A normalization mechanism for estimating visual motion across speeds and scales

Nikos Gekas[1,2], Andrew I. Meso[3,4], Guillaume S. Masson[4], and Pascal Mamassian[1,2]

[1] Laboratoire des Systèmes Perceptifs, CNRS UMR 8248, 29 rue d'Ulm, 75005 Paris, France

[2] Institut d'Etude de la Cognition, Ecole Normale Supérieure - PSL Research University, 75005 Paris, France

[3] Psychology & Interdisciplinary Neuroscience Research, Faculty of Science and Technology, Bournemouth University, Poole, UK

[4] Institut de Neurosciences de la Timone, UMR7289, CNRS, Aix-Marseille Université, Marseille, France

Corresponding authors:

Nikos Gekas and Pascal Mamassian

Laboratoire des Systèmes Perceptifs (CNRS UMR 8248), Ecole Normale Supérieure, 29 rue d'Ulm, 75005 Paris, France

Emails: nikos.gekas@outlook.com and pascal.mamassian@ens.fr

Summary

The use of simple stimuli has been tremendously beneficial to understand the basic properties of primary sensory systems. However, interacting with the natural environment leads to complex stimulations of our senses. Here we focus on the estimation of visual speed, a critical information for the survival of many animal species as they monitor moving prey or approaching dangers. In mammals, and in particular in primates, speed information is conceived to be represented by a set of channels sensitive to different spatial and temporal characteristics of the optic flow [1-5]. However, it is still largely unknown how the brain could accurately infer the speed of complex natural scenes from this set of spatiotemporal channels [6-14]. As complex stimuli, we chose a set of well-controlled moving naturalistic textures called 'Compound Motion Clouds' (CMC) [15, 16] that simultaneously activate multiple spatiotemporal channels. We found that CMC stimuli that have the same physical speed are perceived moving at different speeds depending on which channels are activated. Thanks to a computational model, we show that the activity in a given channel is both boosted and weakened following a systematic pattern over neighboring channels. This pattern of interactions can be understood as a combination of two components oriented in speed (consistent with a slow-speed prior) and scale (sharpening of similar features). Interestingly, the interaction along scale implements a lateral inhibition principle that is usually found to operate in early sensory processing. These results further promote the idea that lateral inhibition along a variety of dimensions is a canonical principle of perceptual processing. Overall, the speed-scale normalization mechanism may reflect the natural tendency of the visual system to integrate complex inputs into one coherent percept.

Results

We created Compound Motion Cloud (CMC) stimuli to stimulate a limited number of speed channels. These naturalistic CMC stimuli have three components that are centered on three slightly different speeds. Each of the three components is a Motion Cloud (MC) stimulus that looks like a texture of a certain scale moving at a certain speed [10, 15, 16]. In the space commonly used to represent speed where the logarithm of temporal frequency is plotted against the logarithm of spatial frequency (so-called 'log-log frequency space'; Figure 1A), an MC stimulus is an ellipse oriented along a speed diagonal. In that log-log space, the simple gratings that are often used as stimuli in motion experiments are single points. MC stimuli are thus more naturalistic generalizations of gratings for which the bandwidths of spatial frequency (B_{sf}) and speed (B_v) can be parametrically manipulated (Supplemental Experimental Procedures). Our CMC stimuli are generated such that its three MC components move at three specific speeds and span at least two different scales (Figure 1B). All CMC stimuli moved at the same physical average speed but each activated distinct channels (two examples of CMC stimuli are shown at full contrast in Figure 1C). In a psychophysical forced-choice task (Figure 1D), human participants had to match the perceived speed of CMC stimuli to that of Random Dot Kinematograms (RDK; a stimulus composed of multiple dots, all moving at the same speed) (Figure 1E). We reasoned that if there are some interactions between speed channels, different CMC stimuli should be matched to different speeds (Figures 1F & 1G).

Psychophysical Results

When the components are far enough apart in the spatiotemporal log space, the CMC stimulus can appear as being composed of multiple interleaved stimuli moving at distinct speeds. Motion segregation is in itself an important topic [17], but we focus here on the perceived speed of coherent motion. Therefore, in a preliminary experiment, we measured the distance at which participants are equally likely to perceive that two superimposed MC stimuli appear to move as a coherent stimulus or as two transparent layers (Figure S1). Using these boundary distance values (Figure 2A), we generated CMC stimuli for 3 distinct mean spatial frequencies. Adding a third (middle) component to the compound should increase the perception of coherency of the stimulus but still allow for a large enough distance for the interactions between components to be meaningful. We considered 6 conditions (Figure 2B): C1 and C2 in which all components had the same mean spatial or temporal frequency; C3 and C4 in which only two of the components had the same mean spatial or temporal frequency, and C5 and C6 in which all components had different mean spatial and temporal frequencies. Out of 12 possible combinations of components (Figure S2), the 6 selected conditions capture all 14 possible relative interactions between components (Table S1), thus making the use of additional conditions redundant.

Figure 2C illustrates the matching speed for each of the 6 conditions and the 3 mean spatial frequencies. There was a significant effect of condition on perceived matching speed for $\Delta = 0.75$ ($F(5,35) = 54.03$, $p < 0.001$), $\Delta = 0.55$ ($F(5,35) = 19.24$, $p < 0.001$), and for $\Delta = 0.25$ ($F(5,35) = 3.18$, $p = 0.018$). While participants were less sensitive to stimuli with lower mean spatial frequencies on average, we did not observe any significant differences between the 6 conditions (Figure 2D). To ensure that these biases were not due to differences in the motion energy of our stimuli, we analyzed the experimental CMCs with a standard computational model of neurons in the middle temporal (MT) visual area [18]. The simulations show that any differences between conditions are minimal (Figure S3).

Matching speed differences between conditions increase as Δ increases. Many variables can have strong effects on the perceived speed of a stimulus such as contrast [19], spatial frequency [8, 20], luminance and chrominance [21] or even attention [22]. Particularly for spatial frequency up to 2 c/deg [23, 24], it has been shown that stimuli with higher spatial frequency are perceived to move faster than stimuli with lower frequency. As our stimuli were in the range of

0.15 to 1.19 c/deg, we also observed that components with the same actual speed were perceived as moving faster for higher frequencies than for lower frequencies. To test whether these differences are sufficient to explain all the perceptual biases in our results, we built a set of simple models that include or exclude interactions between speed channels.

Computational Models

In order to quantify the potential interactions between speed channels, we designed different variants of a model that computes the speed likelihood of a neural population [25] (Figure 3A). In the basic model, each CMC stimulus consists of 3 components whose input activities are assumed to be equal. The activities are first normalized through a gain control procedure

$$n_i = \frac{m_i^2}{c^2 + \sum_j m_j^2}, \qquad \text{(Equation 1)}$$

where m is the input activity and c is the semi-saturation parameter [26]. The normalized activity n is then passed through an interaction matrix, so that each component can be boosted or weakened by the activity of the neighboring components

$$\log o_i = \log n_i + \sum_{j \neq i} w_j \log n_j, \qquad \text{(Equation 2)}$$

where w is the interaction weight. Finally, the output activity o_i is multiplied with the log likelihood of each of the components and the log likelihood of the speed for that stimulus is obtained by adding the products

$$\log L(s) = \sum_i o_i \log L_i, \qquad \text{(Equation 3)}$$

where L_i is the speed likelihood of one component. Using the same speed discrimination procedure as for the CMC stimuli, we measured the psychometric functions for each of the seven individual components across all mean spatial frequencies for the same human participants. From these psychometric functions, we built the speed likelihood for each component by taking the derivative of the fitted cumulative distribution function. Then, the combined likelihood was used to match the psychometric functions of each participant. The interaction matrix includes different weights for temporal and spatial frequency and for different distances between components in log space. It is assumed that the weights are symmetric with respect to the origin (Table S1). More details on how the models fit the experimental data can be found in Supplemental Information.

An example of three component likelihoods and the combined likelihood predicted by the *Interaction* model is shown in Figure 3B (left). We also considered two simpler models: a *Maximum-likelihood estimation* (*MLE*) model in which the combined likelihood is the product of the component likelihoods (Figure 3B center), and an *Averaging* model in which the combined likelihood is the average of the component likelihoods (Figure 3B right). These models have no free parameters and ignore any potential interactions between components.

The *Interaction* model outperforms the other two models in both predicted matching speed (Figure 3C) and sensitivity (Figure 3D). It captures the biases across conditions and better matches the overall sensitivity of the combined stimulus at the highest *Δ* value. It has 5 more free parameters than the other two models so better fitting of the data is anticipated. We evaluated the goodness-of-fit of all models by measuring the Akaike information criterion (AIC) values of each model for each participant and mean spatial frequency. The preferred model is the one that minimizes the AIC values. The *Interaction* model has the minimum value in 20 out of 24 comparisons, while the *Averaging* model in 4 out of 24 comparisons (Figure S4). Overall, the *Interaction* model thus provides a better fit than the other models and is the best option among the three presented. We also considered variations of the *Interaction* model; for example, implementing weights only for the temporal or spatial frequency dimensions, or weights that are not symmetric with respect to the origin. The version of the model presented here provides an overall minimum AIC value compared with these later variations (Figure S4).

Normalization Mechanism

Next, we estimated the shape of this interaction pattern. To increase the power of our analysis, we collapsed the interaction weights across all three mean spatial frequencies. We then applied cubic surface interpolation [27] on the resulting 42 interaction weights to create a continuous 2D surface that passes through all of them. Contour plots of the weights for the average participant are illustrated in Figure 4B (individual plots are presented in Figure S5). The values indicate the effect that neighboring channels have on the central channel depending on their distance in spatial and temporal frequency. Positive values (light grey) indicate excitation (a boost in activity) and negative values (dark grey) inhibition (a weakening). Excitation/inhibition interaction effects appear stronger at around half an octave away from the origin and dissipate as distance increases.

The pattern of interactions may appear as overly complex when considered along the spatial and temporal frequency axes, but it is in fact considerably simpler when expressed along the scale and speed axes. The scale axis is the diagonal of the log-log space that correspond to all the combinations of spatial and temporal frequencies matching the same speed [28]. The speed axis is orthogonal to the scale axis. The interaction between channels can be seen as a combination of two components along these two axes (Figure 4C). Along the speed axis (Figure 4C up left), there is a sine-like component where channels that encode higher speeds boost the activity of the central channel, while channels that encode lower speeds weaken it. This component decreases the average speed of a CMC stimulus and may represent a form of a slow speed prior [29]. Along the scale axis (Figure 4C up right), there is a cosine-like component where outer channels inhibit the central channel, while channels in the middle of the axis excite the central channel. This component may help in the fine-tuning of the spectral properties of the stimulus and, subsequently, in the perception of one single coherent percept rather than a broadband noisy stimulus. When the two components are multiplied together (Figure 4C bottom), the pattern of channel interaction approximates the pattern of the average weights of our computational model. At the implementation level, the product of these two components can be expressed differently, for instance as a simple difference [30] (Figure S6).

Discussion

How speed information is encoded in mammalian brains including the human perceptual system remains a mystery. In their seminal paper, Adelson and Bergen [31] suggested that velocity may be derived by comparing the outputs of several spatiotemporal channels within the same spatial frequency band. A specific class of models known as Weighted Intersection Mechanism (WIM) [12, 14, 32-34] assume separate magno- (band-pass temporal) and parvo-cellular (low-pass temporal) components that offer a mechanism to explain the origin of spatiotemporal channels. These models nicely account for neural results such as cortical sensitivity changes with contrast, but they are not applicable in our case because the large bandwidth of the filters assumed in these models would blur our MCs inputs and conceal the interaction patterns we sought to measure. A more recent model by Perrone [9] includes inhibitory localized interactions between direction-selective units prior to the estimation of velocity. This local inhibition is assumed to take place within monkey middle temporal (MT) cortical area, and it is consistent with measured properties in actual MT neurons [35, 36]. However, this process is more likely to apply only for very simple stimuli like gratings, which contain only one spatial and temporal frequency pair. Our stimuli contain multiple spectral components and so our results may correspond to the next stage in Perrone's model where the outputs of velocity channels are integrated to derive the actual velocity of the stimulus. This stage is assumed to take place somewhere between MT and the medial superior temporal (MST) area, where multiple MT channels feed their activity to MST neurons that pool inputs from several of such channels across a large area of the visual field. Up to now, this pooling mechanism was usually assumed to only be a weighted average of the output of the MT activity

[9, 13, 37]. We have shown here that this integration process might in fact be more complex than currently assumed.

In a study investigating integration along the same speed line [13], the equivalent averaging and MLE models predicted different matching speeds for composite grating stimuli but equal sensitivities. Here, the two models give different predictions for both matching speeds and sensitivities, thus providing more evidence towards a MLE decoding scheme. However, as the components diverge further in the speed dimension (e.g., for $\Delta = 0.75$), the MLE model fails to predict the decrease in sensitivity. Integration along the same speed line may be consistent with the model presented in [13] but our proposed normalization mechanism is required to explain the integration of different speed channels into a global speed percept. Moreover, the speed component of our proposed mechanism suggests a neurally plausible way of encoding perceptual priors in the same early cortical areas that provide sensory evidence [29, 37].

The interaction pattern that we found over the scale dimension belongs to the generic class of lateral inhibition. This principle is commonly found to explain early sensory interactions, such as the light interactions for nearby receptors in the retina [38], orientation interactions in the visual cortical neurons [39], tone interactions in auditory cortical neurons [40], and other interactions in the olfactory bulb [41] and in the primary somatosensory cortex [42]. However, it has so far been possible to identify these lateral interaction patterns in human perception only for low level detection tasks [43]. Here, we show that lateral inhibition is playing a critical role for a more complex perceptual task, speed processing. This result further corroborates the postulate from von Békésy that lateral inhibition is a general characteristic of the nervous system [44], possibly even more widespread than other mechanisms such as gain control [26]. Our study also opens the door for modeling the response properties of MT neurons to complex motion (e.g. [3]) as well as to unveil the excitation/inhibition interactions between these cells.

Our findings contribute to a novel understanding of the neural mechanisms serving speed perception. To the best of our knowledge, the interaction over scale to estimate speed has never been reported. We speculate that it contributes to our biased impression of cohesiveness when we look at complex optic flows. Very similar mechanisms are seen throughout the perceptual and motor systems implemented via lateral inhibition, and may be a recurring canonical principle in the efforts of the central nervous system to achieve more and more complex computations.

Experimental Procedures

Observers

Eight participants took part in the main experiment. They all had normal or corrected-to-normal vision and all gave informed consent before the experiment. All but one participants were naive with regard to the purpose of the study.

Procedure

The experimental task was a 2 Alternative Forced Choice task (Figure 1D) where participants had to compare the speed of a CMC stimulus with that of a random-dot kinematogram (RDK: contrast: 20%, density: 4 dots/deg^2, and dot diameter: 0.1 deg) moving at 1 of 6 possible speeds (1.64, 2.08, 2.66, 3.39, 4.32, and 5.5 deg/s). All stimuli were shown at fixation and inside a circular aperture (radius: 5.3 deg) for 300 ms. A raised cosine filter was used at the fringes of the aperture. The CMC stimulus was always shown first, followed by the RDK after an inter-stimulus interval of 450ms. We did not randomize the order of presentation of the 2 types of stimuli in order to reduce variability related to the order of presentation (e.g., [45]). Since we were interested in a comparison across different CMC stimuli and not between CMC stimuli and the RDKs, any pre-existing response bias would apply equally to all CMC stimuli. Participants did 40 trials for each condition and RDK speed in three one-hour sessions (1440 trials) for each of the three mean spatial frequencies on separate days. The order of the sessions was randomized between participants. We pre-generated multiple instances of MC stimuli for all components, creating a database of CMC stimuli with the same frequency properties but different phase, and we presented them in a randomized order. All stimuli were generated using the Matlab programming language with the psychophysics toolbox [46] and displayed on a CRT monitor with a resolution of 1280 X 960 pixels at 100 Hz. Participants viewed the display in a darkened room at a viewing distance of 60 cm, and a chin rest was used to maintain a constant head location and viewing distance.

Supplemental Information

Supplemental Information includes six figures and one table.

Acknowledgements

This work was first presented at the annual conference of the Vision Sciences Society in St. Pete Beach, Florida, in May 2016 [47]. The research has been funded by a grant from the Agence National de la Recherche (SPEED, ANR-13-SHS2-0006) to GSM and PM. Additional support came from ANR-10-LABX-0087 IEC and ANR-10-IDEX-0001-02 PSL* to PM. We thank Laurent Perrinet for providing us the source code of the Motion Clouds generation algorithm. We thank John Perrone for helpful discussions.

References

1. Watson, A.B. and Ahumada Jr, A.J. (1983). A look at motion in the frequency domain. In Motion: Perception and Representation, J.K. Tsotsos, ed. (New York: Association for Computing Machinery), pp. 1-10.

2. Perrone, J.A. and Thiele, A. (2001). Speed skills: measuring the visual speed analyzing properties of primate MT neurons. Nat. Neurosci. *4*, 526-532.

3. Priebe, N.J., Cassanello, C.R., and Lisberger, S.G. (2003). The neural representation of speed in

macaque area MT/V5. J. Neurosci. *23*, 5650-5661.

4. Priebe, N. J., & Lisberger, S. G. (2004). Estimating target speed from the population response in visual area MT. J. Neurosci. *24*, 1907-1916.

5. Priebe, N.J., Lisberger, S.G., and Movshon, J.A. (2006). Tuning for spatiotemporal frequency and speed in directionally selective neurons of macaque striate cortex. J. Neurosci. *26*, 2941-2950.

6. Perrone, J.A. (2006). A single mechanism can explain the speed tuning properties of MT and V1 complex neurons. J. Neurosci. *26*, 11987-11991.

7. Inaba, N., Shinomoto, S., Yamane, S., Takemura, A., and Kawano, K. (2007). MST neurons code for visual motion in space independent of pursuit eye movements. J. Neurophysiol. *97*, 3473-3483.

8. Brooks, K.R., Morris, T., & Thompson, P. (2011). Contrast and stimulus complexity moderate the relationship between spatial frequency and perceived speed: Implications for MT models of speed perception. J. Vis. *11*, 19-19.

9. 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*, 1-1.

10. Simoncini, C., Perrinet, L.U., Montagnini, A., Mamassian, P., and Masson, G.S. (2012). More is not always better: adaptive gain control explains dissociation between perception and action. Nat. Neurosci. *15*, 1596–1603.

11. Meso, A.I. and Simoncini, C. (2014). Towards an understanding of the roles of visual areas MT and MST in computing speed. Front. Comput. Neurosci. *8*.

12. Hassan, O., & Hammett, S.T. (2015). Perceptual biases are inconsistent with Bayesian encoding of speed in the human visual system. J. Vis. *15*, 9-9.

13. Jogan, M. and Stocker, A.A. (2015). Signal integration in human visual speed perception. J. Neurosci *35*, 9381–9390.

14. Hassan, O., Thompson, P., & Hammett, S.T. (2016). Perceived speed in peripheral vision can go up or down. J. Vis. *16*, 20-20.

15. Leon, P.S., Vanzetta, I., Masson, G.S., and Perrinet, L.U. (2012). Motion clouds: model-based stimulus synthesis of natural-like random textures for the study of motion perception. J. Neurophysiol. *107*, 3217–3226.

16. Vacher, J., Meso, A.I., Perrinet, L.U., and Peyre, G. (2015). Biologically inspired dynamic textures for probing motion perception. In Advances in Neural Information Processing Systems 28, C. Cortes, N.D. Lawrence, D.D. Lee, M. Sugiyama, and R. Garnett, eds., pp. 1918–1926.

17. Hedges, J.H., Stocker, A.A., & Simoncelli, E.P. (2011). Optimal inference explains the perceptual

coherence of visual motion stimuli. J. Vis. *11*, 14-14.

18. Simoncelli, E.P., & Heeger, D.J. (1998). A model of neuronal responses in visual area MT. Vision Res. *38*, 743-761.

19. Thompson, P. (1982). Perceived rate of movement depends on contrast. Vision Res. *22*, 377-380.

20. Smith, A. and Edgar, G. (1990). The influence of spatial frequency on perceived temporal frequency and perceived speed. Vision Res. *30*, 1467-1474.

21. Cavanagh, P., Tyler, C.W., and Favreau, O.E. (1984). Perceived velocity of moving chromatic gratings. J. Opt. Soc. Am. A *1*, 893-899.

22. Turatto, M., Vescovi, M., and Valsecchi, M. (2007). Attention makes moving objects be perceived to move faster. Vision Res. *47*, 166-178.

23. Diener, H., Wist, E., Dichgans, J., and Brandt, T. (1976). The spatial frequency effect on perceived velocity. Vision Res. *16*, 169-176.

24. McKee, S.P., Silverman, G.H., and Nakayama, K. (1986). Precise velocity discrimination despite random variations in temporal frequency and contrast. Vision Res. *26*, 609-619.

25. Jazayeri, M. and Movshon, J.A. (2006). Optimal representation of sensory information by neural populations. Nat. Neurosci. *9*, 690-696.

26. Carandini, M., & Heeger, D.J. (2012). Normalization as a canonical neural computation. Nat. Rev. Neurosci. *13*, 51-62.

27. De Boor, C. (1978). A practical guide to splines. (New York: Springer-Verlag).

28. Watson, A.B. and Ahumada, A.J. (1985). Model of human visual-motion sensing. J. Opt. Soc. Am. A *2*, 322-342.

29. Stocker, A.A., & Simoncelli, E.P. (2006). Noise characteristics and prior expectations in human visual speed perception. Nat. Neurosci. *9*, 578-585.

30. Simoncelli, E.P. (2003). Local analysis of visual motion. In The Visual Neurosciences, L.M. Chalupa and J.S. Werner, ed. (MIT Press), pp. 1616-1623.

31. Adelson, E.H. and Bergen, J.R. (1985). Spatiotemporal energy models for the perception of motion. J. Opt. Soc. Am. A *2*, 284–299.

32. Perrone, J.A., & Thiele, A. (2002). A model of speed tuning in MT neurons. Vision Res. *42*, 1035-1051.

33. Hammett, S.T., Champion, R.A., Morland, A.B., & Thompson, P.G. (2005). A ratio model of

perceived speed in the human visual system. Proc. R. Soc. B *272*, 2351-2356.

34. Meese, T.S., & Baker, D.H. (2009). Cross-orientation masking is speed invariant between ocular pathways but speed dependent within them. J. Vis. *9*, 2-2.

35. Xiao, D., Raiguel, S., Marcar, V., and Orban, G. (1997). The spatial distribution of the antagonistic surround of MT/V5 neurons. Cereb. Cortex *7*, 662-677.

36. Pack, C.C., Hunter, J.N., and Born, R.T. (2005). Contrast dependence of suppressive influences in cortical area MT of alert macaque. J. Neurophysiol. *93*, 1809-1815.

37. Vintch, B., & Gardner, J.L. (2014). Cortical correlates of human motion perception biases. J. Neurosci. *34*, 2592-2604.

38. Hartline, H.K., Wagner, H.G., & Ratcliff, F. (1956). Inhibition in the eye of Limulus. J. Gen. Physiol. *39*, 651-673.

39. Blakemore, C., & Tobin, E.A. (1972). Lateral inhibition between orientation detectors in the cat's visual cortex. Exp. Brain Res. *15*, 439-440.

40. Shamma, S.A. & Symmes, D. (1985). Patterns of inhibition in auditory cortical cells in awake squirrel monkeys. Hear. Res. *19*, 1-13.

41. Rall, W., Shepherd, G.M., Reese, T.S., & Brightman, M.W. (1966). Dendrodendritic synaptic pathway for inhibition in the olfactory bulb. Exp. Neurol. *14*, 44-56.

42. Gardner, E.P., & Spencer, A. (1972). Sensory funneling. II. Cortical neuronal representation of patterned cutaneous stimuli. J. Neurophysiol. *35*, 954-977.

43. De Valois, R.L., and De Valois, K. (1988). Spatial Vision. (Oxford: Oxford University Press).

44. von Békésy, G. (1967). Sensory Inhibition. (Princeton, NJ: Princeton University Press).

45. Raviv, O., Lieder, I., Loewenstein, Y., and Ahissar, M. (2014). Contradictory behavioral biases result from the influence of past stimuli on perception. PLOS Comput. Biol. *10*, 1-10.

46. Brainard, D.H. (1997). The psychophysics toolbox. Spat. Vis. *10*, 433-436.

47. Gekas, N., Meso, A., Masson, G.S., and Mamassian, P. (2016). Speed channel interactions in naturalistic motion stimuli. J. Vis. *16*, 1131-1131.

Figure captions

Figure 1. Experimental paradigm. (A) In the log-log space representing spatial against temporal frequency, oblique lines indicate combinations of spatial and temporal frequency that correspond to the same speed. A moving grating corresponds to a single point in that space (red circle), whereas a motion cloud stimulus corresponds to an ellipse of which the spatial frequency (B_{sf}) and speed (B_v) bandwidths can be manipulated. (B) Test stimuli were made of 3 component MCs, one moving at the central speed, one moving faster (by Δ octaves) and one slower (likewise). (C) Two examples of CMC stimuli and their components are shown at full contrast. (D) Experimental procedure. In each trial, a CMC stimulus was always followed by a random dot stimulus moving at one out of 6 possible speeds. Participants were asked to indicate which stimulus was faster. (E) Psychometric functions provide the matching speed and sensitivity for each CMC stimulus against the same RDKs. (F, G) The disparities in matching speeds and sensitivities between CMC stimuli may show the nature of interaction between the components.

Figure 2. Psychophysical results. (A) From a preliminary psychophysical experiment, we measured the minimal distance Δ in spatial and temporal frequency at which two superimposed motion cloud stimuli appear to move transparently over each other. This distance decreases as mean spatial frequency increases. The red ellipses indicate the components that make up the CMC of condition 1 (C1) for each of the mean spatial frequencies and Δ values. (B) Test stimuli were generated from 6 combinations of 3 components moving at distinct speeds. Depending on the combination, components shared the same mean spatial and/or temporal frequency with other components or none at all. (C) Matching speeds and (D) sensitivities (i.e., the inverse of the standard deviation of the psychometric function) of the test stimuli (closed circles) are plotted for each condition and mean spatial frequency. Data are represented as mean ± 95% CI.

Figure 3. Computational models. (A) *Interaction* model. Each CMC stimulus consists of 3 components that make up the input, which is normalized with a gain control procedure. Then, the model assumes an interaction phase where each component's activity boosts or weakens the activity of the other 2 components. The output is multiplied with the log likelihoods of each component and the sum of the products is the log likelihood of the CMC stimulus. (B) Examples of CMC likelihood functions for each of the models. Grey curves indicate 3 example component likelihoods and colored curves the CMC likelihoods for each model. (left) *Interaction* model, (center) *MLE* model, (right) *Averaging* model. (C) Fitted matching speeds and (D) sensitivities are plotted for each condition and mean spatial frequency for each model (colored circles) along with the experimental data (white circles). Data are represented as mean ± 95% CI.

Figure 4. Model weights and normalization mechanism. (A) Interaction weights. Black dots indicate all 14 possible interactions of a channel and its neighbors from the 6 experimental conditions. (B) Weight contour plot of the average participant. Values indicate how neighboring channels affect the central channel based on their distance in spatial and temporal frequency. Positive values (in light grey) indicate excitation and negative values (in dark grey) inhibition. Colored dots indicate the coordinates in distance space that the model fitted the experimental data. The underlying surface was calculated using cubic surface interpolation from the weights by combining the fits from all 3 mean spatial frequencies. (C) Normalization mechanism. We reframe the interaction of channels in log-log frequency space along two orthogonal axes; speed and scale. Light grey areas indicate excitation and dark grey areas inhibition. Along the speed axis (top left), there is a sine-like component where channels of higher speeds excite the central channel, while channels of lower speeds inhibit the central channel. Along the scale axis (top right), there is a cosine-like component where outer channels inhibit the central channel, while channels in the middle of the axis excite the central channel. When the two components are multiplied together (bottom), the pattern of channel interaction approximates the pattern of the average weights of our computational model.

Figure 1

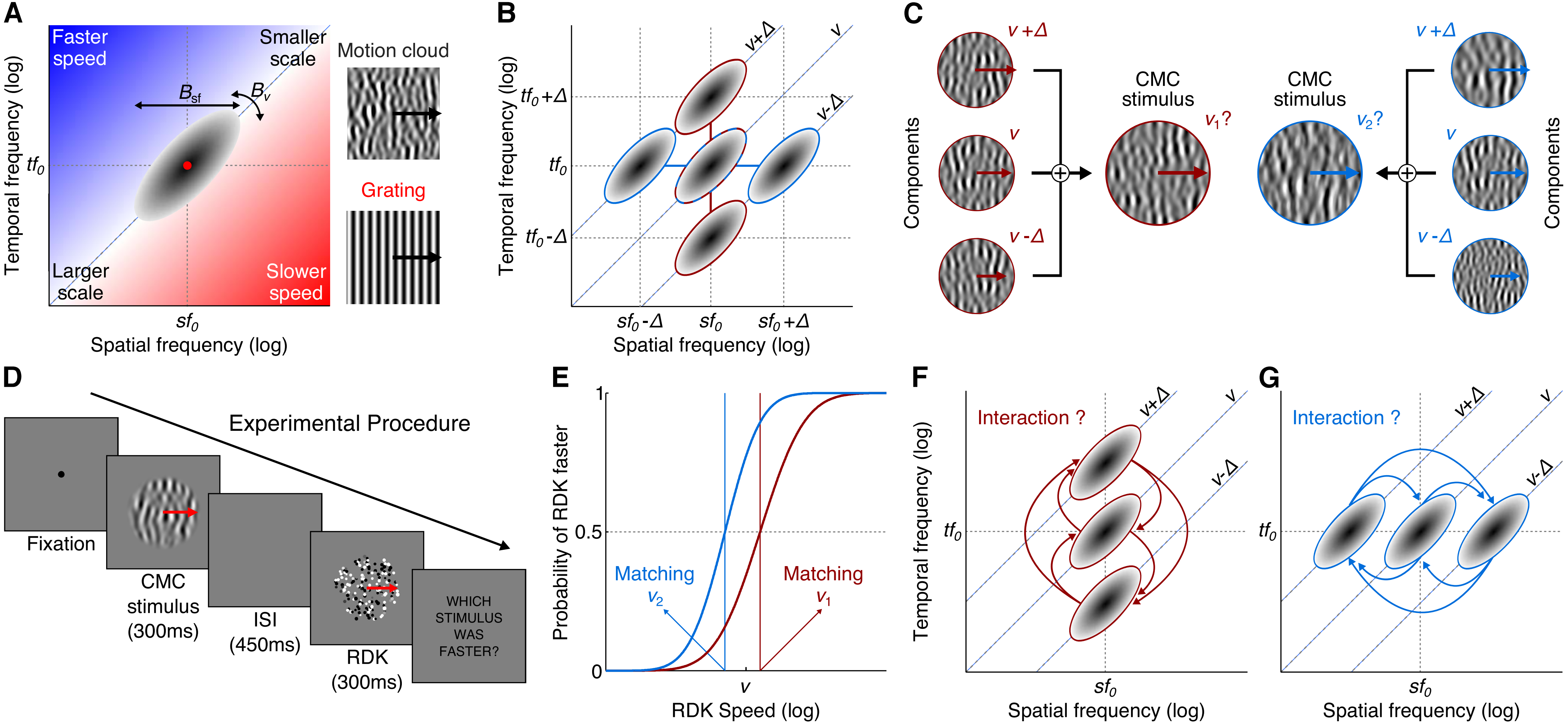
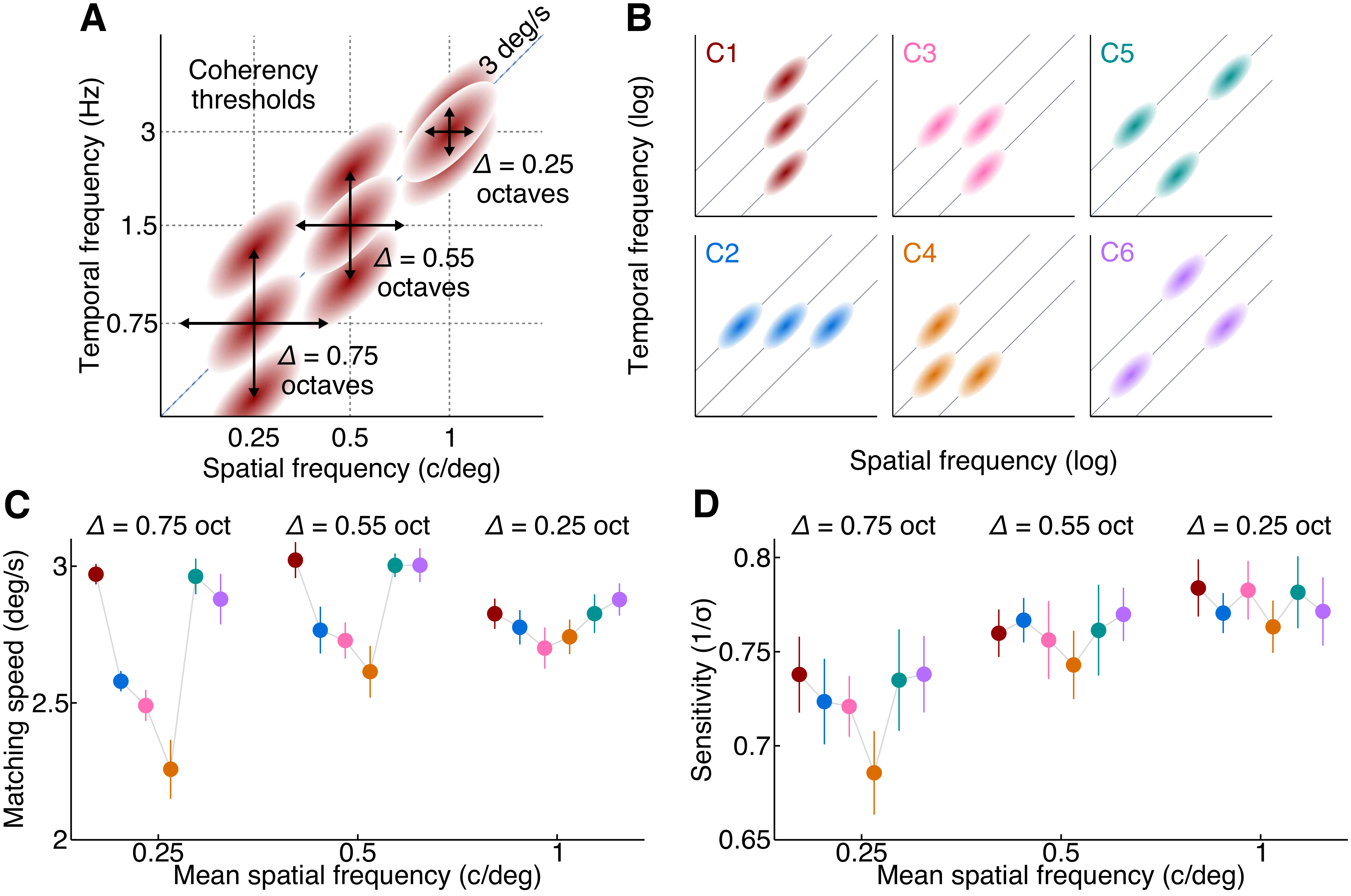

A
Temporal frequency (log)
tf_0
Faster speed
Smaller scale
Larger scale
Slower speed
B_{sf}
B_v
sf_0
Spatial frequency (log)
Motion cloud
Grating
B
Temporal frequency (log)
$tf_0+\Delta$
tf_0
$tf_0-\Delta$
$v+\Delta$
v
$v-\Delta$
$sf_0-\Delta$
sf_0
$sf_0+\Delta$
Spatial frequency (log)
C
Components
$v+\Delta$
v
$v-\Delta$
CMC stimulus
v_1?
CMC stimulus
v_2?
$v+\Delta$
v
$v-\Delta$
Components
D
Experimental Procedure
Fixation
CMC stimulus (300ms)
ISI (450ms)
RDK (300ms)
WHICH STIMULUS WAS FASTER?
E
Probability of RDK faster
1
0.5
0
Matching v_2
Matching v_1
v
RDK Speed (log)
F
Temporal frequency (log)
tf_0
Interaction ?
$v+\Delta$
v
$v-\Delta$
sf_0
Spatial frequency (log)
G
tf_0
Interaction ?
$v+\Delta$
v
$v-\Delta$
sf_0
Spatial frequency (log)

Figure 2



A
Coherency thresholds
3 deg/s
Temporal frequency (Hz)
3
1.5
0.75
Δ = 0.25 octaves
Δ = 0.55 octaves
Δ = 0.75 octaves
0.25
0.5
1
Spatial frequency (c/deg)
B
C1
C3
C5
C2
C4
C6
Temporal frequency (log)
Spatial frequency (log)
C
Δ = 0.75 oct
Δ = 0.55 oct
Δ = 0.25 oct
Matching speed (deg/s)
3
2.5
2
0.25
0.5
1
Mean spatial frequency (c/deg)
D
Δ = 0.75 oct
Δ = 0.55 oct
Δ = 0.25 oct
Sensitivity (1/σ)
0.8
0.75
0.7
0.65
0.25
0.5
1
Mean spatial frequency (c/deg)

Figure 3

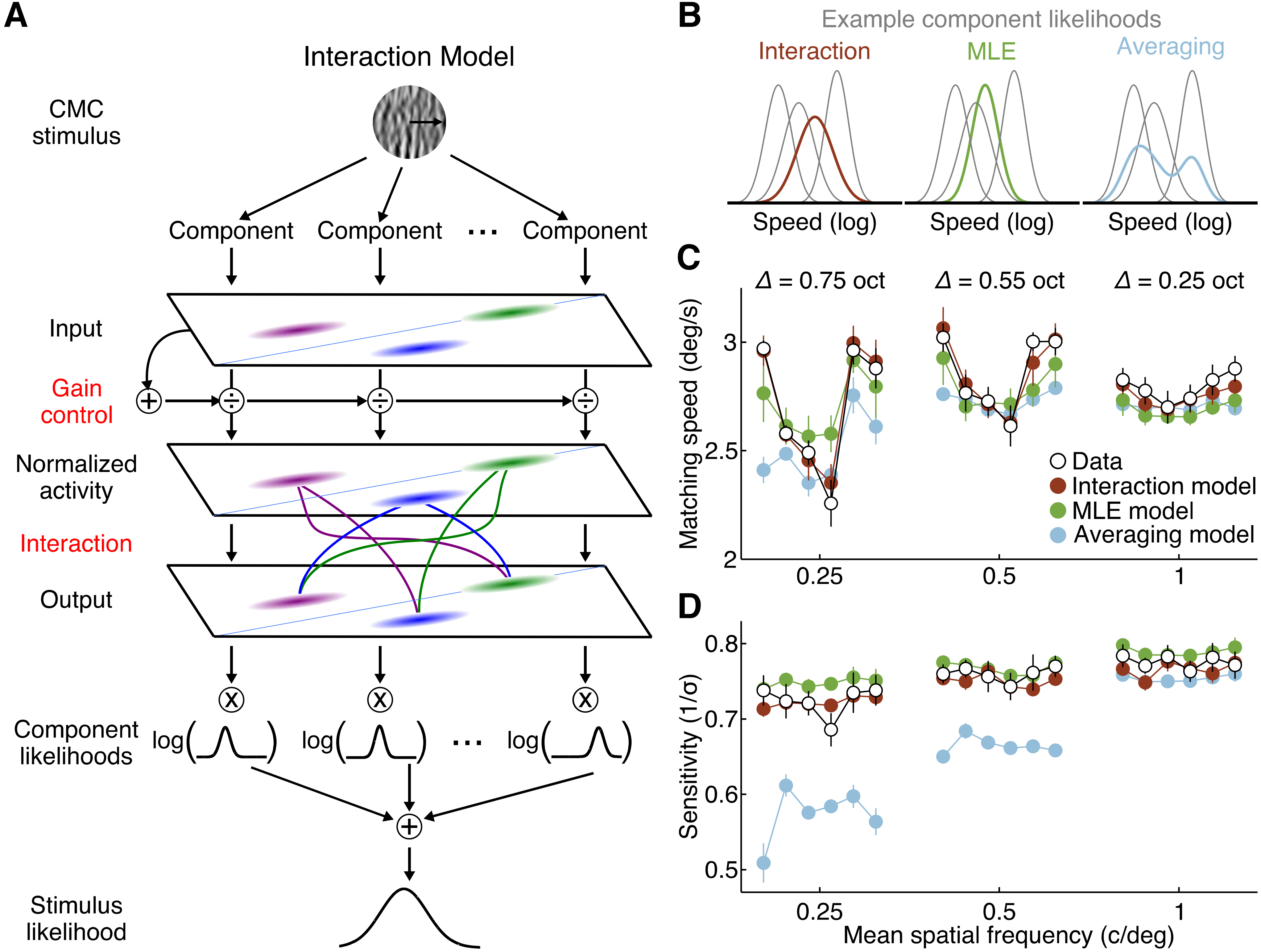
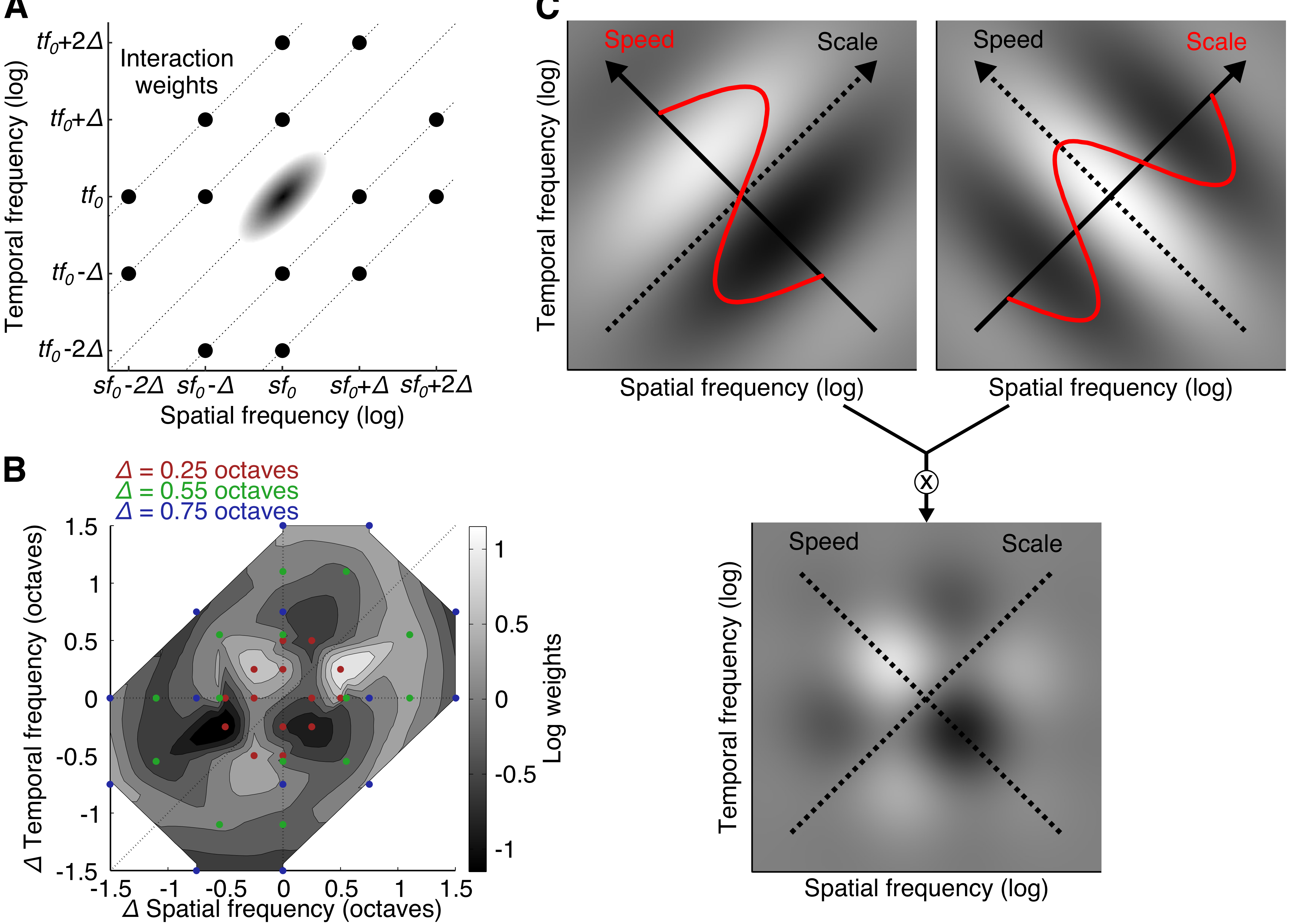

A
Interaction Model
CMC
stimulus
Component
Component
...
Component
Input
Gain
control
Normalized
activity
Interaction
Output
Component
likelihoods
log( )
log( )
...
log( )
Stimulus
likelihood
B
Example component likelihoods
Interaction
MLE
Averaging
Speed (log)
Speed (log)
Speed (log)
C
Δ = 0.75 oct
Δ = 0.55 oct
Δ = 0.25 oct
Matching speed (deg/s)
3
2.5
2
0.25
0.5
1
○ Data
● Interaction model
● MLE model
● Averaging model
D
Sensitivity (1/σ)
0.8
0.7
0.6
0.5
0.25
0.5
1
Mean spatial frequency (c/deg)

Figure 4



A
Interaction weights
$tf_0+2\Delta$
$tf_0+\Delta$
tf_0
$tf_0-\Delta$
$tf_0-2\Delta$
Temporal frequency (log)
$sf_0-2\Delta$
$sf_0-\Delta$
sf_0
$sf_0+\Delta$
$sf_0+2\Delta$
Spatial frequency (log)
B
Δ = 0.25 octaves
Δ = 0.55 octaves
Δ = 0.75 octaves
1.5
1
0.5
0
-0.5
-1
-1.5
Δ Temporal frequency (octaves)
-1.5
-1
-0.5
0
0.5
1
1.5
Δ Spatial frequency (octaves)
1
0.5
0
-0.5
-1
Log weights
C
Speed
Scale
Temporal frequency (log)
Spatial frequency (log)
Speed
Scale
Spatial frequency (log)
X
Speed
Scale
Temporal frequency (log)
Spatial frequency (log)

Supplemental Information for:

A normalization mechanism for estimating visual motion across speeds and scales

Nikos Gekas, Andrew I. Meso, Guillaume S. Masson, and Pascal Mamassian

Inventory of Supplemental Information:

- Figure S1. Preliminary experiment for identifying coherency thresholds.

- Figure S2. All possible combinations of 3-component CMCs constructed from seven component MCs along three speed lines.

- Table S1. Relative distances in spatial and temporal frequency space between channels across all 6 experimental conditions

- Figure S3. Velocities of the experimental CMC stimuli measured with a computational model of area MT.

- Figure S4*.* Model comparison.

- Figure S5. Weight contour plots for all participants.

- Figure S6. Equivalent implementations of the normalization mechanism.

- Supplemental Experimental Procedures.
 - Motion Clouds stimulus construction.
 - Computational model fitting procedure.

- Supplemental References.

- Supplementary Movie. An example of Compound Motion Cloud stimulus (condition 3) is provided as a supplemental movie file.

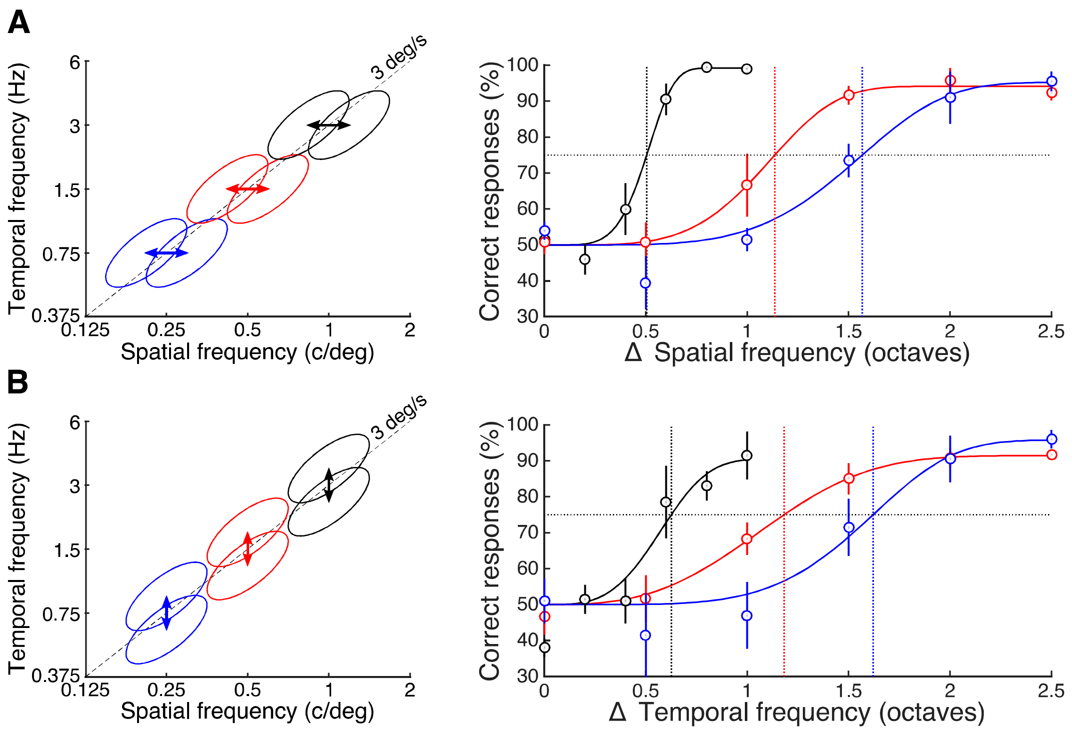


Figure S1. Preliminary experiment for identifying coherency thresholds. In an ABX task, 4 participants were presented with three stimuli in succession and were asked to match the third stimulus (X) with either the first (A) or the second (B). Stimuli A were always composed from two components with the same mean spatial and temporal frequency, whereas the components of stimuli B had diverging mean spatial (A left) or temporal frequencies (B left) at steps of 0.5 (or 0.2 for sf_0 = 1 c/deg) octaves. We plot the percentage of correct responses as a function of the distance between the mean spatial (A right) or temporal frequencies (B right) of the two component stimuli at 3 different mean spatial frequencies: 0.25 c/deg (blue), 0.5 c/deg (red), and 1 c/deg (black). The minimum value across conditions in which participants had 75% accuracy (vertical dotted lines) was chosen as the threshold to generate the CMC stimuli of the main experiment. At 20% stimulus contrast, we found thresholds of 1.5, 1.1, and 0.5 octaves for spatial frequency for mean spatial frequency of 0.25 c/deg, 0.5 c/deg, and 1 c/deg respectively and slightly higher thresholds for temporal frequency. Data are represented as mean ± SEM.

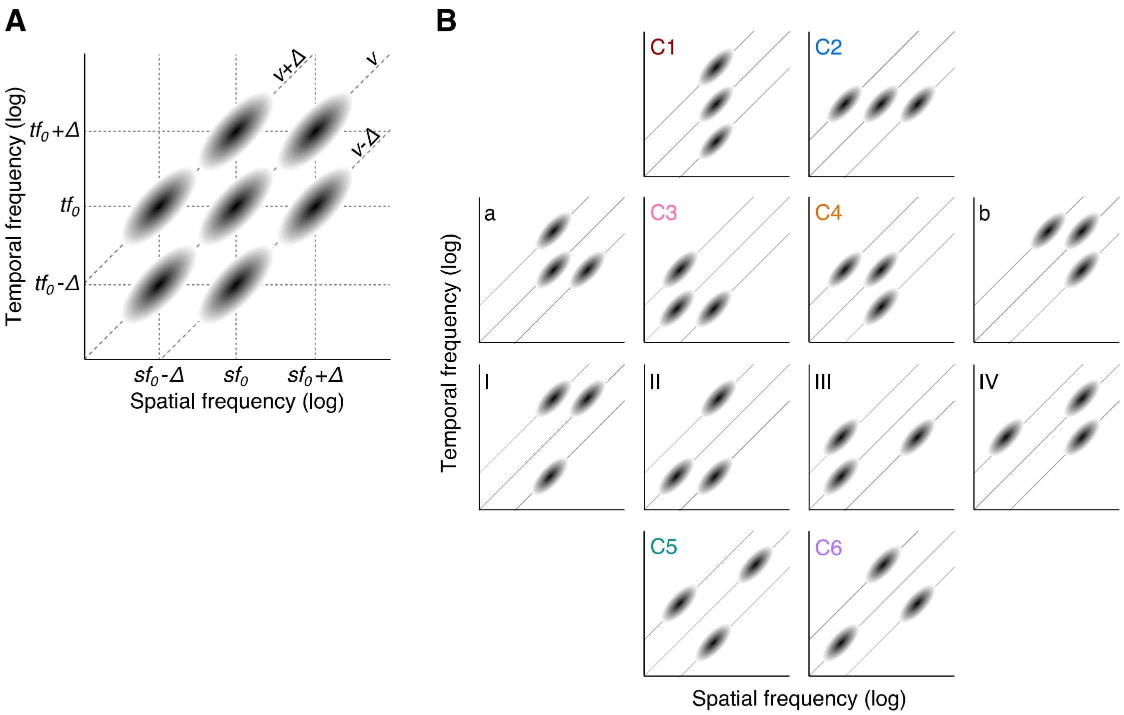


Figure S2. All possible combinations of 3-component CMCs constructed from seven component MCs along three speed lines. (A) Seven component MCs arranged along three speed lines. (B) All 12 possible combinations of the seven components so that the average speed of the CMC is the central speed. There are 14 possible interactions between components (see Table S1). All possible interactions between two components can be captured by our selection of combinations (C1-C6). The unused combinations do not offer any additional information. In particular, combinations a and b are identical to C3 and C4 only transposed to higher temporal and spatial frequencies, while combinations I to IV include interactions that are captured by conditions C1, C2, C5, and C6. Furthermore, combinations I to IV are unbalanced and thus possibly present additional complexities in the interactions between components.

Table S1. Relative distances in spatial and temporal frequency space between channels across all 6 experimental conditions along with the model interaction weights and the applicable conditions.

Spatial frequency distance (*Δ* units)	Temporal frequency distance (*Δ* units)	Weight	Applicable conditions
1	**0**	a_{sf}	2, 3, 4
-1	**0**	$1/a_{sf}$	2, 3, 4
0	**1**	a_{tf}	1, 3, 4
0	**-1**	$1/a_{tf}$	1, 3, 4
2	**0**	b_{sf}	2
-2	**0**	$1/b_{sf}$	2
0	**2**	b_{tf}	1
0	**-2**	$1/b_{tf}$	1
1	**-1**	a_{sf}/a_{tf}	3, 4, 5, 6
-1	**1**	a_{tf}/a_{sf}	3, 4, 5, 6
1	**2**	$a_{sf}*b_{tf}$	5, 6
-1	**-2**	$1/(a_{sf}*b_{tf})$	5, 6
2	**1**	$a_{tf}*b_{sf}$	5, 6
-2	**-1**	$1/(a_{tf}*b_{sf})$	5, 6

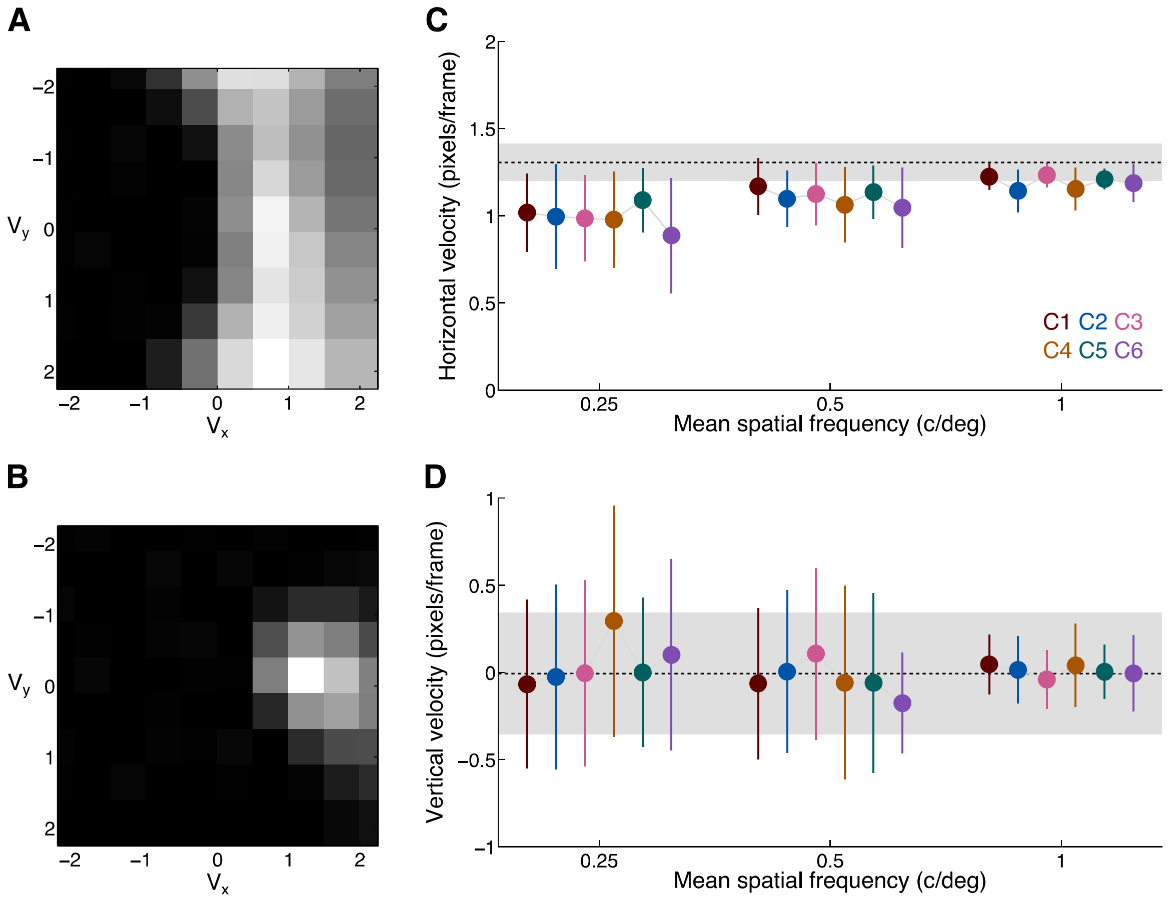


Figure S3. Velocities of the experimental CMC stimuli measured with a computational model of area MT. We used a computational model of areas V1/MT [S1] to measure the response of a population of MT neurons to our CMC experimental stimuli and checked that the reported pattern of speed sensitivities was not explained by differences in the raw motion energy patterns. There were 20 stimuli for each condition and mean spatial frequency (10 moving leftwards and 10 moving rightwards). We calculated the responses of 81 MT neurons (tuned to velocities from -2 to 2 pixels/frame in steps of 0.5 pixels/frame both horizontally and vertically) to each stimulus movie (30 frames). The perceptual estimate for each movie was extracted from the population average: $v_{\text{estimate}} = \Sigma_i v_i r_i \ / \ \Sigma_i r_i$, where v_i is the preferred velocity of neuron i and r_i is its response to the stimulus. An example of model MT neuronal responses to a CMC stimulus is shown in (A) and to a RDK stimulus in (B). Intensity is proportional to a neuron's response. For our stimuli, 3 deg/s translates to 1.023 pixels/frame. The pattern of activation is different between the two types of stimuli because the CMC stimulus has an orientation component that, due to the aperture problem, results in local motion ambiguity and therefore activation along the V_y axis. Estimated horizontal and vertical velocities for each condition and mean spatial frequency are shown in (C) and (D) respectively. Leftward and rightward stimuli are grouped together. Results are represented as mean ± SD. The horizontal dashed line indicates the mean estimate for 20 generated instances of the RDK stimulus moving at 3 deg/s and the gray area indicates ± SD. The model horizontal velocities for each condition do not match the biases observed in the psychophysical data (Figure 2C).

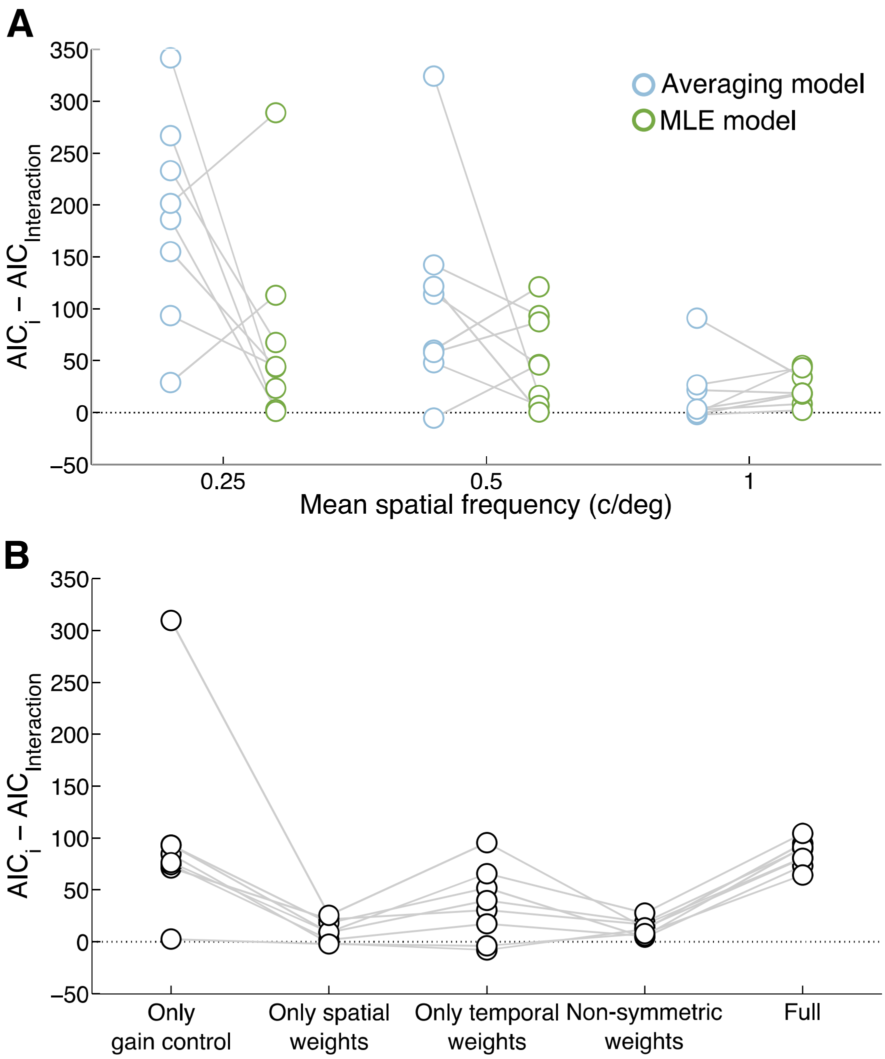


Figure S4. Model comparison. (A) Akaike information criterion values for the *Averaging* and *MLE* models subtracted by the values of the *Interaction* model are shown for all participants at all mean spatial frequencies. Positive values indicate that the Interaction model provides better goodness-of-fit than the other model, and vice versa for negative values. (B) Akaike information criterion values for 5 variations of the *Interaction* model subtracted by the values of the model presented in the main text are shown for all participants. The AIC values are summed for all mean spatial frequencies. The 'Full' model assumes a different weight for each of 14 possible interactions.

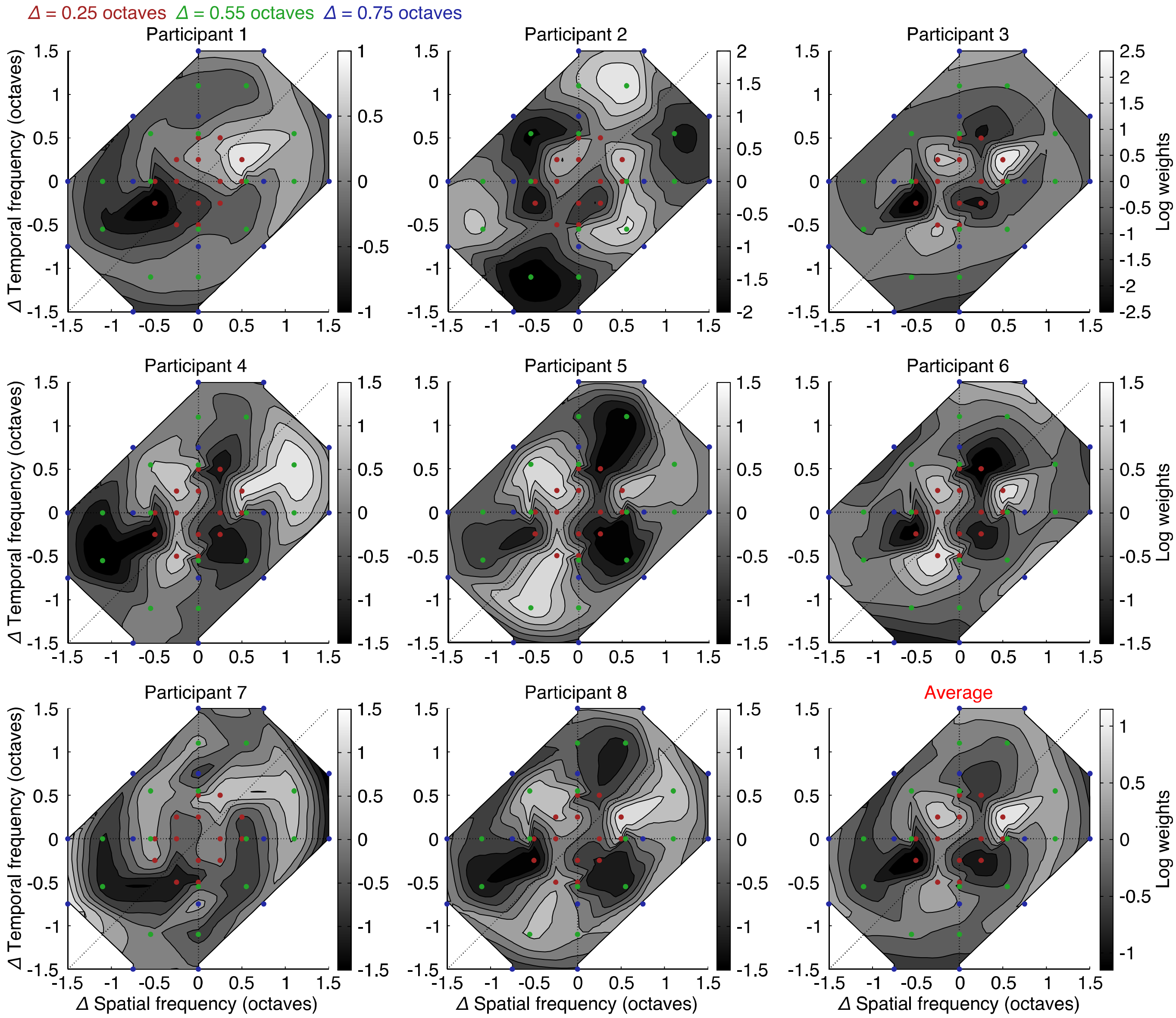


Figure S5. Weight contour plots for all participants. The values indicate how neighboring channels affect the central channel based on their distance in spatial and temporal frequency. Positive values (in light grey) indicate excitation and negative values (in dark grey) inhibition. Colored dots indicate the coordinates in distance space that the model fitted the experimental data. The underlying surfaces were calculated using cubic surface interpolation from the log weights by combining the fits from all 3 mean spatial frequencies. The use of cubic interpolation instead of polynomial interpolation limits potential distortions near the edges of the surface, and this is important because we do not want to make further assumptions about the weights outside the distance values used in the psychophysical experiments. While a few participants' weights have high values even at large distances, these values might be a result of model overfitting, and additional data points further away would be required to make a stronger prediction.

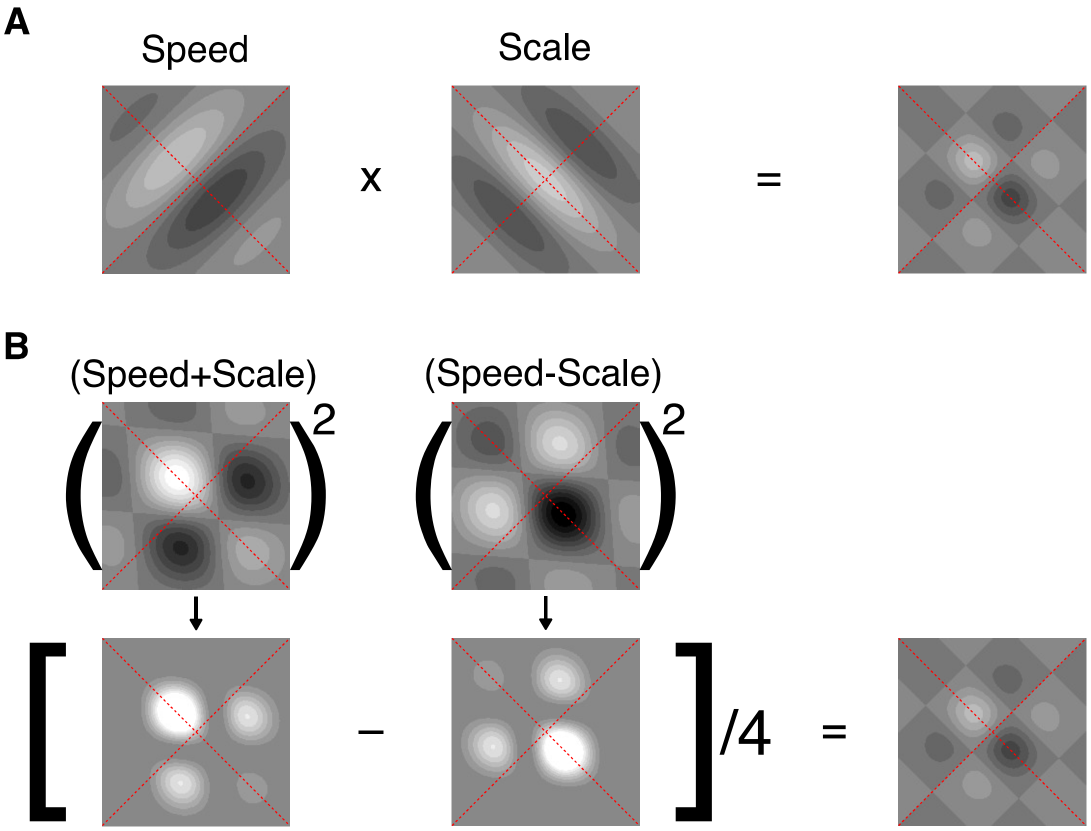


Figure S6. Equivalent implementations of the normalization mechanism. (A) When the two proposed components (speed and scale) are multiplied together, the product approximates the pattern of the average weights of our computational model. (B) A potentially more biologically plausible, and completely equivalent, result can be obtained as a difference of two squared entities. These two entities are the sum of the speed and scale components and their difference.

Supplemental Experimental Procedures

Motion Clouds stimulus construction

It has been shown that natural images can be decomposed into, and therefore be well represented by, a superimposed set of Gabor-like units with a range of positions within an image, orientations, contrasts and scales (or frequencies). This encoding and decoding framework closely resembles the organization of units of sensitivity identified in mammalian primary visual cortex, V1 [S2, S3]. Spatiotemporal scales within such images approximately follow a 1/f distribution, where the amplitude reduces with increasing frequency [S4, S5] and the specific local phase relationships between Gabor units differentiate one image from another [S6]. In the current work, we wanted to tackle the question of how speed might be computed for dynamic scenes rich with superimposed naturalistic objects.

Motion clouds (MCs) were proposed to study dynamic integrative processes within the visual system under natural stimulation. They serve as a generative model of naturalistic images in which the V1 inspired basis set of localized drifting Gabor elements G_i of critical characteristics C_i (defining its orientation θ_i, spatial frequency sf_i, and temporal frequency tf_i) are linearly combined with randomized phases ϕ_i to remove specific object information. An envelope distribution E is then applied to the elements in Fourier space to constrain the overall stimulus and define the spatiotemporal ellipses illustrated in Figures 1A. The result of this dense mixing is dynamic band-pass filtered luminance noise stimuli in which distribution parameters can be well controlled [S7-S9]. The MCs are fully characterized by a small set of parametric vectors M and U over orientation, spatial frequency, and speed

$$M = [\theta_o, sf_o, v_o], \qquad \text{(Equation S1)}$$

$$U = [\Delta\theta, Bsf, Bv], \qquad \text{(Equation S2)}$$

where *M* determines the central characteristics of these dimensions, and *U* their spread. The envelope distribution E is then defined as

$$E = P[M, U], \qquad \text{(Equation S3)}$$

that is as a probability distribution function centered on *M* and with spread *U*. Orientation is defined as a von Mises distribution, while both spatial frequency and speed are defined as Log Normal distributions. Sampling this envelope distribution gives us a set of Gabor characteristics

$$C_i = [\theta_i, sf_i, tf_i], \qquad \text{(Equation S4)}$$

where the temporal frequency is simply computed from the sampled spatial frequency and speed as

$$tf_i = \ sf_i \cdot v_i. \qquad \text{(Equation S5)}$$

A large, finite number *N* of vector elements G_i are defined with the characteristics C_i centered on the location p_i [p_x,p_y] which is uniformly distributed over the size of the image. The phase of each element, ϕ_i is also uniformly distributed over [0,2π]. Each element is then scaled by an amplitude a_i to control contrast, and then summed to define the MC as the luminance I at each spatial location (*x, y*) and time *t*

$$I(x, y, t) = \sum_{i=1}^{N} a_i G_i(p_i, \phi_i - t, C_i). \qquad \text{(Equation S6)}$$

More detailed descriptions of these stimuli, as well as example implementations, can be found in previous work [S8, S9].

Notably, MCs differ from gratings because they have a distributed frequency in Fourier space rather than a point (see Figure 1A). In addition, and importantly, unlike a very sparse array of Gabors, MC elements cannot be perceptually segregated. During experiments, MCs used have their contrast energy controlled by fixing their RMS contrast which closely matches perception. CMCs are created as a linear superposition of MCs intended to simulate complex composite scenes with each MC serving as a complex object within it.

Computational model fitting procedure

Using the same speed discrimination procedure as for the CMCs (see Experimental Procedures in the main text), we measured the psychometric functions for each of the seven individual components across all mean spatial frequencies. Naturally, the six speeds of the RDK were transposed to higher or lower values for the faster or slower components. From the psychometric functions, we build the speed likelihood for each component by taking the derivative of the cumulative distribution function, and we construct a combined likelihood for each of the 6 CMC. We use the combined likelihood to create a receiver operating characteristic (ROC) curve for each of the 6 RDK speeds. This ROC is obtained by taking an arbitrary criterion, computing the probabilities that the CMC and the RDK exceed this criterion (which is one point on the ROC curve), and repeating the procedure for other criteria. The probability *p*(RDK > CMC) that the RDK is perceived faster than the CMC is given by the area under the curve (AUC). We calculated the AUC of each ROC curve, and we reconstructed the psychometric function for each CMC. This procedure is used across all models.

The *Averaging* model uses a simple average of the component likelihoods and has no free parameters. The *Maximum-likelihood estimation* (MLE) model uses a simple product of the component likelihoods and has no free parameters either.

The *Interaction* model presented in the main text has 5 free parameters, 1 for gain control and 4 for the interaction matrix (a_{sf}, b_{sf}, a_{tf}, and b_{tf}). The values of the weights indicate the effect that channels have on each other, and they can be positive (excitation) or negative (inhibition). A potential alternative way to model the interaction between components would be to assume a uniform level of facilitation between all components and only selective inhibition between some of them [S10].

For each participant and mean spatial frequency, 5 free parameters are fitted to 36 data points (6 RDK speeds times 6 conditions). Different variations of the Interaction model have varying number of free parameters based on the number of interaction weights (Only gain control: 1, Only spatial weights: 3, Only temporal weights: 3, Non-symmetric weights: 9, and Full: 15).

Supplemental References

S1. Simoncelli, E.P., & Heeger, D.J. (1998). A model of neuronal responses in visual area MT. Vision Res. *38*, 743-761.

S2. Berkes, P., Turner, R.E., & Sahani, M. (2009). A Structured Model of Video Reproduces Primary Visual Cortical Organisation. PLoS Comput. Biol. *5*. doi:10.1371/journal.pcbi.1000495.

S3. van Hateren, J.H., & Ruderman, D.L. (1998). Independent component analysis of natural image sequences yields spatio-temporal filters similar to simple cells in primary visual cortex. Proc. R. Soc. Lond. B *265*, 2315-2320.

S4. Field, D.J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. J. Opt. Soc. Am. A *4*, 2379-2394.

S5. Dong, D.W., & Atick, J.J. (1995). Statistics of natural time-varying images. Netw. *6*, 345-358.

S6. Yoonessi, A., & Kingdom, F. A. A. (2008). Comparison of sensitivity to color changes in natural and phase-scrambled scenes. J. Opt. Soc. Am. A *25*, 676-684.

S7. Simoncini, C., Perrinet, L.U., Montagnini, A., Mamassian, P., and Masson, G.S. (2012). More is not always better: adaptive gain control explains dissociation between perception and action. Nat. Neurosci. 15, 1596–1603.

S8. Leon, P.S., Vanzetta, I., Masson, G.S., and Perrinet, L.U. (2012). Motion clouds: model-based stimulus synthesis of natural-like random textures for the study of motion perception. J. Neurophysiol. *107*, 3217–3226.

S9. Vacher, J., Meso, A.I., Perrinet, L.U., and Peyre, G. (2015). Biologically inspired dynamic textures for probing motion perception. In Advances in Neural Information Processing Systems 28, C. Cortes, N.D. Lawrence, D.D. Lee, M. Sugiyama, and R. Garnett, eds., pp. 1918–1926.

S10. Meese, T.S., & Holmes, D.J. (2007). Spatial and temporal dependencies of cross-orientation suppression in human vision. Proc. R. Soc. B *274*, 127-136.

Supplemental Movies for Condition 3

Click here to access/download

Supplemental Movies & Spreadsheets

CMC_cond3.mp4