Computational modeling of non-Fourier motion: further evidence for a single luminance-based mechanism

Christopher P. Benton and Alan Johnston

Department of Psychology, University College London, Gower Street, London, WC1E 6BT, UK

Peter W. McOwan

Department of Computer Science, Queen Mary, University of London, Mile End Road, London, E1 4NS, UK

Jonathan D. Victor

Department of Neurology and Neuroscience, Cornell University Medical College, 1300 York Avenue, New York, New York 10021

Received November 27, 2000; revised manuscript received March 5, 2001; accepted April 19, 2001

It is generally assumed that the perception of non-Fourier motion requires the operation of some nonlinearity before motion analysis. We apply a computational model of biological motion processing to a class of non-Fourier motion stimuli designed to investigate nonlinearity in human visual processing. The model correctly detects direction of motion in these non-Fourier stimuli without recourse to any preprocessing nonlinearity. This demonstrates that the non-Fourier motion in some non-Fourier stimuli is directly available to luminance-based motion mechanisms operating on measurements of local spatial and temporal gradients. © 2001 Optical Society of America

OCIS codes: 000.1430, 000.4920.

1. INTRODUCTION

Standard computational approaches to biological low-level motion perception are based on motion energy analysis,[1–3] autocorrelation,[4–6] and spatiotemporal gradient techniques.[7,8] Non-Fourier motion is generally perceived to be opaque to standard computational techniques[9,10] but can, however, be perceived by human observers. This has been taken to imply that some additional motion mechanism must underlie perception of non-Fourier motion.

The general computational approach to non-Fourier motion has been to propose that some nonlinearity is applied to the stimulus before motion analysis. Coupled with linear filtering operations, this can make non-Fourier motion detectable by standard computational techniques.[9] Many non-Fourier stimuli can be thought of as comprising an envelope and a carrier. The effect of the nonlinear transformation is to demodulate the signal, effectively translating non-Fourier motion into Fourier motion. Much of the debate surrounding non-Fourier motion has focused around the issue of whether we have a specialized non-Fourier mechanism that deliberately contains a substantial early nonlinearity (the two-channel hypothesis) or whether we have a single low-level motion mechanism that includes some early nonlinearity (the distortion-product hypothesis).

One example of a non-Fourier stimulus can be created by multiplying a static binary noise pattern by a translating sinusoid,

$$I(x, y, t) = I_0[1 + \cos(ky + \omega t)R(x, y)], \qquad (1)$$

where $I(x, y, t)$ is luminance at point (x, y, t), I_0 is mean luminance, k is spatial frequency, ω is temporal frequency, and $R(x, y)$ is a static binary noise pattern with an expected mean of zero. Such a stimulus may also be thought as resulting from the sampling of two sine-wave gratings with identical spatial and temporal frequencies, separated by half a cycle from each other. Figure 1(a) shows a slice through a binary noise pattern $R(x, y)$, and Fig. 1(b) shows two sinusoids separated by half a cycle from each other. In Fig. 1 the two sinusoids are sampled by the noise pattern. If the value of the noise is $+1$, then the dotted sinusoid is chosen; if the noise has a value of -1, then the sinusoid indicated by the solid curve is chosen.

What is interesting about thinking in terms of sampling sine waves separated by phase differences is that we can start to consider what happens when, for example, we sample three sine waves separated from one another by one third of a cycle. In this case the noise that determines which sine wave occupies which particular spatial position does not contain just two values but contains three values each with a one third chance of occurring. Similarly, one can create stimuli made from sampling four, five or n sine waves. These sine waves are separated by $2\pi/n$ rad, and the noise used to sample them has n possible values, each with an equal chance of occurring. This family of stimuli was developed and described by Taub *et al.*,[11] who denote a stimulus created from sam-

0740-3232/2001/092204-05$15.00 © 2001 Optical Society of America

pling n sine waves as P_n. For $n \geq 2$ these are microbalanced non-Fourier stimuli. This means that for any space–time-separable function that is applied to the stimuli, they remain balanced in terms of their expected motion energy.[9]

Each static noise element samples a sinusoid that has the same amplitude and temporal frequency as that sampled by any other noise element. This means that the temporal profile for any point in the image is sinusoidal and has the same amplitude and temporal frequency as the temporal profile of any other point in the image. There is therefore only one temporal frequency present in each stimulus. Consequently, there are no flicker cues to motion and the stimuli are unaffected (except by attenuation) by temporal filtering operations. Additionally, for $n = 1$ the stimulus is a luminance modulation, and for $n = 2$ the stimulus is a contrast modulation, however, for $n > 2$, the motion of the stimulus is defined neither by a change in luminance nor by a change in local rms contrast.

Taub *et al.*[11] show that for a stimulus of order n, a polynomial nonlinearity with order greater than or equal to n is necessary to introduce a distortion product into the image. Taub *et al.* measured direction-discrimination thresholds in the presence and absence of noise. They found a clear trend over order of stimulus with the most detectable stimulus being P_1 and the least detectable being P_4. For stimuli of order 5 and above, direction-discrimination thresholds could not be obtained. Taub *et al.* propose that the distortion-product hypothesis can account for the perception of motion elicited by their stimuli. On the basis of their psychophysical findings they derive a nonlinear function that can account for the data.

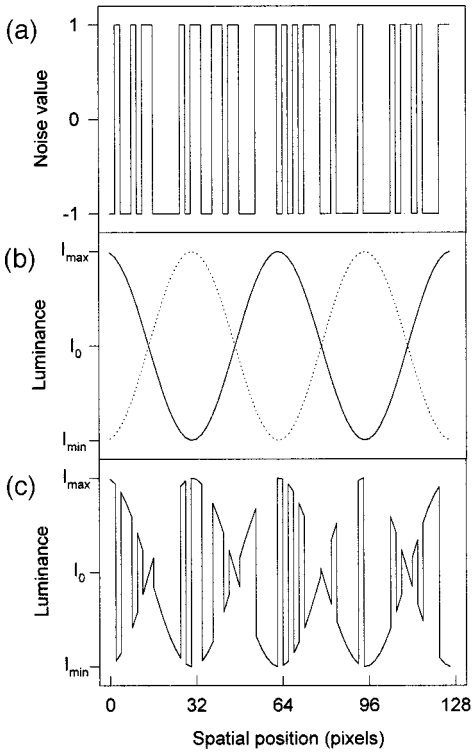

Fig. 1. (a) Binary noise pattern. (b) Two sine waves separated by half a cycle. (c) Snapshot of a contrast modulation of noise created by sampling the two sine waves shown in (b) with the noise sample shown in (a). When the value of the noise is 1, then the dotted sine wave is selected; when the value of the noise is -1, then the sine wave indicated by the solid curve is selected. Note that in the example shown here there is spatial variation within noise elements. However, for the stimuli used in this study there was no spatial variation within noise elements.

The analysis that Taub *et al.*[11] present is firmly rooted in the notion that some early nonlinearity underlies the perception of motion in their class of non-Fourier stimuli. In this paper we demonstrate that this early nonlinearity is not a necessity and that a gradient-based computational approach can detect non-Fourier motion.

2. COMPUTATIONAL MODELING

We apply a computational model of biological motion processing to a number of non-Fourier stimuli. The model is described in detail elsewhere[8,12] and is an extension of the gradient-model approach in which image velocity is calculated by taking the ratio of the local temporal gradient to the local spatial gradient.[7,13,14] A central design feature of the model is that the estimation of velocity can be stabilized by increasing the number of local gradient measures included in the calculation. To this end the model integrates the outputs of a number of spatial frequency, temporal frequency and orientation-tuned filters to produce its final estimate of image velocity. The filters used in the model are derived by differentiating a single blur kernel,

$$K(r,t) = \frac{1}{4\pi\sigma}\exp(-r^2/4\sigma)\frac{1}{\sqrt{\pi}\tau\alpha\exp(\tau^2/4)} \times \exp\{-[\ln(t/\alpha)/\tau]^2\}, \quad (2)$$

where t is time, r is radial distance from the center of the filter, $\sigma = 1.5$, $\alpha = 10$, and $\tau = 0.275$. In the case of the spatial parameter (σ), this value is given in pixels; in the case of the temporal parameters (α and τ), values are given in number of frames. 128 pixels is equivalent to 1°, and 128 frames is equivalent to 1 s. The parameters were fixed in previous work.[15]

Following Taub *et al.*,[11] our stimuli are of the form

$$I(x,y,t) = I_0[1 + P_n(x,y,t)], \quad (3)$$

where

$$P_n(x,y,t) = \cos[ky + \omega t + (2\pi/n)R(x,y)]. \quad (4)$$

$R(x,y)$ is a random function that takes a value between 0 and $n - 1$ (inclusive) at each position (x,y), and I_0 signifies mean luminance. The final term in Eq. (4) simply adds a phase shift to the underlying sine wave $[\cos(ky + wt)]$ and is equivalent to the sampling of sine waves with various phase separations in a manner that is dependent on an underlying static noise sample (see Fig. 1).

Input stimuli comprised 128 frames of 128 by 128 pixel images. Spatial frequency k was 1 cycle/deg, and temporal frequency ω was 4 Hz. Direction of motion was upward. Noise check size was 16 pixels horizontally and 4 pixels vertically; there was no spatial variation within

noise elements. These values provide a close fit to the stimuli investigated by Taub *et al.*[11] Note that the pattern of results described below is robust and has been replicated across a number of envelope spatial and temporal frequencies and across a number of noise-element sizes.

Output images measured 106 × 106 pixels (spatiotemporal filters were 23 × 23 pixels). For each sequence we calculated only one frame of model output. The model produces speed and direction, which can be expressed as a vector that has a vertical component of velocity and a horizontal component of velocity. We discard the horizontal component, as this indicates a velocity component orthogonal to stimulus motion. We calculate a direction index by taking the sum of the velocity components in the upward direction ($V_\uparrow$) and the sum of components in the downward direction ($V_\downarrow$) and then by taking the following measure of contrast between the two:

$$\text{Directional index} = (|V_\uparrow| - |V_\downarrow|)/(|V_\uparrow| + |V_\downarrow|). \quad (5)$$

The directional index ranges between a maximum of +1, indicating correct stimulus motion, and a minimum of −1, indicating reversed motion. A directional index close to zero indicates no overall predominance of either forward or reversed motion.

Figure 2 shows the results of our simulations. Each point and its associated error bar are drawn from 100 directional indexes, each calculated over a sample frame, and indicate the mean and standard deviations of those measures. Each instantiation of a stimulus contained a fresh noise sample, and the starting phase of the underlying sinusoid was randomized. Our results show that the model detects motion in the correct direction and that as n increases, the strength with which the model output signals this direction is reduced. Psychophysical results show that motion is detected correctly for stimuli from P_1 to P_4 and that motion direction thresholds increase as n increases.[11] The model therefore provides a qualitative account of the psychophysical data.

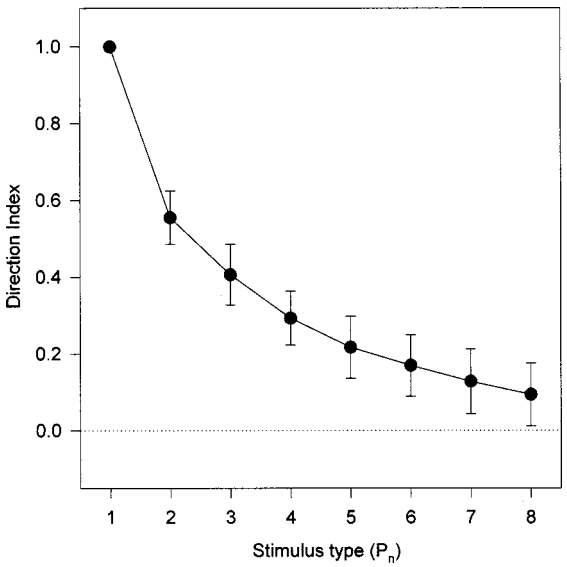

Fig. 2. Mean directional index as a function of stimulus type [see Eq. (5)]. Error bars show standard deviations. The dotted horizontal line marks a directional index of zero, where no overall motion is indicated by the model.

In the directional index [see Eq. (5)] the numerator is a measure of directionality and the denominator is a measure of the total amount of motion detected in the stimulus. The reduction in direction index over increases in n that we see in Fig. 2 is primarily the result of a decrease in the numerator. As n increases there is also a decrease in the denominator that, taken alone, would lead to an increase in the directional index. The reduction of magnitude of the numerator is by far the larger effect. We also examined an additional measure of directionality in the model output. In this we simply counted the number of instances of upward motion and the number of instances of downward motion. A simple difference between these two measures gives an alternative measure of directionality. When plotted against order, this followed a pattern similar to that shown in Fig. 2.

Quantitatively, the decline in directional index over increases in n is more gradual than would be expected if directional index were directly proportional to the reciprocal of the direction-discrimination threshold. There should, however, be no strong expectation of a direct mathematical relationship between our directional indexes and psychophysical direction-discrimination thresholds. A direct mapping of model output onto contrast threshold data would require the implementation of a theory of threshold determination that is beyond the scope of the current paper. Directional index is a measure of the directionality in the output of our computational model. As the directional index approaches zero, there is less predominance of motion in any one particular direction in the model output. This clearly implies that the net direction of motion in the stimulus should be less easily resolved. We might therefore reasonably expect that as the directional index nears zero, discrimination thresholds should increase. This would result in the qualitative agreement that we find between modeling and psychophysical data. More important, at least in terms of the issues addressed in this paper, the results of our simulations indicate that non-Fourier motion can be detected in these microbalanced stimuli without recourse to an early nonlinearity before motion processing.

3. DISCUSSION

Taub *et al.*[11] introduced a class of non-Fourier stimuli created by randomly sampling translating sinusoids separated from one another by $2\pi/n$ rad, where n is the number of sampled sinusoids. These stimuli are referred to as P_n, where P_1 is a single translating sinusoid, a Fourier motion stimulus, and P_2 is a contrast modulation of noise, a non-Fourier motion stimulus. For all values of n larger than 1, the stimuli are non-Fourier. As n increases, the perception of coherent motion that the stimuli elicits becomes weaker. By P_5, the direction of non-Fourier motion cannot be reliably determined. We show that a single unified computational approach can detect both Fourier and non-Fourier motion in these stimuli. In common with psychophysical findings, as n increases, the strength with which the model signals the correct direction of motion (the directional index) is reduced. The findings demonstrate that the information for non-Fourier motion is present in the local spatial and

temporal luminance gradients within the images that we examine. The non-Fourier motion in these stimuli can be extracted without recourse to a nonlinear preprocessing stage.

The extended gradient model that we employ in this study combines the outputs of a number of spatiotemporal filters to calculate velocity.[8,12] From the perspective of Fourier-based image analysis, one might well consider that these form different spatiotemporal channels and that the model is therefore a multiple-channel model. However, this observation applies to all major classes of low-level biological motion models. Even a simple Reichardt detector must utilize two temporal filters and therefore contains multiple temporal channels. In the context of non-Fourier motion, we characterize the model as a single-channel model because it applies a single computational strategy that does not differentiate between Fourier and non-Fourier motion. Critically, there is no separate non-Fourier channel within the model.

Clearly, the model utilized in this study employs a number of nonlinearities to calculate local image motion. Again, this observation is true of other classes of motion models. Reichardt detectors use a multiplicative nonlinearity to combine the outputs of channels with different temporal filters. Energy models utilize a squaring nonlinearity to produce phase invariance. The important point here is that these nonlinearities exist in order to extract image motion. They are part and parcel of the underlying computational strategy. In contrast, the preprocessing nonlinearity generally proposed for non-Fourier motion models is simply bolted onto the front of the motion-detection algorithm. The purpose of that nonlinear transformation is to translate non-Fourier motion into Fourier motion so that it becomes readily detectable through the subsequent computations. However, in the model that we employ in this paper, the detection of non-Fourier motion occurs as a direct consequence of the operation of the computational strategy for analyzing image motion. As we have argued previously,[12,16] it is the nonlinearity involved in the taking of the ratios of temporal and spatial gradients that underlies the perception of non-Fourier motion in this model.

A number of studies have examined stimuli in which there is a reversal in perceived direction of motion over some change of stimulus parameter.[17–22] These reversals have been taken as indicative of two motion-processing streams: a Fourier channel signaling motion in one direction and some non-Fourier channel signaling motion in the opposite direction. We have shown previously that in each case these direction reversals can be accounted for by a single computational strategy for motion perception.[12,15] Thus motion reversal *per se* cannot necessarily be taken as evidence of separate motion-processing streams.

Ledgeway and Smith[23] describe an experimental manipulation that has been used as evidence against single-channel models. Alternate frames of a luminance-defined (Fourier) motion sequence and a contrast-defined (non-Fourier) motion sequence are interleaved. In each sequence the sinusoidal modulation translates by a quarter of a cycle on every frame. For this particular condition, Ledgeway and Smith argue that a model that detects both luminance-defined and contrast-defined motion should see motion in the interleaved case. For one single-channel model this assertion is perfectly reasonable. Under the distortion-product hypothesis, small nonlinearities before motion analysis introduce the demodulated contrast-defined motion into the signal. Contrast-defined motion is therefore processed as if it were low-amplitude luminance-defined motion. It has recently been shown, however, that a gradient-based approach can detect both luminance- and contrast-defined motion yet fails to detect motion in the interleaved quarter-cycle phase-shift condition.[12] It is therefore not necessarily true that a single-channel model must integrate Fourier and non-Fourier frames to arrive at a coherent motion percept. The prediction from the distortion-product hypothesis, that non-Fourier motion should be processed as if it were Fourier motion, cannot be extended to single-channel models in general.

In the gradient-based approach that we used in our simulations,[8,12] Fourier and non-Fourier motion are processed by the same mechanisms. A non-Fourier stimulus is not processed as if it were an equivalent Fourier stimulus. The central tenet underlying theories of non-Fourier motion processing is that an early nonlinearity must underlie the perception of motion in non-Fourier stimuli. Our simulations offer an alternative view in which both Fourier and non-Fourier motion may be processed by a single low-level motion mechanism operating without any substantial nonlinearity before motion analysis.

ACKNOWLEDGMENTS

The research was supported by a project grant from the Engineering and Physical Sciences Research Council/Biotechnology and Biological Sciences Research Council Mathematical Modelling Initiative. J. D. Victor was supported in part by National Institutes of Health grant EY7977.

Corresponding author Chris Benton can be reached at Department of Experimental Psychology, University of Bristol, 8 Woodland Road, Bristol, BS8 1TN, UK, and by phone, 44 (0)117-9288450; fax, 44 (0)117-9288588; and e-mail, chris.benton@bristol.ac.uk.

REFERENCES

1. A. B. Watson and A. J. Ahumada, Jr., *A Look at Motion in the Frequency Domain* (Ames Research Center, Moffett Field, Calif., 1983).
2. A. B. Watson and A. J. Ahumada, "Model of human visual-motion sensing," J. Opt. Soc. Am. A **2**, 322–341 (1985).
3. E. H. Adelson and J. R. Bergen, "Spatiotemporal energy models for the perception of motion," J. Opt. Soc. Am. A **2**, 284–299 (1985).
4. W. Reichardt, "Autocorrelation, a principle for the evaluation of sensory information by the central nervous system," in *Sensory Communication*, W. A. Rosenblith, ed. (Wiley, New York, 1961), pp. 303–317.
5. J. P. H. van Santen and G. Sperling, "Temporal covariance model of human motion perception," J. Opt. Soc. Am. A **1**, 451–473 (1984).
6. J. P. H. van Santen and G. Sperling, "Elaborated Reichardt detectors," J. Opt. Soc. Am. A **2**, 300–321 (1985).

7. A. Johnston, P. W. McOwan, and H. Buxton, "A computational model of the analysis of some first-order and second-order motion patterns by simple and complex cells," Proc. R. Soc. London Ser. B **250**, 297–306 (1992).
8. A. Johnston, P. W. McOwan, and C. P. Benton, "Robust velocity computation from a biologically motivated model of motion perception," Proc. R. Soc. London Ser. B **266**, 509–518 (1999).
9. C. Chubb and G. Sperling, "Drift-balanced random dot stimuli: a general basis for studying non-Fourier motion perception," J. Opt. Soc. Am. A **5**, 1986–2007 (1988).
10. C. P. Benton and A. Johnston, "First-order motion from contrast modulated noise?" Vision Res. **37**, 3073–3078 (1997).
11. E. Taub, J. D. Victor, and M. M. Conte, "Nonlinear preprocessing in short-range motion," Vision Res. **37**, 1459–1477 (1997).
12. C. P. Benton, A. Johnston, and P. W. McOwan, "Computational modelling of interleaved first-and second-order motion sequences and translating $3f + 4f$ beat patterns," Vision Res. **40**, 1135–1142 (2000).
13. C. L. Fennema and W. B. Thompson, "Velocity determination in scenes containing several moving objects," Comput. Graph. Image Process. **9**, 301–315 (1979).
14. B. K. P. Horn and B. G. Schunck, "Determining optical flow," Artif. Intel. **17**, 185–203 (1981).
15. A. Johnston and C. W. G. Clifford, "A unified account of three apparent motion illusions," Vision Res. **35**, 1109–1123 (1995).
16. A. Johnston, C. P. Benton, and P. W. McOwan, "Induced motion at texture-define motion boundaries," Proc. R. Soc. London Ser. B **266**, 2441–2459 (1999).
17. C. P. Benton, A. Johnston, and P. W. McOwan, "Perception of motion direction in luminance- and contrast-defined reversed-phi motion sequences," Vision Res. **37**, 2381–2399 (1997).
18. C. Chubb and G. Sperling, "Two motion perception mechanisms revealed through distance-driven reversal of apparent motion," Proc. Natl. Acad. Sci. U.S.A. **86**, 2985–2989 (1989).
19. S. Shioiri and P. Cavanagh, "ISI produces reverse apparent motion," Vision Res. **30**, 757–768 (1990).
20. M. A. Georgeson and M. G. Harris, "The temporal range of motion sensing and motion perception," Vision Res. **30**, 615–619 (1990).
21. A. Pantle and K. Turano, "Visual resolution of motion ambiguity with periodic luminance- and contrast-domain stimuli," Vision Res. **32**, 2093–2106 (1992).
22. S. T. Hammett, T. Ledgeway, and A. T. Smith, "Transparent motion from feature- and luminance-based processes," Vision Res. **33**, 1119–1122 (1993).
23. T. Ledgeway and A. T. Smith, "Evidence for separate mechanisms for first- and second-order motion in human vision," Vision Res. **34**, 2727–2740 (1994).

128.243.2.30 {ts '2016-09-21 05:26:22'}