion direction alone. Likewise, in this formulation of the CD model, the template match between the observed optic flow and each radial cell in MSTd is determined largely by the match in vector directions while differences in speed have a minimum impact. Just as the information about heading is largely invariant over flow-vector magnitude, so is the activity of individual radial cells in MSTd. Taken together, although dynamic tuning affects the encoding of optic flow speed in MT, the signals that carry the information relevant to heading estimation are left intact and the impact on MSTd cell activity is minimal.

This explains why dynamic tuning does not degrade heading estimation, but it does not explain why performance improves. The key insight is that dynamic tuning allows the model to be more sensitive to all the optic flow vectors present. By continually shifting tuning curves to detect a larger portion of the current optic flow, more of the available information about heading direction can be detected. As such, even if dynamic tuning-related changes distort the absolute speed signal, the overall improvement in sensitivity to task-relevant information more than compensates for the loss.

It is also worth noting that some information about flow vector speed is preserved in the signal to MSTd. In speed cells governed by DESE, each unique speed tuning curve captures an equal portion of the probability mass of the stimulus distribution and their ordinal relation remains constant. That is, the model unit that responds to the slowest speeds always encodes the slowest optic flow speeds, no matter what distribution is being used for attunement. As such, although the absolute magnitude of optic flow to which a speed cell is sensitive is continually shifting, each speed cell always represents a relatively constant percentage of the stimulus distribution. For example, the activity of the slowest speed cell might represent the presence of the slowest 10% of speeds in the receptive field of an active neuron. In this sense, DESE converts the signal from speed cells into relative units, where the absolute magnitude of the optic flow is lost but the magnitude relative to recently experienced optic flow is preserved. This could be useful for estimating properties other than heading that may rely on the magnitude of optic flow.

In our simulations, all units are either fully dynamic or static, but it is possible that neural populations may have a mix of dynamic and static units. In such a scenario, both the relative speeds and the absolute speeds would be encoded for use by various downstream populations that serve different purposes. In this work, we strove to elucidate the effects of dynamic efficient sensory encoding on the feedforward signal but do not claim that this is the mechanism used by all speed cells in primate MT. We found that DESE serves as a useful mechanism for simulating the effects of neural adaptation on the feedforward signal while accounting for computational constraints in accordance with the efficient coding hypothesis. From a modeling perspective, DESE represents a promising methodology for automatically regulating tuning curve parameters without modeler intervention.

CONCLUSION

Dynamic efficient sensory encoding offers a principled approach for capturing adaptation in early visual areas. By adjusting the tuning curves of simulated neurons based on the distribution of recently experienced stimuli, DESE produces the optimal encoding of that distribution by a finite number of neurons which results in improved sensitivity to task-relevant information. Implementing this adaptive mechanism in the speed cells of the competitive dynamics model led to measurable gains in performance, decreasing the error and shortening the latency of heading estimates derived from model area MSTd activity. Although some aspects of the approach are specific to the model and task used in the present study, dynamic efficient sensory encoding is more generic and could also be applied to the encoding of stimulus variables in other domains. Because it relies only on the distribution of recently experienced stimulus values, it does not assume prior knowledge of the statistics of the environment, thereby capturing how biological systems and neural models alike could rapidly adapt to unfamiliar environments.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: The raw video dataset and model activity are available on Open Science Framework: https://osf.io/suqxm/.

AUTHOR CONTRIBUTIONS

SS and NP contributed to collection of the video datasets. SS performed the statistical analysis with contributions from OL, NP, and BF and wrote the manuscript. All authors contributed to the conception and design of the study, manuscript revision, read, and approved the submitted version.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fncom.2022.844289/full#supplementary-material

For a playlist of the supplementary videos, please see: https://www.youtube.com/playlist?list=PLhTNB4zwGYlwCSZZzHsMou1KQE1hxg-dB.

REFERENCES

Addams, R. (1834). An account of a peculiar optical phenomenon seen after having looked at a moving body. *Lond. Edinb. Philos. Mag. J. Sci.* 5, 373–374.

Attneave, F. (1954). Some informational aspects of visual perception. *Psychol. Rev.* 61, doi: 183–193. 10.1037/h0054663

Barlow, H. B. (1961). "Possible principles underlying the transformations of sensory messages," in *Sensory Communication*, ed. W. A. Rosenblith (Cambridge, MA: The MIT Press), doi: 10.7551/mitpress/9780262518420.003.0013

Barlow, H. B. (1990). "A theory about the functional role and synaptic mechanism of visual aftereffects," in *Vision: Coding And Efficiency*, ed. C. Blakemore (Cambridge: Cambridge University Press).

Bartlett, L. K., Graf, E. W., and Adams, W. J. (2018). The effects of attention and adaptation duration on the motion aftereffect. *J. Exp. Psychol.* 44, 1805–1814. doi: 10.1037/xhp0000572

Beintema, J. A., and Van den Berg, A. V. (1998). Heading detection using motion templates and eye velocity gain fields. *Vis. Res.* 38, 2155–2179. doi: 10.1016/S0042-6989(97)00428-8

Beyeler, M., Dutt, N., and Krichmar, J. L. (2016). 3D visual response properties of MSTd emerge from an efficient, sparse population code. *J. Neurosci.* 36, 8399–8415. doi: 10.1523/JNEUROSCI.0396-16.2016

Brenner, N., Bialek, W., and De Ruyter Van Stevenick, R. (2000). Adaptive rescaling maximizes information transmission. *Neuron* 26, 695–702. doi: 10.1016/S0896-6273(00)81205-2

Brette, R. (2015). Philosophy of the spike: rate-based vs. Spike-based theories of the brain. *Front. Syst. Neurosci.* 9:151. doi: 10.3389/fnsys.2015.00151

Britten, K. H. (2008). Mechanisms of self-motion perception. *Annu. Rev. Neurosci.* 31, 389–410. doi: 10.1146/annurev.neuro.29.051605.112953

Browning, N. A., and Raudies, F. (2015). "Visual navigation in a cluttered world," in *Biologically Inspired Computer Vision*, (Weinheim: Wiley-VCH Verlag GmbH and Co. KGaA), 425–446. doi: 10.1002/9783527680863.ch18

Browning, N. A., Grossberg, S., and Mingolla, E. (2009). Cortical dynamics of navigation and steering in natural scenes: motion-based object segmentation, heading, and obstacle avoidance. *Neural Netw.* 22, 1383–1398. doi: 10.1016/j.neunet.2009.05.007

Calow, D., and Lappe, M. (2007). Local statistics of retinal optic flow for self-motion through natural sceneries. *Network* 18, 343–374. doi: 10.1080/09548980701642277

Carandini, M., Anthony Movshon, J., and Ferster, D. (1998). Pattern adaptation and cross-orientation interactions in the primary visual cortex. *Neuropharmacology* 37, 501–511. doi: 10.1016/S0028-3908(98)00069-0

Chen, A., DeAngelis, G. C., and Angelaki, D. E. (2011). A comparison of vestibular spatiotemporal tuning in Macaque parietoinsular vestibular cortex, ventral intraparietal area, and medial superior temporal area. *J. Neurosci.* 31, 3082–3094. doi: 10.1523/JNEUROSCI.4476-10.2011

Cullen, K. E. (2019). Vestibular processing during natural self-motion: implications for perception and action. *Nat. Rev. Neurosci.* 20, 346–363. doi: 10.1038/s41583-019-0153-1

Dragoi, V., Sharma, J., and Sur, M. (2000). Adaptation-induced plasticity of orientation tuning in adult visual cortex. *Neuron* 28, 287–298. doi: 10.1016/S0896-6273(00)00103-3

Duffy, C. J., and Wurtz, R. H. (1991). Sensitivity of MST neurons to optic flow stimuli. I. A continuum of response selectivity to large-field stimuli. *J. Neurophysiol.* 65, 1329–1345. doi: 10.1152/jn.1991.65.6.1329

Durgin, F. H., and Gigone, K. (2007). Enhanced optic flow speed discrimination while walking: contextual tuning of visual coding. *Perception* 36, 1465–1475. doi: 10.1068/p5845

Farnebäck, G. (2003). "Two-frame motion estimation based on polynomial expansion," in *Image Analysis. SCIA 2003. Lecture Notes in Computer Science*, Vol. 2749, eds J. Bigun and T. Gustavsson (Berlin: Springer), 363–370. doi: 10.1007/3-540-45103-X_50

Felleman, D. J., and Kaas, J. H. (1984). Receptive-field properties of neurons in middle temporal visual area (MT) of owl monkeys. *J. Neurophysiol.* 52, 488–513. doi: 10.1152/jn.1984.52.3.488

Foulkes, A. J., Rushton, S. K., and Warren, P. A. (2013). Heading recovery from optic flow: comparing performance of humans and computational models. *Front. Behav. Neurosci.* 7:53. doi: 10.3389/fnbeh.2013.00053

Ganguli, D., and Simoncelli, E. P. (2014). Efficient sensory encoding and bayesian inference with heterogeneous neural populations. *Neural Comput.* 26, 2103–2134. doi: 10.1162/NECO_a_00638

Gibson, J. J. (1950). *The Perception Of The Visual World*. Boston, MA: Houghton Mifflin.

Greenlee, M. W., Frank, S. M., Kaliuzhna, M., Blanke, O., Bremmer, F., Churan, J., et al. (2016). Multisensory integration in self motion perception. *Multisens. Res.* 29, 525–556. doi: 10.1163/22134808-00002527

Gu, Y., Watkins, P. V., Angelaki, D. E., and DeAngelis, G. C. (2006). Visual and nonvisual contributions to three-dimensional heading selectivity in the medial superior temporal area. *J. Neurosci.* 26, 73–85. doi: 10.1523/JNEUROSCI.2356-05.2006

Gutnisky, D. A., and Dragoi, V. (2008). Adaptive coding of visual information in neural populations. *Nature* 452, 220–224. doi: 10.1038/nature06563

Hosoya, T., Baccus, S. A., and Meister, M. (2005). Dynamic predictive coding by the retina. *Nature* 436, 71–77. doi: 10.1038/nature03689

Howard, J., Blakeslee, B., and Laughlin, S. B. (1987). The intracellular pupil mechanism and photoreceptor signal: noise ratios in the fly *Lucilia cuprina*. *Proc. R. Soc. Lond. B. Biol. Sci.* 231, 415–435. doi: 10.1098/rspb.1987.0053

Huang, X., Albright, T. D., and Stoner, G. R. (2007). Adaptive surround modulation in cortical area MT. *Neuron* 53, 761–770. doi: 10.1016/j.neuron.2007.01.032

Lappe, M., and Rauschecker, J. P. (1994). Heading detection from optic flow. *Nature* 369, 712–713. doi: 10.1038/369712a0

Laughlin, S. (1981). A Simple coding procedure enhances a neuron's information capacity. *Z. Naturforsch. C* 36, 910–912. doi: 10.1515/znc-1981-9-1040

Laughlin, S. B. (1989). The role of sensory adaptation in the retina. *J. Exp. Biol.* 146, 39–62.

Layton, O. W., and Fajen, B. R. (2016a). A neural model of MST and MT Explains perceived object motion during self-motion. *J. Neurosci.* 36, 8093–8102. doi: 10.1523/JNEUROSCI.4593-15.2016

Layton, O. W., and Fajen, B. R. (2016c). Sources of bias in the perception of heading in the presence of moving objects: object-based and border-based discrepancies. *J. Vis.* 16:9. doi: 10.1167/16.1.9

Layton, O. W., and Fajen, B. R. (2016b). Competitive dynamics in MSTd: a mechanism for robust heading perception based on optic flow. *PLoS Comput. Biol.* 12:e1004942. doi: 10.1371/journal.pcbi.1004942

Layton, O. W., and Fajen, B. R. (2017). Possible role for recurrent interactions between expansion and contraction cells in MSTd during self-motion perception in dynamic environments. *J. Vis.* 17, 1–21. doi: 10.1167/17.5.5

Layton, O. W., Mingolla, E., and Browning, N. A. (2012). A motion pooling model of visually guided navigation explains human behavior in the presence of independently moving objects. *J. Vis.* 12, 1–19. doi: 10.1167/12.1.20

Li, L., and Cheng, J. C. K. (2011). Heading but not path or the tau-equalization strategy is used in the visual control of steering toward a goal. *J. Vis.* 11, 1–12. doi: 10.1167/11.12.20

Liu, L. D., and Pack, C. C. (2017). The contribution of area MT to visual motion perception depends on training. *Neuron* 95, 436–446.e3. doi: 10.1016/j.neuron.2017.06.024

Longuet-Higgins, H. C., and Prazdny, K. (1980). The interpretation of a moving retinal image. *Proc. R. Soc. Lond. B* 208, 385–397. doi: 10.1098/rspb.1980.0057

Lucas, B., and Kanade, T. (1981). "An iterative image registration technique with an application to stereo vision," in *Proceedings of DARPA Image Understanding Workshop (IUW '81)*, Vancouver, BC.

Müller, J. R., Metha, A. B., Krauskopf, J., and Lennie, P. (1999). Rapid adaptation in visual cortex to the structure of images. *Science* 285, 1405–1408. doi: 10.1126/science.285.5432.1405

Nover, H., Anderson, C., and DeAngelis, G. (2005). A logarithmic, scale-invariant representation of speed in macaque middle temporal area accounts for speed discrimination performance. *J. Neurosci.* 25, 10049–10060. doi: 10.1523/JNEUROSCI.1661-05.2005

Olshausen, B. A., and Field, D. J. (1997). Sparse coding with an overcomplete basis set: a strategy employed by V1? *Vis. Res.* 37, 3311–3325. doi: 10.1016/S0042-6989(97)00169-7

Pack, C. C., and Born, R. T. (2001). Temporal dynamics of a neural solution to the aperture problem in visual area MT of macaque brain. *Nature* 409, 1040–1042. doi: 10.1038/35059085

Perrone, J. A. (2012). A neural-based code for computing image velocity from small sets of middle temporal (MT/V5) neuron inputs. *J. Vis.* 12, 1–31. doi: 10.1167/12.8.1

Royden, C. S. (2002). Computing heading in the presence of moving objects: a model that uses motion-opponent operators. *Vis. Res.* 42, 3043–3058. doi: 10.1016/S0042-6989(02)00394-2

Shapley, R., and Enroth-Cugell, C. (1984). Visual adaptation and retinal gain controls. *Eng. Sci.* 3, 263–346.

Simoncelli, E. P. (2003). Vision and the statistics of the visual environment. *Curr. Opin. Neurobiol.* 13, 144–149. doi: 10.1016/S0959-4388(03)00047-3

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

Steinmetz, S. T., Layton, O. W., Browning, N. A., Powell, N. V., and Fajen, B. R. (2019). An integrated neural model of robust self-motion and object motion perception in visually realistic environments. *J. Vis.* 19:294a. doi: 10.1167/19.10.294a

Wainwright, M. J. (1999). Visual adaptation as optimal information transmission. *Vis. Res.* 39, 3960–3974. doi: 10.1016/S0042-6989(99)00101-7

Warren, W. H., Blackwell, A. W., Kurtz, K. J., Hatsopoulos, N. G., and Kalish, M. L. (1991). On the sufficiency of the velocity field for perception of heading. *Biol. Cybern.* 65, 311–320. doi: 10.1007/BF00216964

Warren, W. H., Kay, B., Zosh, W., Duchon, A., and Sahuc, S. (2001). Optic flow is used to control human walking. *Nat. Neurosci.* 4, 213–216. doi: 10.1038/84054

Warren, W. H., Morris, M. W., and Kalish, M. (1988). Perception of translational heading from optical flow. *J. Exp. Psychol.* 14, 646–660. doi: 10.1037/0096-1523.14.4.646

Webster, M. A. (2015). Visual adaptation. *Annu. Rev. Vis. Sci.* 1, 547–567. doi: 10.1146/annurev-vision-082114-035509

Werblin, F. S. (1974). Control of retinal sensitivity. *J. Gen. Physiol.* 63, 62–87. doi: 10.1085/jgp.63.1.62

Xiao, J., and Huang, X. (2015). Distributed and dynamic neural encoding of multiple motion directions of transparently moving stimuli in cortical area MT. *J. Neurosci.* 35, 16180–16198. doi: 10.1523/JNEUROSCI.2175-15.2015

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

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.