Vision Research xxx (2009) xxx–xxx

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Illusory motion due to causal time filtering

Cornelia Fermüller [a,*], Hui Ji [b], Akiyoshi Kitaoka [c]

[a] Computer Vision Laboratory, Center for Automation Research, Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742-3275, United States
[b] Department of Mathematics, National University of Singapore, Singapore
[c] Department of Psychology, Ritsumeikan University, Japan

ARTICLE INFO

Article history:
Received 8 April 2008
Received in revised form 24 November 2009
Available online xxxx

Keywords:
Illusory motion
Spatio-temporal filtering
Image motion estimate
Computational model

ABSTRACT

A new class of patterns, composed of repeating patches of asymmetric intensity profile, elicit strong perception of illusory motion. We propose that the main cause of this illusion is erroneous estimation of image motion induced by fixational eye movements. Image motion is estimated with spatial and temporal energy filters, which are symmetric in space, but asymmetric (causal) in time. That is, only the past, but not the future, is used to estimate the temporal energy. It is shown that such filters mis-estimate the motion of locally asymmetric intensity signals at certain spatial frequencies. In an experiment the perception of the different illusory signals was quantitatively compared by nulling the illusory motion with opposing real motion, and was found to be predicted well by the model.

1. Introduction

Most observers experience very strong illusory movement when viewing patterns such as Donguri (Fig. 1) and Rotating Snakes (Fig. 2) (Kitaoka, 2003). These patterns are composed of image patches which have an asymmetric intensity profile. For example, consider a narrow slice in the middle region of one of the ovals in Donguri, as shown in Fig. 3a. (The Japanese word "donguri" translates to "acorn".) Its monochromatic intensity image can be described as a white and a dark bar (the boundaries of the oval) next to different shades of gray. Referring to Fig. 3b, from the highest intensity (the white bar) the intensity drops about twice as much on the right than on the left side. Similarly, from the lowest intensity (the dark bar) the intensity rises about twice as much on the right than on the left. Thus, at the two bars the change of intensity in the right and left neighborhood is different. Informally we say that the pattern is asymmetric. Patterns with such intensity profiles create a very strong illusory effect. The perceived movement is a drift from the intensity extremum in the direction of lesser intensity change (i.e. from the white bar to light gray, and from the dark bar to dark gray) (Kitaoka & Ashida, 2003).

The illusory movement is experienced under free viewing conditions when one moves the eyes, and it is perceived in non-central vision. It stops if steadily fixating after about 6–8 s. The perceived motion is a drift, whose direction depends on the intensity relationship of the pattern elements. Chromaticity is not necessary for the illusion, but enhances the effect in some patterns (Backus & Oruç, 2005; Kitaoka, 2006). The illusion depends on the size of the image patterns. For medium sized patches such as Donguri, motion occurs in a patch when it is viewed in the periphery. Smaller patches give illusory motion closer to the center of the retina. Blur reduces the illusion in peripheral vision, but enhances it in central vision. The illusory effect is more forceful if a pattern consists of many patches, and the patches are at multiple sizes. It is stronger when the patches are circularly organized, but also exists for columnar and other arrangements.

It is generally considered that the illusory effect was first observed in patterns with circularly organized sawtooth luminance grating named the staircase illusion (Fraser & Wilcox, 1979) and the peripheral drift illusion (Faubert & Herbert, 1999) (see Fig. 18). Ashida and Kitaoka (2003) showed that the effect is much increased if the sawtooth luminance profile is replaced by step functions with intensities in the same order as in Donguri (i.e. light gray–white–dark gray–black), and if the large patches are replaced by many small ones. In Kitaoka and Ashida (2004) the authors presented patterns with continuously increasing intensity ramp-like profiles, which are perceived in central and close to central vision, and Kitaoka in (2006) proposed a classification of the different intensity profiles giving rise to the illusory effect.

A number of hypotheses for the illusory motion have been proposed. The dominant idea originating from Faubert and Herbert (1999) is that temporal differences in luminance processing produce a signal that tricks the motion system. The theories differ in how this signal is produced. Faubert and Herbert suggest that eye movements or blinks need to trigger an image motion, and the different motion signals (due to differences in intensity) are integrated over large spatial areas. Backus and Oruç (2005) focus on the perception during fixation and hypothesize that motion is not necessary, but a motion signal is triggered from the change

* Corresponding author. Fax: +1 301 414 9115.
E-mail address: fer@cfar.umd.edu (C. Fermüller).

0042-6989/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2009.11.021

Fig. 1. Variation of Donguri pattern. In peripheral vision most observers experience rotary movement. The direction in the circular arrangements alternates, with counter–clockwise direction in the upper left.

of the neural response over time. Differently strong contrasts and intensities cause different neural response curves over time (Albrecht, Geisler, Frazor, & Crane, 2002). As a result the phase of the signal is estimated erroneously as time passes and a motion signal is triggered. Their model also introduces the effect of adaptation which can account for the smooth perception under fixation over a few seconds. This effect may exist in addition to the one discussed here. Conway, Kitaoka, Yazdanbakhsh, Pack, and Livingstone (2005) discuss the illusory motion effect when flashing the pattern. Their viewpoint is that small eye movements refresh retinal stimulation, promoting new onset responses in the stimulated area on the retina. Thus the system has available a stimulus consisting of pattern frames interlaced with frames of the pattern's average intensity. Using psychophysical and physiological experiments Conway et al. (2005) argue that in addition to signals created by the differences in intensity processing, a signal analogous to the reverse phi motion (Anstis, 1970) is created. Reverse phi motion is an image motion effect caused by reversing the contrast in some frames of a video sequence. It is easy to explain that the flashing illusion will produce an apparent motion (phi motion) in the direction of the perceived motion. On appearance of the pattern the net motion of the pattern is in the direction described before and on disappearance it is in the opposite direction. In our opinion the illusory effect under free viewing and the effect when flickering the pattern are not the same. We observed that for the reduced experimental stimuli the latter is experienced much stronger than the former, and it is experienced even by observers who do not have the effect under free viewing. We therefore do not believe that Conway et al. provide a sufficient account of the illusion.

We propose that the main reason for the illusion under free viewing conditions is erroneous estimation of the image motion due to involuntary fixational eye movements. Work by Murakami, Kitaoka, and Ashida (2006) implicates drift eye movements.[1] In particular, the authors showed a correlation between the amplitude of drift movements in different observers and the strength of their illusory perception. Further evidence for the role of drifts in this illusion comes from the fMRI studies of Kuriki, Ashida, Murakami, and Kitaoka (2008). Comparing the snake illusion with a control stimulus, they found significantly increased activity in motion area MT+ (also called V5) when eye movements were present, but no increase in the absence of eye movements.

The small eye movements cause a change of the image on the retina and trigger the estimation of a motion field. This motion field is due to rigid motion and thus has a certain structure. Under normal circumstances the vision system estimates this image motion and compensates for it, i.e. the images are stabilized (Murakami, 2004; Murakami & Cavanagh, 1998). Even for asymmetric signals, the vision system estimates the correct 3D rigid motion using the average of all the motion vectors in the patterns. However, a mis-estimation occurs at certain locations in the image. The difference between the estimated rigid motion field and the erroneously estimated image motion vectors gives rise to residual motion vectors. These residual motion vectors are integrated over time and space causing the perception of illusory motion in the image.

The dominant model for motion processing in humans and other mammals is the motion energy model (Watson & Ahumada, 1985; Adelson & Bergen, 1985), and it has been found to be consistent with the physiological responses in primary visual cortex (Albrecht & Geisler, 1991). Motion is found from the response of multiple spatio-temporal filters, which are separable in space and time. The spatial filters are symmetric. The *temporal filters*, however, are *asymmetric*. This is because real-time systems have *causal* filters, which are filters that receive as input data from the present and the past, but not the future. If such filters were symmetric, the processing would be delayed by half the extent of the filter. Since early responding is valuable, the temporal responses are asymmetric in time, with greater weight given to recent input than older input.

As will be shown, causal filters mis-estimate the image motion in asymmetric image signals for certain spatial frequencies. That is, if we apply differently sized motion filters to some asymmetric pattern, we will get mis-estimation for a range of filter sizes. The resolution of the eye decreases from the center to the periphery. Thus, the size of the motion energy filters increases as we move from the center to the periphery, and their spatial frequency decreases. The illusory motion patterns consist of repeated patches of asymmetric signals, and for some of these patches the resolution of the eye is such that it leads to erroneous motion. For most of the known patterns the mis-estimation occurs at the periphery. For very small patterns with high frequency the perception is closer to the center.

The next two sections will explain in detail the reasons for the mis-estimation of image motion. The reader not interested in the technical details may want to skip these sections. We summarize here the main concept: Fig. 4 illustrates a spatio-temporal filter with symmetric impulse response in the spatial domain, and with asymmetric impulse response (Burr & Morrone, 1993) in the time domain. The spatial response may be modeled as a sinusoid of certain frequency enveloped by a symmetric function. The temporal response may be modeled as a sinusoid enveloped by an asymmetric function. Consider filtering the Donguri signal with a whole range of spatial filters of increasing size (and decreasing frequency). Let us go ahead in the paper and take a look at Fig. 8b–d, which show the amplitude of the response from filtering a single bar in Donguri. A filter of high spatial frequency will respond to the two edges bordering the bar. A filter of low spatial frequency will not recognize the edges, but only have one response to the bar. However, for intermediate frequencies, with

[1] The drift movements, one of the three fixational eye movements, are defined as incessant random fluctuations at about 1–30 Hz, quite large (~10 min of visual angle) and fast (up to 2–3°/s) (Eizenman, Hallett, & Frecker, 1985). They have greater amplitude after a saccade (Ross, Morrone, Goldberg, & Burr, 2001) than during steady fixation.

Please cite this article in press as: Fermüller, C., et al. Illusory motion due to causal time filtering. *Vision Research* (2009), doi:10.1016/j.visres.2009.11.021

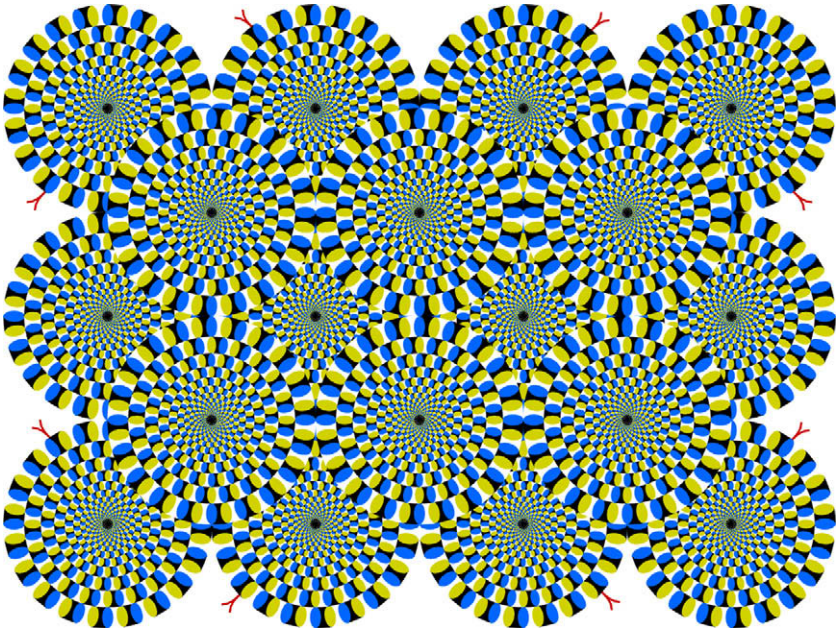

Fig. 2. Rotating snakes.

the period of the sinusoid about as large as the bar, the two edges will effect each other during filtering leading to an amplitude response curve of a larger peak merged with a smaller peak. In essence, for filters of these frequencies there is poor frequency localization. While the filtered signal should have the frequency of the filter everywhere, the actual value varies along the signal. If now we estimate image motion by applying to this signal temporal asymmetric temporal filters, we find that the temporal frequency responses from a movement to the left and a movement to the right will be significantly different. The image motion estimated as the average over the signal is larger for left motion than for right motion.

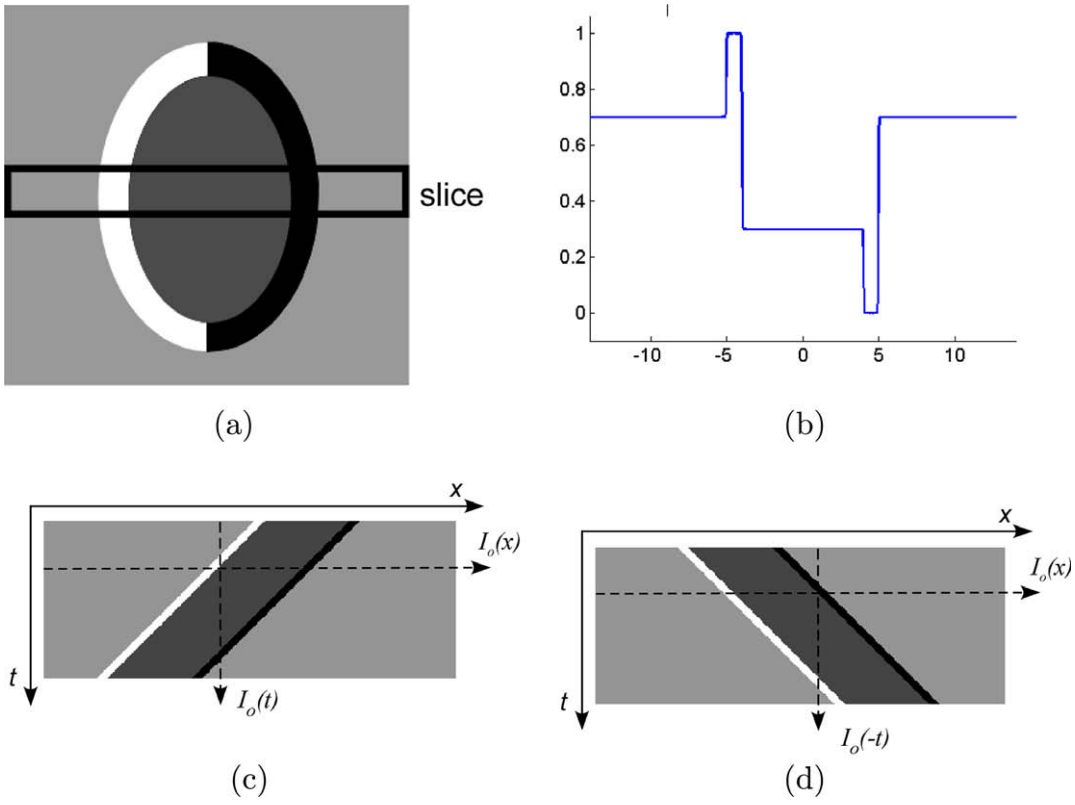

Fig. 3. (a) Slice through a patch in the Donguri pattern. (b) Its intensity profile. (c and d) Spatio-temporal picture of the patch moving to the left and to the right. $I_0(x)$ denotes a static image. $I_0(t)$ and $I_0(-t)$ denote the signal at a point over time, where the $-$ sign in $I_0(-t)$ indicates that the profile in (d) can be obtained by reflecting the profile in (c) to obtain the inverted motion direction.

Please cite this article in press as: Fermüller, C., et al. Illusory motion due to causal time filtering. *Vision Research* (2009), doi:10.1016/j.visres.2009.11.021

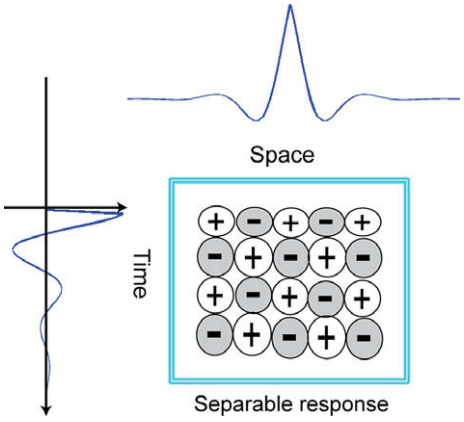

Fig. 4. Illustration of biological implementation of spatio-temporal filter (similar to Fig. 6 of Adelson and Bergen (1985)). The spatial and temporal impulse responses are shown along the margins. Their product is shown schematically in the center.

The idea of anisotropic temporal filtering was first proposed by Ashida and Kitaoka (2003), who modeled image motion estimation using a differential local motion model with asymmetric temporal derivative filters. This local model, however, requires the filters to have larger weight in the past and smaller weight in the present. Furthermore, it cannot explain the estimation at different resolutions of the pattern. For this we need to look at the different frequencies.

2. Motion estimation in the frequency domain

The monochromatic light distribution on the retina can be described as a function $I(x,y,t)$, which specifies the intensity at a point (x,y) at time t. We refer to the instantaneous light distribution at time $t=0$ as the static image $I_0(x,y)=I(x,y,0)$. Let us assume that within a small interval the change of the image can be described as a translation with constant motion velocity $\vec{v}$ of horizontal and vertical speed components (u,v). Thus, the intensity functions at time t and at time 0 are related as

$$I(x,y,t)=I(x-ut,y-vt,0). \tag{1}$$

From the three-dimensional Fourier transform of this equation, we obtain (Watson & Ahumada, 1985; Adelson & Bergen, 1985)

$$u\omega_x+v\omega_y=-\omega_t. \tag{2}$$

where ω_x, ω_y denote the spatial frequencies and ω_t the temporal frequency. This equation defines a plane through the origin in the three-dimensional frequency space.

To simplify the analysis, in the following sections we consider only images with bar-like structures parallel to the vertical dimension and the motion component perpendicular to the bars. Thus, let us consider a two-dimensional case of $I(x,t)$, that is a signal $I(x)$ which is shifted. Eq. (1) then simplifies to

$$I(x,t)=I(x-ut,0), \tag{3}$$

and the image motion constraint amounts to

$$u\omega_x=-\omega_t, \tag{4}$$

defining a line in the two-dimensional frequency space. The velocity u can be found from the ratio of the temporal and spatial frequency, i.e. as

$$u=-\frac{\omega_t}{\omega_x}. \tag{5}$$

Fig. 3c and d illustrate the spatio-temporal signal for an image line in the Donguri pattern moving with velocity $u=1$ and $u=-1$, respectively. Since the spatio-temporal signal $I(x,t)$ is obtained simply by shifting the signal $I_0(x)$, it has the same structure in the spatial and temporal domain. Referring to Fig. 3c and d, a spatial cross-section through $I(x,t)$ gives a shifted version of $I_0(x)$. A temporal cross-section gives a shifted, stretched and maybe reflected version of I_0. For unit motion to the left $(u=-1)$, the cross-section is a shifted signal $I_0(t)$, and for unit motion to the right $(u=1)$, it is the shifted signal $I_0(-t)$ (i.e. the reflection of $I_0(t)$). The amount of stretch encodes the velocity. Thus, later when we analyze temporal filtering, instead of examining the temporal cross-section, we can look at the spatial cross-section.

3. The filters

The spatio-temporal energy filters for extracting motion are separable in space and time. This just means that the filters can be created as the product between a spatial and a temporal filter. For the analysis this means that the spatio-temporal signal may first be convolved with the spatial filter and the result may then be convolved with the temporal filter.

The filters need to be localized in image space as well as in frequency space. We follow the common formulation of modeling a filter for detecting the local frequency ω_0, as a complex function

$$g(y)=p(y)\cdot exp(2\pi i\omega_0 y). \tag{6}$$

$exp(2\pi i\omega_0 y)=\cos(2\pi\omega_0 y)+i\sin(2\pi\omega_0 y)$, called the carrier function, is a complex sinusoidal for detecting the signal's component of frequency ω_0, and $p(y)$, called the envelope function, localizes the sinusoid in image space. The complex filter really consists of two filters in quadrature, the even cosine components and odd sine components. For example, in the spatial domain, the even component will respond maximally to bar-like signals and the odd component will respond maximally to edges. The magnitude of the output of the combined complex filter does not depend on whether the signal is even or odd, or any mixture thereof. As a result, complex motion filters (Adelson & Bergen, 1985; Watson & Ahumada, 1985) extract motion independent of the phase of the signal, that is independent of the position of the signal within the receptive field at certain time, and independent of the sign of the contrast.

3.1. Modeling the spatial and temporal filters

We model the spatial filters as Gabor functions (see Fig. 5a) with impulse response

$$G(x;\omega_x)=\frac{1}{\sqrt{2\pi}\sigma_x}exp\left(\frac{-x^2}{2\sigma_x^2}\right)\cdot exp(2\pi i\omega_x x), \tag{7}$$

where the envelope is a Gaussian. ω_x is the preferred frequency and σ_x determines the support of the filter, which for convenience is plausibly chosen as $\sigma_x=\frac{1}{\omega_x}$. The transfer function of the Gabor filter, which is obtained as its Fourier spectrum, amounts to $\hat{G}(\omega;\omega_x)=exp(-2\pi^2\sigma_x^2(\omega-\omega_x)^2)$, that is a Gaussian centered at ω_x and of standard deviation $\frac{1}{2\pi\sigma_x}$. The Gabor of frequency ω_x, thus extracts the signal's energy in a small frequency band around ω_x. Fig. 5b illustrates the amplitude of $\hat{G}$. Its phase is zero (i.e. there is no imaginary part), because the envelope of the Gabor is symmetric around 0. In general, symmetric filters around a point different from 0 (for example, a time-shifted Gabor) have a phase response that is linearly related to the frequency.

The temporal filter has an envelope described by a function with first-order exponential decay. We use the formulation proposed in (Chen, Wang, & Qian, 2001; Shi, Tsang, & Au, 2004), which models the envelope as a Gamma probability density function of parameter $\Gamma(2)$, resulting in temporal filters $T(t)$ of the form

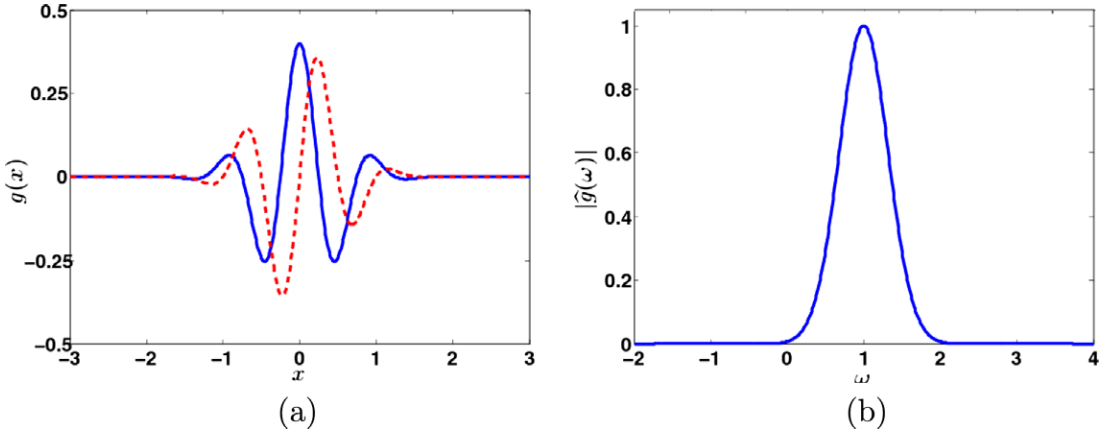

Fig. 5. Gabor filter of $\omega_x = 1$: (a) impulse response. The full (blue) line denotes the real (even) part and the dashed (red) line the imaginary (odd) part. (b) Amplitude spectrum. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$$T(t;\omega_t) = \begin{cases} \frac{t}{\tau^2} exp(-\frac{t}{\tau}) \cdot exp(2\pi i \omega_t t + i\phi_t)) & \text{for} \quad t \geqslant 0 \\ 0 & \text{for} \quad t < 0 \end{cases} \tag{8}$$

where ω_t is the temporal frequency. τ, the decay velocity, is a time constant for the envelope, which we chose as $\frac{1}{4\omega_t}$ to make the wave of the temporal filter similar to the Gabor. ϕ_t is a phase offset of the sinusoid, which is chosen such that the odd components of the filter sum to zero.

Fig. 6a illustrates the impulse response of this filter. Since the temporal filter extends from the past to the present, it actually estimates the frequency in the recent past. This can also be seen from the spectrum of the filter. As can be observed from Fig. 6b and c, the amplitude of $\hat{T}(t)$ is still a hat-type function, centered at ω_t similar to the Gabor. However, the phase of $\hat{T}(t)$ is non-zero, indicating a shift of the response in image domain. It is approximately linear for ω close to ω_t and deviates from linearity for values farther from ω_t.

3.2. Definition of filtering

When analyzing image motion, we can think of the filtering as a spatial filtering followed by a temporal filtering. First, the image sequence $I(x,t)$ is filtered with a spatial Gabor, $G(x;\omega_x)$ to obtain the image sequence $\tilde{I}(x,t;\omega_x)$ as

$$\tilde{I}(x,t;\omega_x) = \int_{-\infty}^{\infty} I(y,t) G(x-y;\omega_x) dy.$$

The idea is that the Gabor obtains the signal's component of frequency ω_x. Thus, at this stage it is assumed that the dominant spatial frequency of $\tilde{I}(x,t;\omega_x)$ at every point (x,t) is ω_x. Second, $\tilde{I}(x,t;\omega_x)$ is filtered with the temporal filter $T(t;\omega_t)$ to obtain $\tilde{\tilde{I}}(x,t;\omega_x,\omega_t)$ as

$$\tilde{\tilde{I}}(x,t;\omega_x,\omega_t) = \int_{-\infty}^{0} \tilde{I}(x,s;\omega_x) T(t-s;\omega_t) ds.$$

Here it is assumed that $|\tilde{\tilde{I}}(x,t;\omega_x,\omega_t)|^2$ returns the motion energy of I at image point (x,t) at frequencies $(\omega_x,\ \omega_t)$.

Since the spatial and temporal filters are complex valued, the complete spatio-temporal filter can be imagined as four separable filters (the even and odd components of each, the spatial and temporal filter), whose outputs are summed according to the rules of complex numbers to arrive at the motion energy.

Fig. 7 illustrates the filtering on the Donguri signal. The spatial and temporal frequencies in this example are set to one (the critical frequencies, to be explained later). Notice, that the temporal filter output is shifted with respect to the signal; to the right for left motion and to the left for right motion.

3.3. Image motion estimation

In the following we will analyze motion estimation as a function of spatial frequency. The image motion of a patch (or in the analysis a line through the patch), is computed from all the measurements in the patch in two computational steps: first, we estimate at every point the (best) velocity. Second, we compute the velocity of the patch as the weighted average of point-wise velocity estimates.

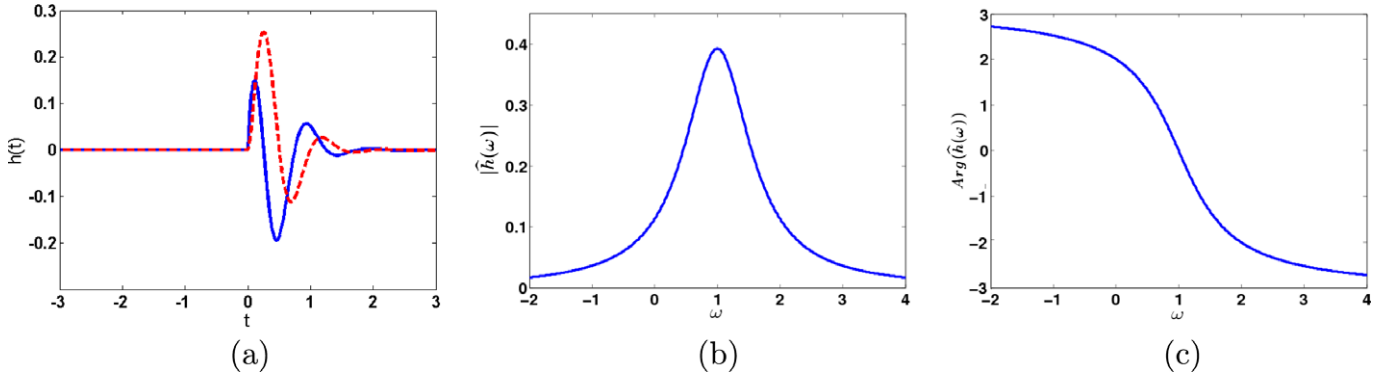

Fig. 6. Temporal filter of $\omega_t = 1$: (a) impulse response. The full (blue) line denotes the real part, the dashed (red) line the imaginary part. (b) Amplitude spectrum. (c) Phase spectrum. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Please cite this article in press as: Fermüller, C., et al. Illusory motion due to causal time filtering. *Vision Research* (2009), doi:10.1016/j.visres.2009.11.021

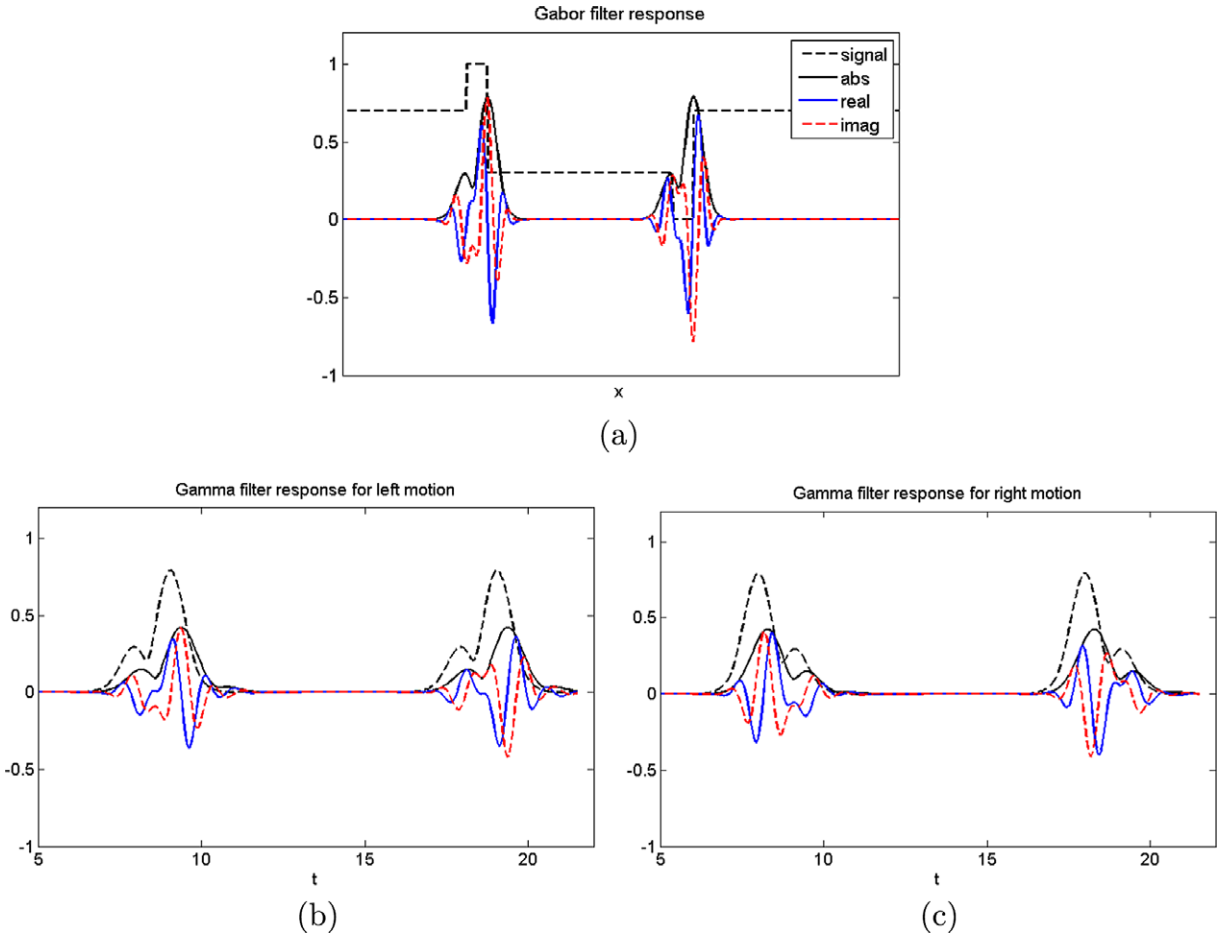

Fig. 7. Filtering of the Donguri signal at the locations of the horizontal and vertical cross-sections in Fig. 3c and d: (a) Spatial filtering the intensity signal with a Gabor of frequency 1. (b and c) Temporal filtering the signal in (a) with a frequency of 1 for left and right motion.

Specifically, given a spatial frequency ω_x, the signal $I(x,t)$ is filtered with the spatial Gabor to obtain $\tilde{I}(x,t;\omega_x)$. Then at some time t (all t are equivalent, since the motion is simply a shift of the spatial signal) at every image point x, we find the dominant temporal frequency ω_{t_0}. To do so we filter with a range of temporal filters $T(t,\omega_t)$ of different frequencies ω_t and choose the filter response with maximum energy. (Ideally according to the motion constraint, only one temporal frequency filter should return non-zero energy.) This way, we find at every point x a local velocity estimate

$$\hat{u}(x;\omega_x) = -\frac{\omega_{t_0}}{\omega_x} \tag{9}$$

and its corresponding energy

$$|\tilde{\tilde{I}}(x,t;\omega_x,\omega_{t_0})|^2.$$

Then the motion of a patch is found as the average of energy weighted velocity measurements:

$$\hat{u}(\omega_x) = \frac{\sum_x \hat{u}(x;\omega_x)|\tilde{\tilde{I}}(x,t;\omega_x,\omega_{t_0})|^2}{\sum_x |\tilde{\tilde{I}}(x,t;\omega_x,\omega_{t_0})|^2}. \tag{10}$$

4. The effect of filtering on Donguri

Fig. 8 illustrates the effect of spatial filtering on Donguri. At frequencies ω_x larger than the reciprocal width of the bar, the Gabor filter detects the two edges at the left and right of the bar (Fig. 8b). At frequencies significantly smaller than the reciprocal width of the bar, the Gabor detects the bar (Fig. 8c). The amplitude of the response thus has either one or two well separated peaks. However, for frequencies of ω_x close to the reciprocal width of the bar, there is something in between one and two responses. The amplitude function becomes asymmetric with two merging peaks, a larger on the right and a smaller on the left (Fig. 8d). Let us call these frequencies the "critical frequencies".

As is well known from the uncertainty principle, there is a limit on the accuracy of localization in image and frequency domain. The Gabor (which is the filter with best localization in joint image and frequency space) cannot guarantee perfect localization of the signal. Because of the "hat" profile of its Gaussian envelope, the filtered signal $\tilde{I}(x,t;\omega_x)$ will not always have local dominant frequency ω_x. We can understand the poor localization of frequencies from the phase responses. Referring to Fig. 8e–g, the phase responses are (nearly) linear for $\omega = \frac{1}{2}$ and $\omega = 2$, but the phase response deviates significantly from linearity for the critical frequencies, which is an indicator for poorly estimated frequencies.

When now estimating on the asymmetric signal $\tilde{I}(x,t;\omega_x)$ image motion with asymmetric temporal filters, left and right motion are estimated of different value. Fig. 9 shows that for the critical frequencies, motion to the left $(u=-1)$ leads to larger velocity estimates than motion to the right $(u=1)$.

We can intuitively understand this estimation from the amplitudes of the signal and the filter. We convolve two asymmetric signals. The temporal filter amplitude has more weight for larger t and smaller weight for smaller t. Referring to Fig. 7b and c, for

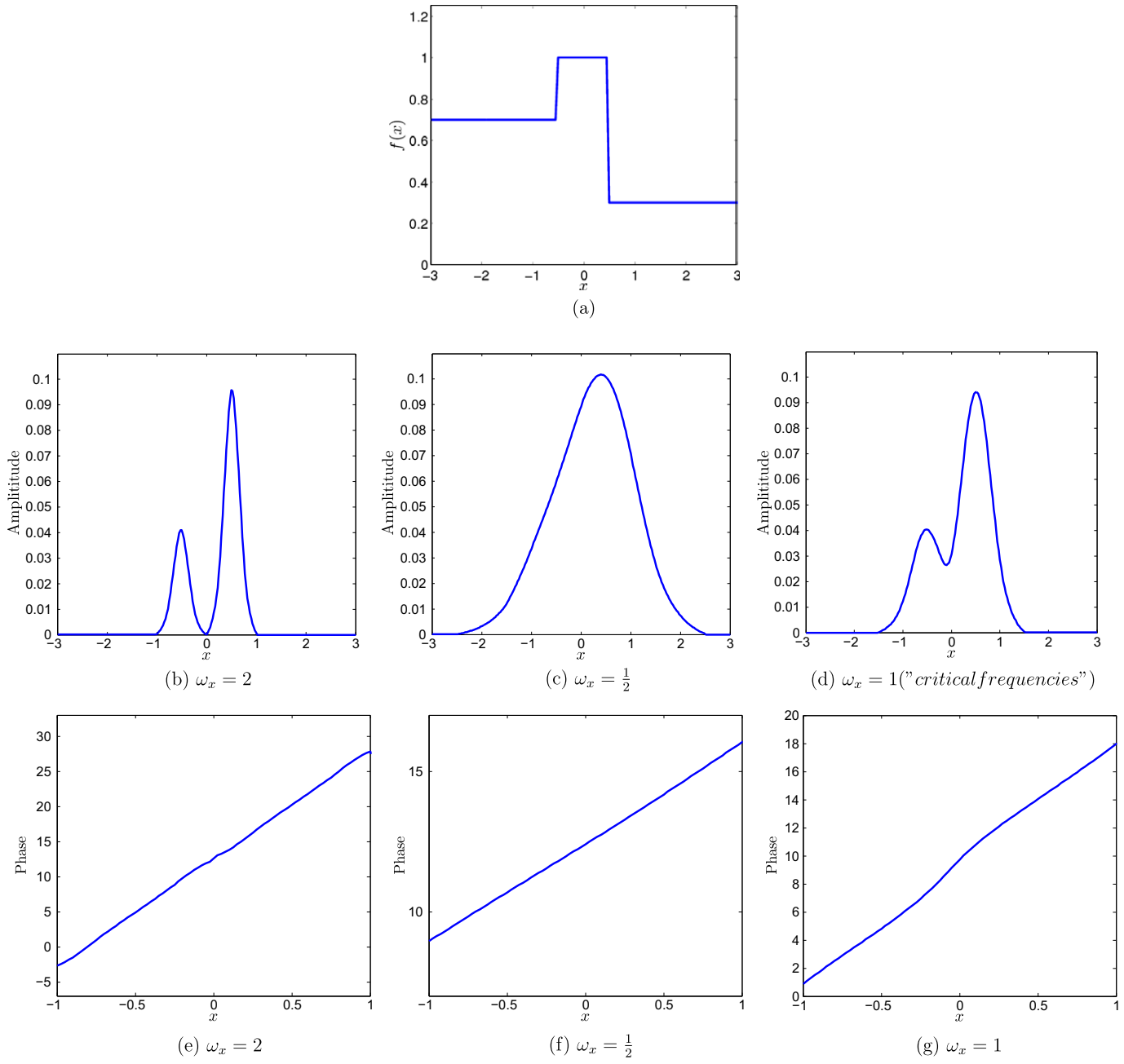

Fig. 8. (a) Bar in Donguri pattern. (b–d) Amplitude of bar filtered with Gabor of different frequencies ω_x. (e–g) Corresponding phase response. For better illustration, the range of the phase (shown on the *y*-axis), since it depends linearly on the frequency has been scaled, so that in (e) it is twice and in (f) half the size of (g).

a left motion, signal $\tilde{I}$ has a larger lobe for larger t and smaller lobe for smaller t, but for right motion the order is reversed.

Fig. 9 shows the local estimated velocity (as full, green line) at every point on the bar. The corresponding amplitude is shown as dot-dashed, red line, and the amplitude of signal $\tilde{I}$ is shown as dashed, blue line (in the spatial domain). Both amplitudes have been scaled to allow for better visualization.

Because of interaction of the regions under the two peaks with each other during temporal filtering, the local velocity (Eq. (9)) varies significantly along the signal. Most significant, there is overestimation of velocity at the right peak for left motion, and underestimation of temporal energy at the left peak for right motion.

This is further demonstrated in Fig. 10, which shows the estimated energy for three different temporal frequencies. As a result of this local mis-estimation, the average velocity (Eq. (10)) is larger for left motion than for right motion.

For higher spatial frequencies (Fig. 8b) the peaks are well separated and do not interact, and for lower spatial frequencies (Fig. 8c) there is only one peak. Thus, in both cases there is no significant difference between left and right motion.

Two final notes: Throughout the demonstration we have used a normalized speed of 1 unit, but the findings apply to any velocity. A different velocity, of say value α, amounts to stretching/compressing the signal $\tilde{I}(t)$ to $\tilde{I}(\frac{t}{\alpha})$. Then a temporal response of ω_t for the unit velocity will correspond to a temporal frequency response of $\alpha\omega_t$ in the stretched signal. Thus, all velocities will be mis-estimated by the same percentage.

The size of the motion filter (with non-vanishing energy) in our implementation is about five times the bar width. (The spatial

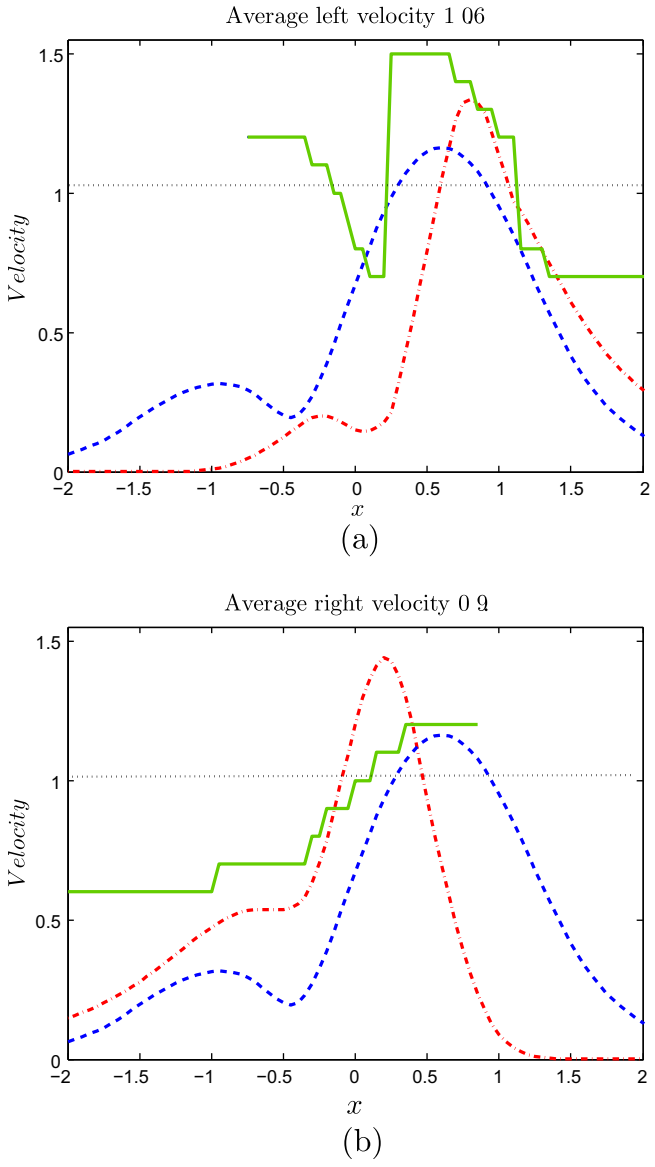

Fig. 9. Velocity estimation at critical frequencies. The dashed (blue) line denotes the scaled amplitude of $\tilde{I}$. The full (green) line denotes the local estimated velocity (Eq. (9)) and the dot-dashed (red) line denotes the corresponding scaled amplitude. (*Note:* because the temporal filter estimates the motion at a point earlier in time, the maximum value for $\tilde{I}$ is found to the right of the stronger edge for left motion and to the left of the stronger edge for right motion). The estimated average velocity (estimated using Eq. (10)) is larger for left than for right motion. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

component with significant energy is four times and the temporal component is two times the size of the bar, see Fig. 5). Our analysis is a simulation of the continuous derivation. Clearly, our vision system does not have motion filters for every position on the retina. But if we assume a non-biased distribution of the filter locations (for example, uniform, or random), we can say that statistically the filter outputs should approximate the continuous signal.

5. Experimental evaluation

5.1. Donguri and rotating snakes

The following figures show the estimated image velocity as a function of spatial frequency. The estimates were obtained by simulations as described in Section 3.3. That is, the motion of a pattern element is computed as the energy weighted average of the local velocities of all the points on the signal (Eq. (10)). Fig. 11a shows the estimated velocity for left and right motion for a large range of frequencies, demonstrating that significant differences occur in a small range around frequency $\omega = 1$. Fig. 11b zooms in on a neighborhood around the critical frequencies, but shows only the difference in velocity between estimated left and right motion. Let us clarify, higher frequencies (of the Gabor filter) in our plot correspond to higher resolution of the perceived image, that is filters located closer to the fovea.

Referring to Fig. 11a, the estimates fluctuate in the neighborhood of the critical frequencies. There is an overestimate for left motion and an underestimate for right motion at $\omega = 1$. Both velocities are overestimated for a bit larger ω (1.25), and both are underestimated for a bit smaller ω (0.75), but at these frequencies their differences are not significant.

To test the validity of the approach, we experimented by varying the parameters in the motion estimation. In particular, we varied the range of possible temporal frequencies (with the smallest range [0.6...1.4] and the largest unlimited), the size of the spatial and temporal filters, and the weighting of the local velocity estimates. Besides the energy, we used the absolute value of the filter response and its cube for weighting. We found that for some parameter settings, both left and right motion were underestimated at $\omega = 1$. However, for all settings, there was a significant difference at the critical frequency, with the left motion being larger than the right. Based on these experiments, we state the gist of our finding as: *Estimated left motion is larger than estimated right motion for the critical frequencies.*

Next, consider the Rotating Snakes pattern (Fig. 2) and take three cross-sections through one of its units to obtain three qualitatively different profiles. Referring to Fig. 12, the first cross-section is at the center, providing a profile just like Donguri's, the second cross-section is in the upper half, where all intensity regions have equal width, and the third is close to the top of the unit, where the intermediate intensity regions (yellow and blue) become narrow bars and the black and white regions have large extent. We refer to these profiles as "Donguri", "Bars", and "Steps", respectively. Linear arrangements of the corresponding monochromatic signals are shown in Fig. 12b and c. Fig. 13 plots the difference in estimated image velocity between left and right motion for the three functions. The simulations show that for all three signals there is a range of frequencies for which left motion is significantly larger than right motion. The difference is significantly larger in Donguri than Bars, and is larger in Bars than Steps.

5.2. Nulling experiment

The strength of the illusory perception varies significantly between observers. The relative strength of the perceived motion in different signals, however, can be used to evaluate the model. We quantitatively compared the perception of the three signals above by nulling the illusory motion with opposing real motion, similar as in Murakami et al. (2006). Nine naive subjects participated in the experiment.

5.2.1. Methods

The signals were arranged on three concentric rings, with the middle ring three times the width of the inner and outer rings. Each ring consisted of 40 signal elements, and the flanking rings were phase shifted with respect to the middle ring by a quarter of the element (Fig. 15). The linearly calibrated intensity values of the four regions were 0.3, 1, 0.7, 0, where 0 is black and 1 is white. The width of the bar was 1 unit and the other regions were 3 units in Donguri and Stairs, and all regions were 2 units in Bars (as shown in Fig. 12). In addition to these three signals, we also

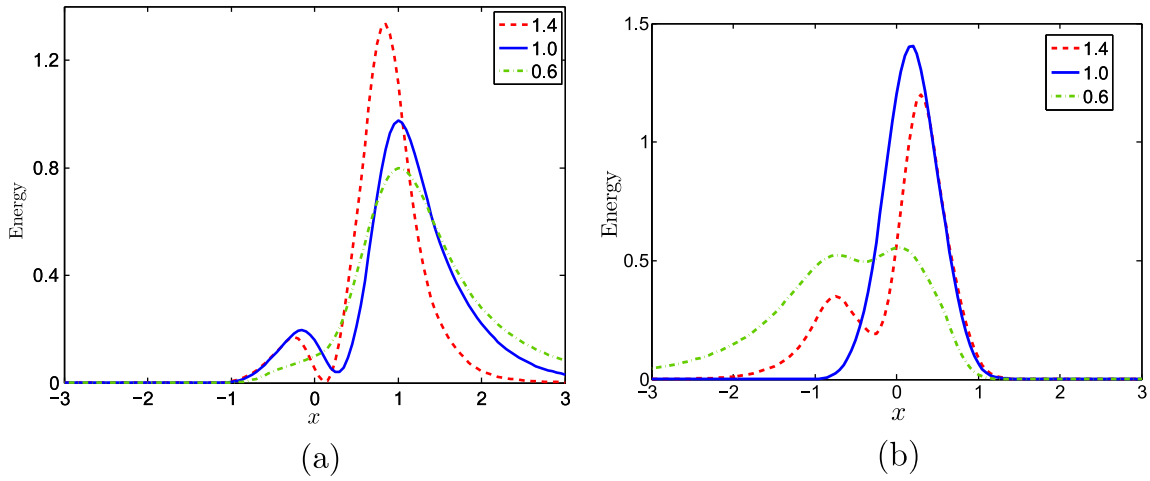

Fig. 10. Amplitude of $\tilde{I}$ for temporal frequencies $\omega_t = 0.6, 1.0$ and 1.4. Left motion is characterized by overestimated temporal frequency on the right edge, and right motion is characterized by underestimated temporal frequency on the left edge.

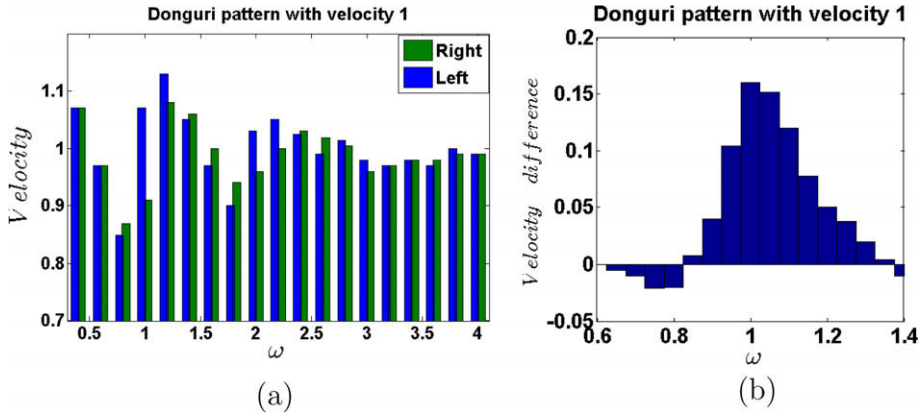

Fig. 11. (a) Velocity estimation for Donguri. (b) Difference in velocity estimates between left and right motion. Higher values of ω correspond to higher resolution images.

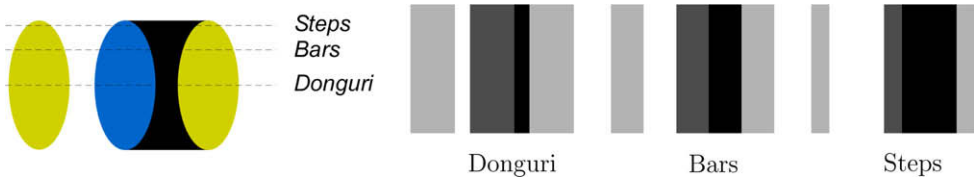

Fig. 12. Three different illusory intensity signals in Snake.

tested a Donguri signal of reduced contrast with the intensities 3/8, 1, 5/8, and 0. Simulations for this signal (Fig. 14) show that the range of frequencies with left and right motion being significantly different is smaller than in the original Donguri, but the predicted value at the maximum is nearly the same, actually slightly larger.

Observers were sitting at a distance of 45 centimeters in front of the 17″ LCD screen and observed the patterns binocularly. At this distance the rings covered 14° of visual field with the width of the three rings covering 1.7°. At the center of the patterns was a ring of 1° filled with random black and white dots for gaze control. Subjects were instructed to look inside the disc freely. For each signal two patterns were created, one with the intensity regions in the order shown above inducing counterclockwise motion, and one obtained as the mirror reflection of the former, inducing clockwise motion.

Using Matlab, an interface was created that allowed to play videos showing these patterns rotating slowly clockwise or counterclockwise. The speed of motion could be set in the range of 0.06—0.6°/*s* by the step of 0.06°, where 1°/*s* corresponds to one degree of polar angle per second.

The speed of motion that gave the subjective stationary percept was found with the Method of Adjustment. Observers were first presented with the static pattern. They then adjusted on a slider the speed of motion, upon which a video of the pattern drifting at the selected speed appeared in the location of the static pattern. Observers increased and decreased the speed until they found the speed, which gave rise to the perception of a stationary pattern.

5.2.2. Results

Fig. 16 displays the measurements. The speed nulling illusory motion in Donguri was in the range of 0.12—0.36°/*s*. All participant perceived Donguri the strongest, and Snake stronger than Stairs. None of the subjects measured a significant difference between

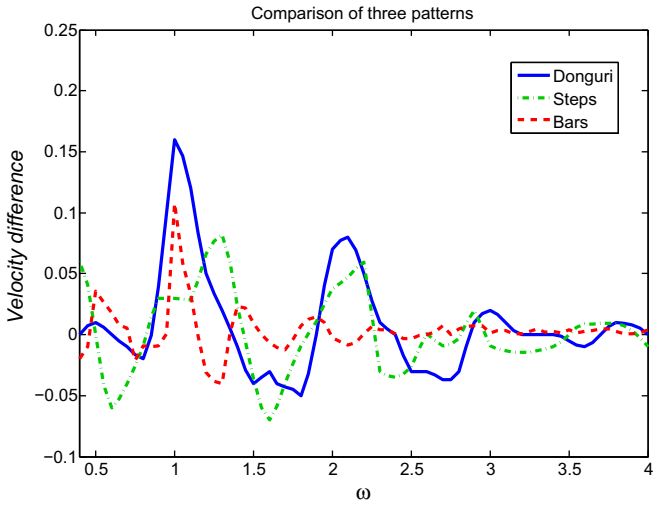

Fig. 13. Difference in estimated velocity between left and right motion for the three illusory signals in Snake.

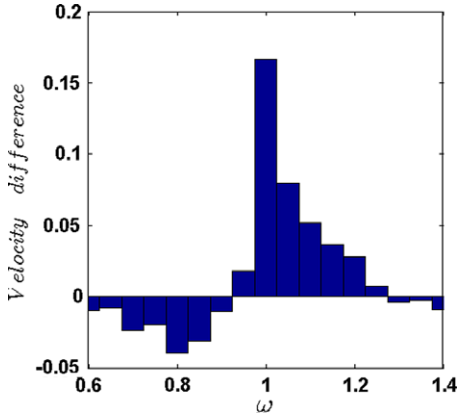

Fig. 14. Difference in estimated velocity for Reduced Donguri.

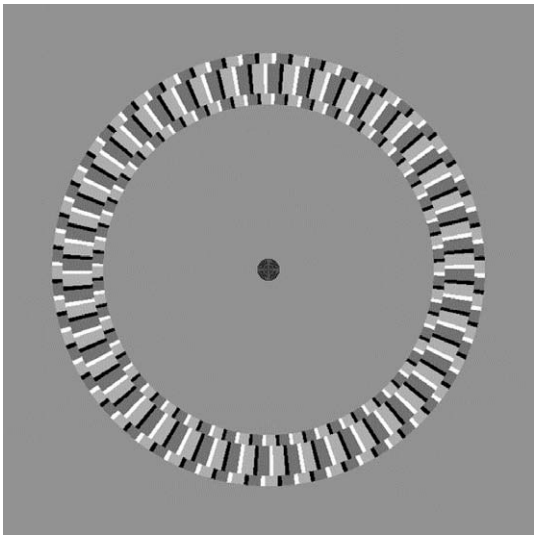

Fig. 15. Example of experimental stimulus.

the original Donguri and Donguri with decreased contrast, and between clockwise and counterclockwise stimuli. Fig. 17 compares the predictions to the mean of measurements. The values shown are the ratio of the nulling motion in Donguri to Snake, Bars and Reduced Donguri, and were obtained as the average over clockwise and counterclockwise patterns over all subjects. The predicted motions were found as the estimated maximum difference between left and right motion over all frequencies. The figure demonstrates that our model predicts observed ratios between different conditions of the experiment very well.

5.3. Peripheral drift and central drift

It is generally considered that the illusory motion effect was first observed in the peripheral drift illusion (Fraser and Wilcox, 1979) (see Fig. 18). The intensity profile in this pattern is a sawtooth function. Such a function would not give rise to erroneously estimated motion according to our model. However, luminance recorded at the neural level is usually modeled as a non-linear function of the actual intensity of the image. Following Backus and Oruç (2005), we consider two factors in our model: first, luminance adaption, a logarithmic function modeling the relationship between recorded luminance and physical intensity (changes at higher intensity values are recorded with a smaller value than changes at lower intensity values); second contrast adaptation, a sigmoid function modeling greater sensitivity to the middle range than the high and low ranges of intensities (see Fig. 19).

Kitaoka and Ashida (2004) created a series of patterns, which they call central drift illusions, as they are perceived in central as well as peripheral vision. For examples, see Sakura and Cendri in Fig. 21. These patterns contain elements (the petal and ovals) with (close to) linear intensity profiles, but in comparison to the peripheral drift illusion, the individual elements are separated by uniform background. This separation increases the illusory effect.

Applying our luminance model to the actual intensities, we obtain the luminance profiles shown in Fig. 20. Our model's predicted velocity differences between left and right motion for Sakura are shown in Fig. 22. The petals in the model are four units (the bar in Donguri is one unit). Thus, the critical frequencies of $\omega = 1$ in Fig. 22 corresponds to the period of the sinusoid being a $\frac{1}{4}$ of the petal size. Fig. 22b compares the velocity differences in peripheral drift, Sakura, and Cendri for a small range around the critical frequencies. Our model predicts perceived illusory motion in the patterns, with the effect being stronger in Sakura and Cendri than in peripheral drift.

A similar signal, which Kitaoka (2006) calls Type I, consisting of either white to medium gray elements on dark background or

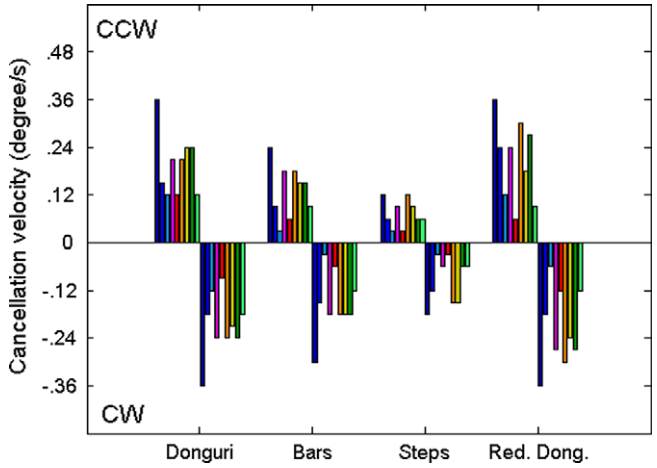

Fig. 16. Cancelation velocity in 9 subjects for clockwise and counterclockwise illusory motion. Each color corresponds to one subject. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 17. Comparison of prediction to experimentally obtained velocity ratios of Bars: Donguri, Steps: Donguri, and Reduced Donguri: Donguri.

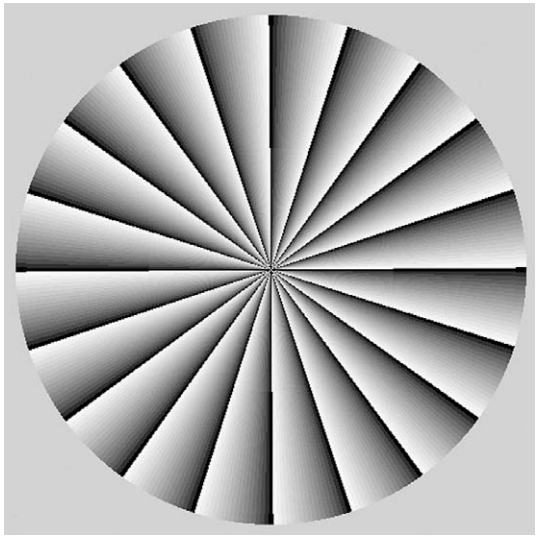

Fig. 18. The peripheral drift illusion (Fraser & Wilcox, 1979). In peripheral vision, the circle appears to rotate slowly in clockwise direction.

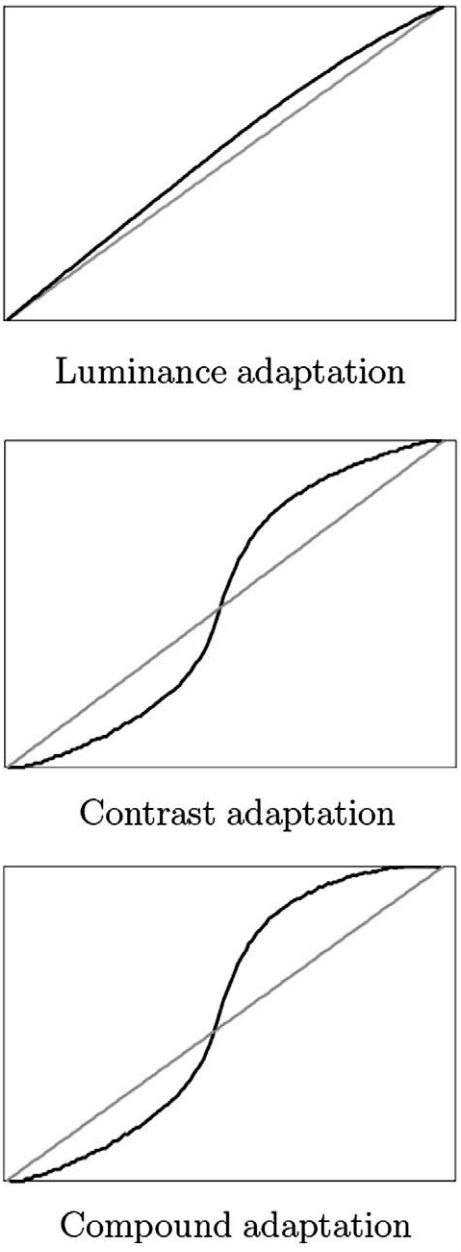

Fig. 19. Model of neural luminance function.

medium to dark gray elements on white background, causes illusory motion in the periphery (Fig. 23a). The corresponding profiles may be considered smooth versions of the Donguri-profile. A simulation of the motion estimate, considering our luminance model predicts the perceived motion (Fig. 23c). In this pattern one element is chosen three units. Fig. 24 shows one of Kitaoka's patterns from this class, in which the two elements are combined for an even stronger effect.

Let us note that contrast and luminance adaptations would not effect the motion estimation in Donguri and Snake. The estimation is very robust over luminance changes. As shown, the reduced Donguri signal gives rise to very similar motion prediction.

6. 2D motion

The estimation of instantaneous 2D image motion still can be imagined as a two-stage computational process. In the first stage, causal filters estimate point-wise erroneous motion in the direction perpendicular to the spatial filter orientation (the 1D motion component, also called normal flow, which is the projection of the 2D motion vector on the tuning direction of the filter). In the second stage normal flow estimates in different directions within spatial local neighborhoods are combined and the 2D image motion of the patch is estimated.

We implemented the following simple motion algorithm to demonstrate that the residual motion vectors are consistent with the perceived illusory motion: at the critical frequency of the pattern, at every image point we obtain the spatial frequency responses using a standard set of Gabor filters, and we estimate the corresponding temporal frequency of maximum energy using causal filters. Thus, we arrive at n equations of the form

$$\omega_{x_i} + \omega_{y_i} v = -\omega_{t_i} \quad \text{with} \quad i = 1, \ldots, n. \tag{11}$$

Then we compute the flow (u, v) of every pattern element by solving the over-determined system of n equations in (Eq. (11)) using weighted least squares estimation, with the weights the energy responses of the filter outputs.

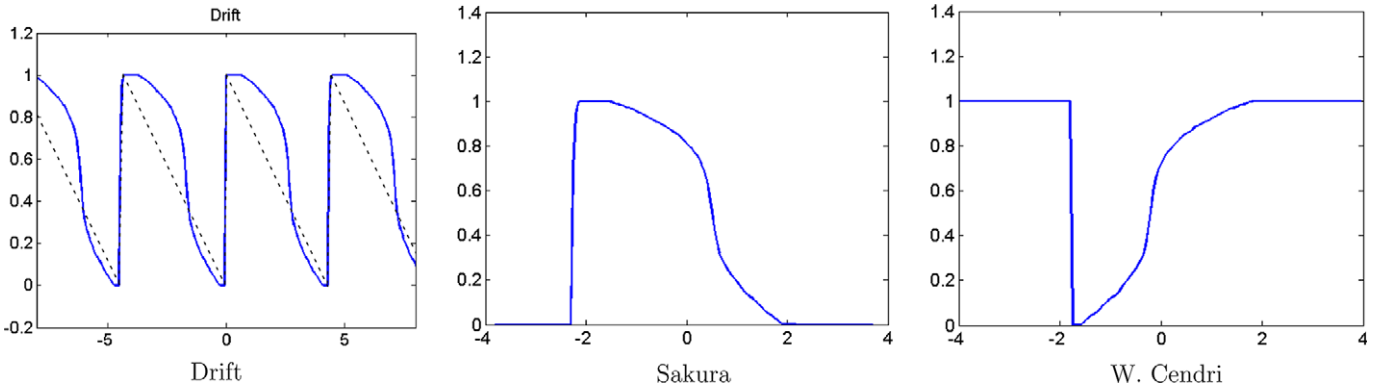

Fig. 20. Modeled luminance profile in peripheral and central drift patterns.

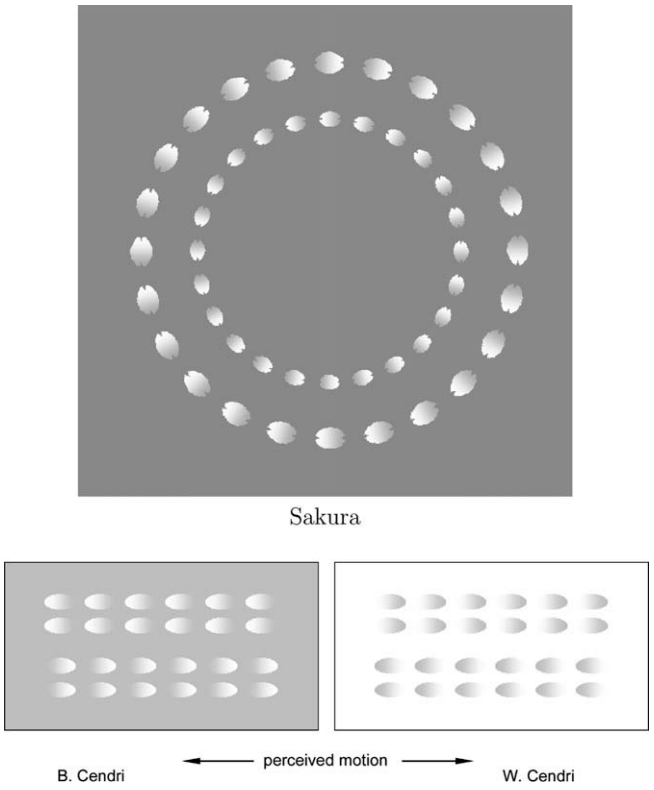

Fig. 21. Sakura in gray: when fixating on the center, the outer petals appear to move slowly clockwise and the inner petals move counterclockwise. B. Cendri (same as Sakura): the ovals on gray background move from gray to white. W. Cendri: the ovals on white background move from white to gray.

Fig. 25a and b show for the Donguri pattern the residual flow vectors resulting from a horizontal and a vertical movement, respectively. Each of the movements produces image motion on most of the individual pattern elements. The motion is largest for the elements with dominant edge direction perpendicular to the movement, but even on elements oriented 60° away from that direction there is some image motion. If we combine image motions from only two movements spaced at least 30° apart, we will obtain image motion on all elements. We expect that our vision system integrates the motion signals over a time interval of a few eye movements. Since each movement produces image motion in a large range of directions, this process should not be sensitive to the particular directions of eye movements. Fig. 25c shows the vector sum of the flow fields due to the horizontal and vertical movements.

7. Discussion

7.1. Signal integration

Estimation of local image motion is the first step in visual motion analysis. The local signals are input to many visual processes. Some of the very basic processes are the estimation of our own motion, that is the relative motion of the eye with respect to the scene, and the segmentation of the scene into different objects. While causal filters create local erroneous motion signals in this illusion, the strong perception of rotating patterns is due to these further processes. First, using the retinal motion signals over the whole visual field, a 3D motion estimation process obtains the eye movements (and the head and body movements if there are any), and stabilizes the image. Second, a segmentation process using as input the local motion signals together with information from static cues, such as edges, texture and color, performs a grouping into circular elements of rotational motion.

According to our motion model all image motion in asymmetric signals should be estimated with error. However, the erroneous estimation in the illusions is due to the motion signal from drift movements (Murakami et al., 2006). We speculate that the role of drift movements for this illusion lies in a better temporal integration of the motion signal when compared to signals from other movements. We know that images are computationally stabilized. The drift motion is computed from the local motion signals over the whole visual field. Then the drift is discarded, and the image signals over a time interval are integrated. This is computationally feasible, because the drift motion is mostly a rotation and does not depend on the structure of the scene. By fitting to the whole image motion field a rotational motion field, which only depends on three parameters, local motion vectors can be estimated very accurately and reliably. On the other hand, head motions and scene motions also involve translation, and the image motion field then depends on the scene. Therefore, local motion estimation cannot be that accurate, and integration over a time interval is more difficult.

The illusion appears a bit stronger when viewing the patterns binocularly versus monocularly. This may be attributed to the responses of binocular motion signals. The drift movements in the two eyes are independent. Any single directional movement gives rise to erroneous residual image motion only on some parts of the patterns (where the edges are perpendicular to the movement, as can be seen in Fig. 25). Two different drift movements, thus, cause erroneous image motion on more parts, and provide more information for the integration into rotational motion.

Some observers, and especially many older people do not perceive this illusion. We speculate that it is not a lack of involuntary eye movements, but decreased sensitivity to motion in the temporal

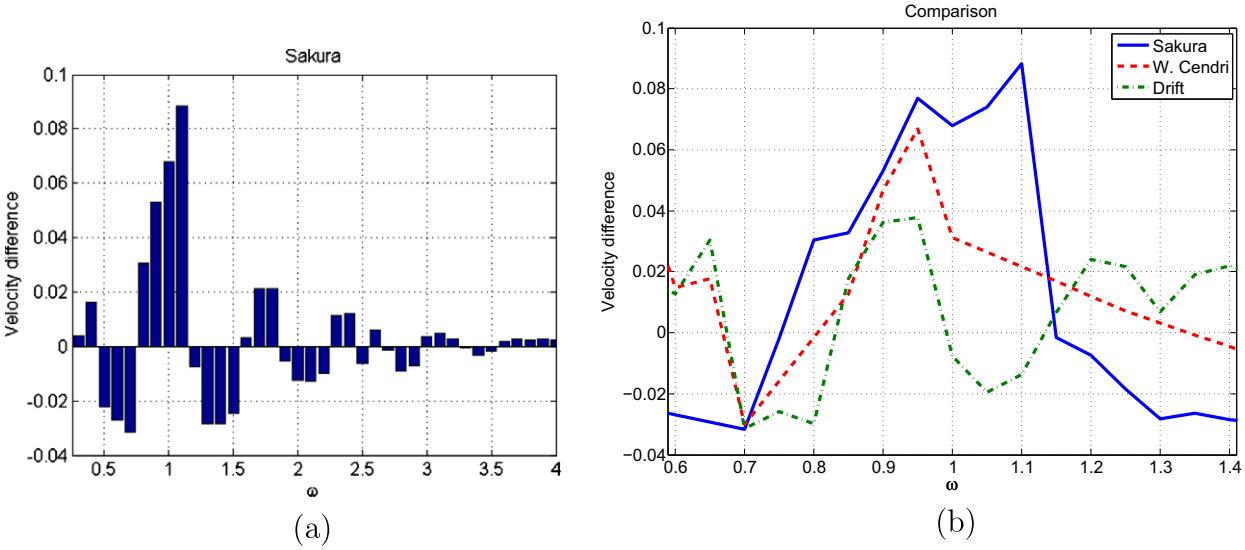

Fig. 22. (a) Difference in estimated velocity between left and right motion in Sakura. (b) Comparison of motion estimation between peripheral and central drift illusions.

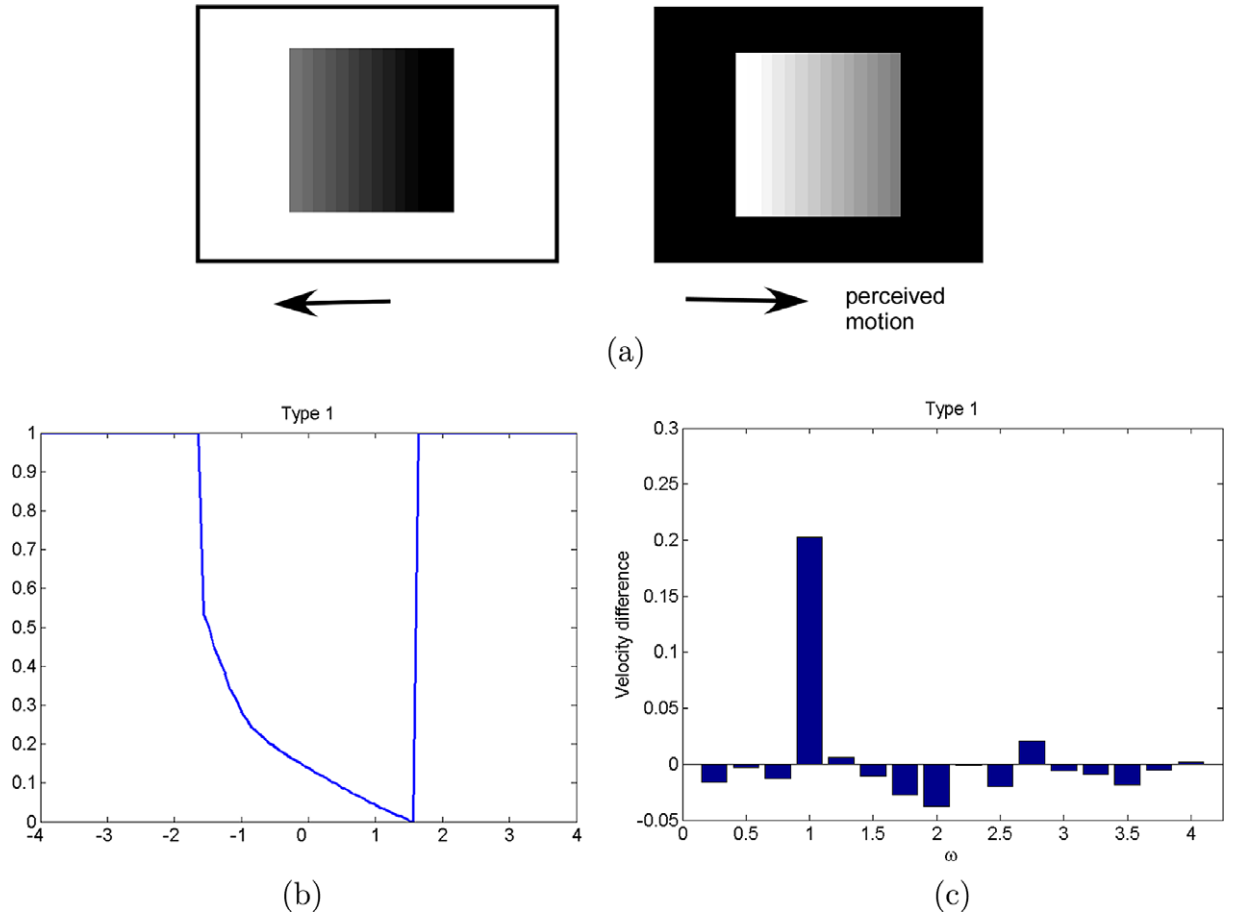

Fig. 23. (a) Gray shaded elements on bright and dark background. Dark elements on white move from dark to light. Light elements on black move from light to dark. (This is opposite to the peripheral drift illusion!) (b) Luminance profile for the dark element on white background. (c) Difference in estimated velocity between left and right motion.

high and middle frequency range, as has been measured in older people (Shinomori and Werner, 2003; Shinomori and Werner, 2006).

7.2. Relationship to geometric optical illusions

The concept of smoothing at certain scale as an explanation for optical illusions is not new to the literature. Morgan and Moulden (1986) and Morgan and Casco (1990) have proposed that bandpass filtering (that is edge detection by computing derivatives on a smoothed image) is the cause of a number of (static) geometric optical illusions. For an example see Fig. 26a. The illusory elements in this pattern are bars. As discussed in Fermüller and Malm, 2004, if we smooth a bar with a Gaussian of σ large enough to effect both edges of the bar but not large enough for the two edges to merge,

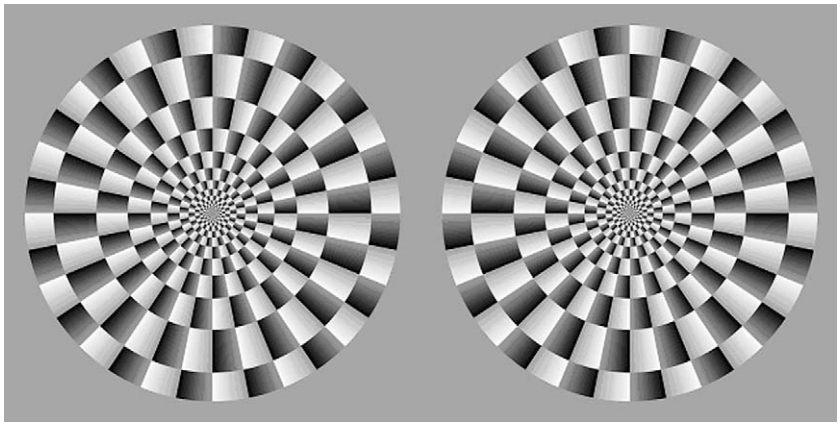

Fig. 24. Type I illusion (Kitaoka, 2006).

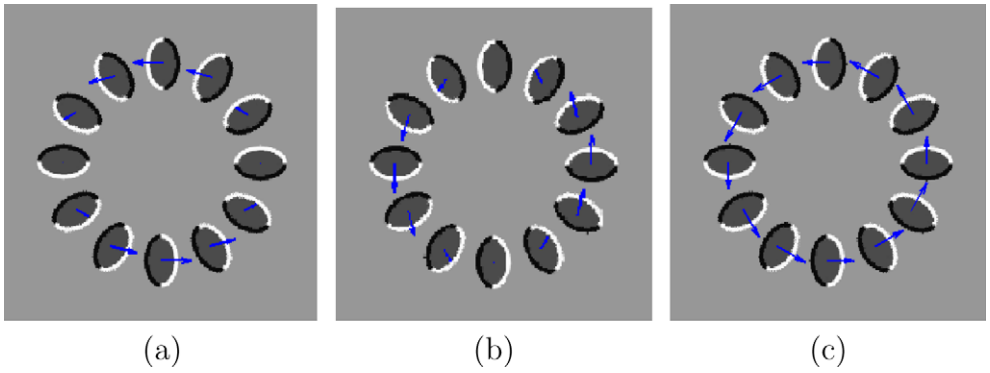

Fig. 25. (a) Estimated residual flow for Donguri at critical frequencies for (a) horizontal, (b) vertical, (c) combined horizontal and vertical motion.

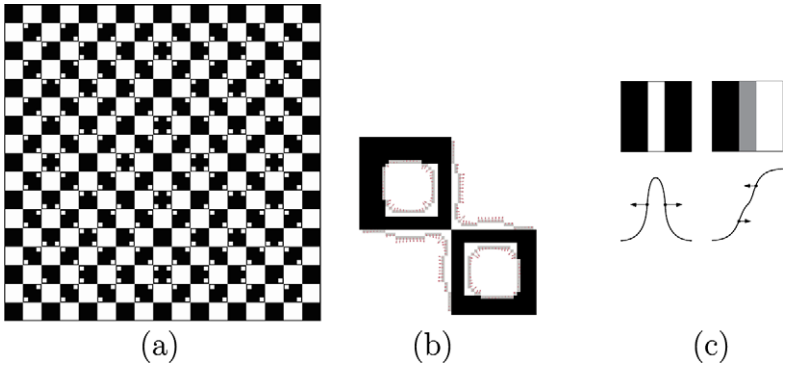

Fig. 26. (a) Illusory pattern "waves" – a perfect checkerboard pattern with superimposed squares – causes the perception of wavy lines (Kitaoka, 1998). (b) Demonstration of the movement of edges under smoothing for a small part of the pattern. (c) A schematic description of the behavior of edge movement in scale space. The first row shows the intensity functions of the two different bars, and the second row shows the profiles of the (smoothed) functions with the dots denoting the location of edges, which either drift apart or get closer.

the location of the edges changes, as illustrated in Fig. 26c. For a bright bar in a dark region (or a dark bar in a bright region) the two edges drift apart. For a bar of medium brightness next to a bright and a dark region the two edges move toward each other. The latter case corresponds to the Donguri profile. The σ in the Gaussian is the same as the σ in the Gabor of the "critical frequencies". Thus, at the "critical frequencies" the interaction of the two edges causes a change in the location of the edges (defined as the extrema in the first-order derivatives or zero-crossings in the second-order derivatives). In this paper we showed, that at the same time local frequencies are poorly estimated, which has an effect if image sequences are filtered asymmetrically in temporal domain.

7.3. Summary of the paper

Temporal image motion filters are causal, i.e. they use data from the past, but do not use data from the future. Such filters are asymmetric giving greater weight to recent input than older input. In this paper we showed that this asymmetry in the filters leads to erroneous estimation of image motion for asymmetric signals at certain scale. This is simply because of the universal uncertainty

in estimating signals. We demonstrated the mis-estimation using simulations. Then we tested our model quantitatively using different signals with bar-like structures and found that it very well predicts the illusory motion perception. Based on these findings, we hypothesize that this erroneous estimation explains the illusory perception of motion in static patterns with repeated asymmetric pattern elements under free viewing conditions.

References

Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A, 2*, 284–299.

Albrecht, D., & Geisler, W. (1991). Motion selectivity and contrast response function of simple cells in the visual cortex. *Visual Neuroscience, 7*, 531–546.

Albrecht, D., Geisler, W., Frazor, R., & Crane, A. (2002). Visual cortex neurons of monkeys and cats: Temporal dynamics of the contrast response function. *Journal of Neurophysiology, 88*, 888–913.

Anstis, S. (1970). Phi movement as a subtraction process. *Vision Research, 10*, 1411–1430.

Ashida, H., & Kitaoka, A. (2003). A gradient-based model of the peripheral drift illusion. In *Proceedings of the ECVP, Paris*.

Backus, B. T., & Oruç, I. (2005). Illusory motion from change over time in the response to contrast and luminance. *Journal of Vision, 5*(11), 1055–1069.

Burr, D., & Morrone, M. (1993). Impulse response functions for chromatic and achromatic stimuli. *JOSAA, 10*, 1706.

Chen, Y., Wang, Y., & Qian, N. (2001). Modeling V1 disparity tuning to time-varying stimuli. *Journal of Neurophysiology, 86*, 143–155.

Conway, B., Kitaoka, A., Yazdanbakhsh, A., Pack, C., & Livingstone, M. (2005). Neural basis for a powerful static motion illusion. *Journal of Neuroscience, 25*, 5651–5656.

Eizenman, M., Hallett, P., & Frecker, R. (1985). Power spectra for ocular drift and tremor. *Vision Research, 25*, 1635–1640.

Fermüller, C., & Malm, H. (2004). Uncertainty in visual processes predicts geometrical optical illusions. *Vision Research, 44*, 727–749.

Fraser, A., & Wilcox, K. J. (1979). Perception of illusory movement. *Nature, 281*, 565–566.

Faubert, J., & Herbert, A. M. (1999). The peripheral drift illusion: A motion illusion in the visual periphery. *Perception, 28*, 617–621.

Kitaoka, A. (1998). <http://www.ritsumei.ac.jp/akitaoka/cushione.html>.

Kitaoka, A. (2003). <http://www.psy.ritsumei.ac.jp/akitaoka/rotsnakee.html>.

Kitaoka, A. (2006). The effect of color on the optimized Fraser–Wilcox illusion. Gold prize at the 9th L'OR+AL Art and Science of Color Prize.

Kitaoka, A., & Ashida, H. (2003). Phenomenal characteristics of the peripheral drift illusion. *Vision, 15*, 261–262.

Kitaoka, A., & Ashida, H. (2004). A new anomalous motion illusion: The central drift illusion. In *Winter Meeting of the Vision Society of Japan*.

Kuriki, I., Ashida, H., Murakami, I., & Kitaoka, A. (2008). Functional brain imaging of the rotating snakes illusion by fmri. *Journal of Vision, 8*(10), 1–10.

Morgan, M. J., & Casco, C. (1990). Spatial filtering and spatial primitives in early vision: An explanation of the Zöllner–Judd class of geometrical illusions. *Proceedings of the Royal Society, London B, 242*, 1–10.

Morgan, M. J., & Moulden, B. (1986). The Münsterberg figure and twisted cords. *Vision Research, 26*(11), 1793–1800.

Murakami, I. (2004). Correlations between fixation stability and visual motion sensitivity. *Vision Research, 44*, 251–261.

Murakami, I., & Cavanagh, P. (1998). A jitter after-effect reveals motion based stabilization of vision. *Nature, 395*, 798–801.

Murakami, I., Kitaoka, A., & Ashida, H. (2006). A positive correlation between fixation instability and the strength of illusory motion in a static display. *Vision Research, 46*, 2421–2431.

Ross, J., Morrone, M., Goldberg, M., & Burr, D. (2001). Changes in visual perception at the time of saccades. *Trends in Neurosciences, 24*(2), 113–121.

Shi, B. E., Tsang, E. K. C., & Au, P. S. P. (2004). An on–off temporal filter circuit for visual motion analysis. In *ISCAS* (Vol. 3, pp. 85–88).

Shinomori, K., & Werner, J. (2003). Senescence of the temporal impulse response to a luminous pulse. *Vision Research, 43*, 617–627.

Shinomori, K., & Werner, J. (2006). Impulse response of an S-cone pathway in the aging visual system. *Journal of the Optical Society of America A, 23*(7), 1570–1577.

Watson, A. B., & Ahumada, A. J. (1985). Model of human visual motion sensing. *Journal of the Optical Society of America, 2*, 322–342.

Please cite this article in press as: Fermüller, C., et al. Illusory motion due to causal time filtering. *Vision Research* (2009), doi:10.1016/j.visres.2009.11.021