

# Spatiotemporal filtering and motion illusions

**Arezoo Pooresmaeili**

Berlin School of Mind and Brain, Berlin, Germany

**Guido Marco Cicchini**

Institute of Neurosciences, National Research Council,  
Pisa, Italy

**Maria Concetta Morrone**

Department of Physiological Sciences, Università di Pisa,  
Pisa, Italy  
Scientific Institute Stella Maris, Pisa, Italy

**David Burr**

Institute of Neurosciences, National Research Council,  
Pisa, Italy  
Università degli Studi di Firenze, Florence, Italy



We are perplexed by Clarke et al.'s (2013) criticisms on our recent contribution to *Journal of Vision* (Pooresmaeili, Cicchini, Morrone, & Burr, 2012). Our group has long championed the idea that perceptual processing of information can be anchored in a dynamic coordinate system that need not correspond to the instantaneous retinal representation. Our recent evidence shows that temporal duration (Burr, Tozzi, & Morrone, 2007; Morrone, Cicchini, & Burr, 2010), orientation (Zimmermann, Morrone, Fink, & Burr, 2013), motion (Melcher & Morrone, 2003; Turi & Burr, 2012) and saccadic error-correction (Zimmermann, Burr, & Morrone, 2011) are all processed to some extent in spatiotopic coordinates. Imaging studies reinforce these studies (d'Avossa et al., 2007; Crespi et al., 2011). Much earlier, we showed that the processing of smoothly moving objects was not anchored in instantaneous, retinotopic coordinates, but in the reference frame given by the trajectory of motion. There is an effective interpolation along the trajectory, so temporal offsets in spatially collinear stimuli causes them to appear spatially offset, corresponding to the physical reality of stimuli moving over large regions of space, behind occluders (Burr, 1979; Burr & Ross, 1979). Our explanation for this surprising effect was that it could be a direct consequence of the spatiotemporal orientation of the impulsive response of motion detectors, providing the spatiotemporal reference frame needed to account for the interactions between time and space (Burr & Ross, 1986; Burr, Ross, & Morrone, 1986; Burr & Ross, 2004; Nishida, 2004). Recently, we have applied the concept of

spatiotemporal oriented receptive fields to account for “predictive remapping,” the “nonretinotopic” effects that occur on each saccadic eye-movement (Burr & Morrone, 2010; Burr & Morrone, 2012; Cicchini, Binda, Burr, & Morrone, 2012).

We were most impressed by the compelling demonstrations of Herzog's group, clearly showing that the reference frame of processing is not the instantaneous retinal position, but is flexible, depending not only on real physical motion, but on an illusory apparent motion where the stimuli do not actually move (Boi, Ogmen, Krummenacher, Otto, & Herzog, 2009). This seemed to us important, worthy of quantitative measurement and modeling, particularly to see whether these new effects may fall within the framework that so successfully explained previous demonstrations, such as spatiotemporal interpolation.

It is reassuring that Clarke et al. (2013) confirm our results, albeit with some variability between subjects. But more importantly add a very nice result in showing that our simplified version of the “litmus test” can be enhanced by attending to the motion. This is an excellent point that we overlooked. The strength of this type of motion is well known to depend on attention (Cavanagh, 1992), and it is indeed interesting that the strength of motion-induced effects depends not only on the physical conditions, but on internal states such as attention. Perhaps attention may also provide the flexibility in choosing the most appropriate scale for analysis, which in this case would be lower, given that attention is diverted to the periphery. This would add strength to our model, and an idea worth following up,

Citation: Pooresmaeili, A., Cicchini, G. M., Morrone, M. C., & Burr, D. (2013). Spatiotemporal filtering and motion illusions. *Journal of Vision*, 13(10):21, 1–4, <http://www.journalofvision.org/content/13/10/21>, doi:10.1167/13.10.21.

given that attention has been implicated in another form of nonretinotopic processing: the BOLD response of areas of the dorsal visual system, clearly spatiotopic under free-viewing conditions, becomes retinotopic when attention is diverted from the moving stimulus to the retinotopically stable fovea (Gardner, Merriam, Movshon, & Heeger, 2008; Crespi et al., 2011).

However, we find it odd that Clarke et al. refer to our approach as “indirect:” it is formally equivalent to theirs. They show that the apparent motion of dots can change when the frame on which they are displayed itself moves, causing motion frames to pair with stimuli that are on the motion trajectory but nonaligned retinotopically. We show that a grating can appear to drift in a specific direction by causing the motion frames to pair with stimuli that are on the motion trajectory, but nonaligned retinotopically. It is important to note that there was no directed motion in retinotopic coordinates, confirmed by the fact that the two-bar stimulus showed no bias. The only originality in our stimulus is that it lends itself to measure the magnitude of the effect psychophysically, by annulling either phase or contrast. While methods could have been devised to measure psychophysically the strength of the rotating dot, we believed that the more traditional grating, fully described by phase, frequency and contrast, was more amenable to rigorous control. Obviously, any result obtained with a drifting stimulus can be generalized to rotation, by a simple polar transform.

Perhaps we failed to communicate our very simple message. We believe that “nonretinotopic” processing can be described as a dynamic change in the coordinate system of analysis (the result of the slant of receptive fields in space-time), where successive motion frames are not aligned in retinal coordinates, but along the trajectory of the apparent motion created by the Ternus display. The simplest way of imposing this dynamic reference frame is to apply a motion detector with impulse response slanted in space-time, which is what we did in our model. However, it turned out that this was not sufficient: a nonlinear stage was necessary to compute motion energy. Although we used specific parameters in our simulation (correctly reported and well justified in our paper), the modeling was robust, and worked for a range of parameters. We do not understand in what sense this is not falsifiable: it would be sufficient to devise a motion-based, nonretinotopic effect that could not be explained by spatiotemporal filtering or slanted spatiotemporal receptive fields.

Clarke et al. report many objections to our experimentation and modeling. Unfortunately, many of their complaints were highly subjective, such as the “indirectness” and “unsuitability” of our psychophysical methods (standard, off-the-shelf annulment tech-

niques), or the nebulous criticisms of our (noncrucial) choice of parameters for our toy model. Some statements are simply untrue: such as the width of the receptive field used for the simulation (correctly stated in the paper as  $1^\circ$  half-width at 0.37 height), the number of nonlinearities, and many similar errors. They claim that our filter selection is ad-hoc, overlooking the fact, obvious in their figure 3, that the upwards bias in the three-bar condition occurs in 8 out of 12 filters, and also holds if we chose the same filter for both two and three bars. Our filter selection was to maximize the fit to the behavioral data, while keeping the model as simple and as biologically plausible as possible, but was not crucial for the reproduction of the main effects. Neurons have receptive fields with specific sizes across the visual hierarchy, so assuming that the most appropriate filters perform the spatiotemporal integration is not unreasonable, and quite standard for the most modeling and simulations. To respond to the other purported discrepancies, the scrupulous reader is referred to our original paper, where all details are correctly reported.

The Herzog group (Boi et al., 2009; Clarke et al., 2013) favor a multistage model. We ourselves have no strong views on how many stages are necessary for the analysis. We used a standard motion-energy model with local-maximum selection, with minimal free parameters (Adelson & Bergen, 1985; Santen & Sperling, 1985; Heeger, 1987; Yuille & Grzywacz, 1988; Del Viva & Morrone, 1992). Certainly, it may be conceptually useful to separate the global Ternus/Pikler motion from the vertical “nonretinotopic” motion, but it is difficult to assert whether this necessarily occurs in functionally different stages. This is possible, given the number of distinct areas that have been identified for motion processing, but we prefer not to speculate at this stage.

In conclusion, we simply do not see the point to Clarke et al.’s commentary. We believe that the line of research that Herzog’s group has recently revived is important, well worthy of further study. We also believe it important to try to inter-relate different experimental approaches, to search for a common language to relate findings from the various research traditions. It was in this spirit that we used the standard spatiotemporal filter approach to model Boi et al.’s clever demonstration, after minimal modification to render the two approaches compatible. Only the test of time will validate the usefulness of our approach, either by providing helpful insights or by stimulating further research along this fascinating line. We trust that readers of this journal will consider the utility or otherwise of our contribution with the objectivity and openmindedness that characterizes good science.

## Acknowledgments

This work was supported by the European Research Council (FP7 – “STANIB”) and the Italian Ministry of Research.

Commercial relationships: none.

Corresponding author: David Burr.

Email: dave@in.cnr.it.

Address: Institute of Neurosciences, National Research Council, Pisa, Italy, and Università degli Studi di Firenze, Florence, Italy.

## References

- Adelson, E. H., & Bergen, J. R. (1985). Spatio-temporal energy models for the perception of motion. *Journal of the Optical Society of America, A2*, 284–299.
- Boi, M., Ogmen, H., Krummenacher, J., Otto, T. U., & Herzog, M. H. (2009). A (fascinating) litmus test for human retino- vs. non-retinotopic processing. *Journal of Vision, 9*(13):5, 1–11, <http://www.journalofvision.org/content/9/13/5>, doi:10.1167/9.13.5. [PubMed] [Article]
- Burr, D., & Ross, J. (2004). Vision: The world through picket fences. *Current Biology, 14*(10), R381–382.
- Burr, D. C. (1979). Acuity for apparent vernier offset. *Vision Research, 19*(7), 835–837. doi:0042-6989(79)90162-7 [pii].
- Burr, D. C., & Morrone, M. C. (2010). Spatiotopic coding and remapping in humans. *Philosophical Transactions of the Royal Society A, 366*(1564), 504–515.
- Burr, D. C., & Morrone, M. C. (2012). Constructing stable spatial maps of the world. *Perception, 41*(11), 1355–1372.
- Burr, D. C., & Ross, J. (1979). How does binocular delay give information about depth? *Vision Research, 19*, 523–532.
- Burr, D. C., & Ross, J. (1986). Visual processing of motion. *Trends in Neuroscience, 9*, 304–306.
- Burr, D. C., Ross, J., & Morrone, M. C. (1986). Seeing objects in motion. *Proceedings of the Royal Society (London), B227*, 249–265.
- Burr, D. C., Tozzi, A., & Morrone, M. C. (2007). Neural mechanisms for timing visual events are spatially selective in real-world coordinates. *Nature Neuroscience, 10*(4), 423–425.
- Cavanagh, P. (1992). Attention-based motion perception. *Science, 257*(5076), 1563–1565.
- Cicchini, G. M., Binda, P., Burr, D. C., & Morrone, M. C. (2012). Transient spatiotopic integration across saccadic eye-movements mediates visual stability. *Journal of Neurophysiology*, doi: jn.00478.2012 [pii] 10.1152/jn.00478.2012.
- Clarke, A. M., Repnow, M., Ögmen, H., & Herzog, M. H. (2013). Does spatio-temporal filtering account for nonretinotopic motion perception? Comment on Pooresmaeili, Cicchini, Morrone, and Burr (2012). *Journal of Vision, 13*(10):20, 1–14, <http://www.journalofvision.org/content/13/10/20>, doi:10.1167/13.10.20. [Article]
- Crespi, S., Biagi, L., d’Avossa, G., Burr, D. C., Tosetti, M., & Morrone, M. C. (2011). Spatiotopic coding of bold signal in human visual cortex depends on spatial attention. *PLoS One, 6*(7), e21661. doi:10.1371/journal.pone.0021661 PONE-D-11-02036 [pii].
- d’Avossa, G., Tosetti, M., Crespi, S., Biagi, L., Burr, D. C., & Morrone, M. C. (2007). Spatiotopic selectivity of bold responses to visual motion in human area MT. *Nature Neuroscience, 10*(2), 249–255.
- Del Viva, M., & Morrone, M. C. (1992). Feature detection and non-Fourier motion. *Perception, 21*(Supp 2), 41.
- Gardner, J. L., Merriam, E. P., Movshon, J. A., & Heeger, D. J. (2008). Maps of visual space in human occipital cortex are retinotopic, not spatiotopic. *Journal of Neuroscience, 28*(15), 3988–3999.
- Heeger, D. J. (1987). Model for the extraction of image flow. *Journal of the Optical Society of America, 4A*, 1455–1471.
- Melcher, D., & Morrone, M. C. (2003). Spatiotopic temporal integration of visual motion across saccadic eye movements. *Nature Neuroscience, 6*(8), 877–881.
- Morrone, M. C., Cicchini, M., & Burr, D. C. (2010). Spatial maps for time and motion. *Experimental Brain Research, 206*, 121–128.
- Nishida, D. (2004). Motion-based analysis of spatial patterns by the human visual system. *Current Biology, 14*(10), 830–839, doi:10.1016/j.cub.2004.04.044.
- Pooresmaeili, A., Cicchini, G., Morrone, M., & Burr, D. (2012). “Non-retinotopic processing” in Ternus motion displays modeled by spatiotemporal filters. *Journal of Vision, 12*(1):10, 1–15, <http://www.journalofvision.org/content/12/1/10>, doi:10.1167/12.1.10. [PubMed] [Article]

- Santen, J. P. H. v., & Sperling, G. (1985). Elaborated Reichardt detectors. *Journal of the Optical Society of America, A2*, 300–321.
- Turi, M., & Burr, D. (2012). Spatiotopic perceptual maps in humans: Evidence from motion adaptation. *Proceedings of the Royal Society B: Biological Sciences*, 279(1740), 3091–3097, doi:10.1098/rspb.2012.0637 rspb.2012.0637 [pii].
- Yuille, A. L., & Grzywacz, N. M. (1988). A computational theory for the perception of coherent visual motion. *Nature*, 333, 71–74.
- Zimmermann, E., Burr, D., & Morrone, M. C. (2011). Spatiotopic visual maps revealed by saccadic adaptation in humans. *Current Biology*, 21(16), 1380–1384, doi: S0960-9822(11)00659-2 [pii] 10.1016/j.cub.2011.06.014.
- Zimmermann, E., Morrone, M. C., Fink, G. R., & Burr, D. (2013). Spatiotopic neural representations develop slowly across saccades. *Current Biology*, 23(5), R193–194, doi:10.1016/j.cub.2013.01.065 S0960-9822(13)00134-6 [pii].