

# Motion clouds: model-based stimulus synthesis of natural-like random textures for the study of motion perception

Paula Sanz Leon, Ivo Vanzetta, Guillaume S. Masson, and Laurent U. Perrinet

*Institut de Neurosciences de la Timone, Centre National de la Recherche Scientifique/Aix-Marseille University, Marseille, France*

Submitted 24 August 2011; accepted in final form 12 March 2012

**Sanz Leon P, Vanzetta I, Masson GS, Perrinet LU.** Motion clouds: model-based stimulus synthesis of natural-like random textures for the study of motion perception. *J Neurophysiol* 107: 3217–3226, 2012. First published March 14, 2012; doi:10.1152/jn.00737.2011.—Choosing an appropriate set of stimuli is essential to characterize the response of a sensory system to a particular functional dimension, such as the eye movement following the motion of a visual scene. Here, we describe a framework to generate random texture movies with controlled information content, i.e., Motion Clouds. These stimuli are defined using a generative model that is based on controlled experimental parametrization. We show that Motion Clouds correspond to dense mixing of localized moving gratings with random positions. Their global envelope is similar to natural-like stimulation with an approximate full-field translation corresponding to a retinal slip. We describe the construction of these stimuli mathematically and propose an open-source Python-based implementation. Examples of the use of this framework are shown. We also propose extensions to other modalities such as color vision, touch, and audition.

low-level sensory systems; eye movements; optimal stimulation; natural scenes; motion detection; Python

ONE OF THE OBJECTIVES OF SYSTEM neuroscience is to understand how sensory information is encoded and represented in the central nervous system, from single neurons to population of cells forming columns, maps, and large-scale networks. Unveiling how sensory-driven behaviors such as perception or action are elaborated implies to decipher the role of each processing stage, from peripheral sensory organs up to associative sensory cortical areas. There is a long tradition of probing each of these levels using standardized stimuli of low dimension and simple statistics. They are based on a powerful, but stringent, theoretical approach that considers the visual system as a spatiotemporal frequency analyzer (Graham 1979; Watson et al. 1983). Accordingly, visual neurons have long been tested with drifting gratings to characterize both their selectivities and some of nonlinear properties of their receptive fields (DeValois and DeValois 1988). A similar approach was applied at both mesoscopic and macroscopic scales to define functional properties of cortical maps (e.g., Blasdel and Salama 1986; Ts'o et al. 1990) and areas (e.g., Henriksson et al. 2008; Singh et al. 2000), respectively. A more recent trend has been to consider sensory pathways as complex dynamical systems. As such, these are able to process high-dimensional sensory inputs with complex statistics such as encountered during natural life. As a consequence, the objective is to understand how the visual brain encodes and processes natural visual

scenes (Dan et al. 1996). This has led to new theoretical approaches of neuronal information processing (Field 1999), as well as to the search for new sets of stimuli for measuring neuronal responses to complex sensory inputs (see Touryan 2001; Wu et al. 2006). Controversial opinions have been proposed on whether natural scenes and movies should be used straightforwardly for visual stimulation as in Felsen and Dan (2005) or whether one should rather develop new sets of “artificial” stimuli. Importantly, the latter approach has the advantage of being relatively easy to parametrize and to customize at different spatial and temporal scales (Rust and Movshon 2005). In brief, it has become a critical challenge to elaborate new visual stimuli that fulfill these two constraints: being both efficient and relevant to probe high-dimension dynamical systems on the one hand and, on the other, being easily tailored so that they can be used to conduct quantitative experiments at different scales, from single neuron to behavior.

Here, our aim is to provide such a set of stimuli cast into a well-defined mathematical framework. We decided to focus on motion detection, as a good illustration for the search for optimal high-dimension stimuli. Visual motion processing is critically involved in several essential aspects of low and middle vision such as scene segmentation, feature integration and object recognition (see Braddick 1993; Bradley and Goyal 2008; Burr and Thompson 2011 for reviews). It also provides essential aspect of visual information for motor systems such as speed and direction of moving objects, as well as about self motion. Lastly, it is one of the few systems for which an integrated approach from single neuronal activity to complex behaviors can be achieved using nearly identical experimental conditions to elucidate the neural bases of perceptual decisions (Newsome 1997) and motor responses (see Masson and Ilg 2010 for a collection of examples).

However, motion perception is highly dynamical and the classical toolbox of standard motion stimuli (such as dots, bars, gratings, or plaids) is now largely outdated and insufficient to understand how the primate brain achieves visual motion processing with both high efficiency and short computing time. To be optimal, a new set of stimuli should be rooted in theoretical assumptions about how motion information is processed (Watson and Turano 1995). A large bulk of experimental and theoretical evidences supports the view that local motion information is extracted through a set of spatiotemporal frequency analyzers, the outputs of which are then integrated to yield motion direction and amplitude (Adelson and Bergen 1985; Simoncelli and Heeger 1998). However, we still lack a deep understanding of several linear (L) and nonlinear (NL) operations needed to extract the global motion from the local

Address for reprint requests and other correspondence: L. U. Perrinet, Institut de Neurosciences de la Timone, CNRS/Aix-Marseille Univ., Marseille, France (e-mail: Laurent.Perrinet@univ-amu.fr).

luminance changes (see Derrington et al. 2004 for a recent review). For instance, it remains unclear how MT neurons can encode speed and direction independently of the local spatiotemporal frequency or orientation content of the image (see Bradley and Goyal 2008 for a recent review). It is also hard to predict MT neurons responses to dense noise patterns or natural scenes from their spatiotemporal frequency selectivity as explored with low-dimension stimuli (Nishimoto and Gallant 2011; Priebe et al. 2006). Lastly, neuronal responses to natural movies are more reliable and sparse than when driven by low-dimensional stimuli such as drifting gratings (Vinje and Gallant 2000).

To overcome these limits, several recent studies have proposed that linearly combining several frequency channels can partly account for pattern direction and speed selectivity (Nishimoto and Gallant 2011; Rust et al. 2006, 2005). Still, such multistage L-NL models (Heeger et al. 1996; Simoncelli and Heeger 1998) fail to account most of the response properties seen with natural scenes (see Carandini et al. 2005 for a review). One key issue is to understand how motion information gathered at different scales is normalized and weighted before integration as in the divisive normalization version of the L-NL model of motion detection. Natural-like stimuli are good probes to further explore the performance of these models (Schwartz and Simoncelli 2001). However, “raw” natural scenes have the major drawback that information content is poorly controlled: their dimensionality is extremely high and the interstimulus variability in the information content with respect to sensory parameters is large (Rust and Movshon 2005). Popular alternatives to natural scenes are dense and sparse noise. However, those are often irrelevant to the sensory system and most often fail to drive strong neuronal responses. Here, we explore a new approach for the characterization of the first-order motion system. Our stimuli are equivalent to a subclass of random phase textures (RPTs; Galerne et al. 2010), which are increasingly attracting interest in exploring neural mechanisms of texture perception (e.g., Solomon et al. 2010).

The study is organized as follows. In METHODS, we first recall the main properties of RPTs as originally defined in computer vision for texture analysis. Next, we define their dynamical version, called thereafter Motion Clouds (MCs), and provide their complete mathematical formulation. We briefly describe the architecture of our implementation, all technical details being available as Supplementary Material, including the source code (Supplemental Material for this article is available online at the *J Neurophysiol* website). In RESULTS section, we illustrate the practical use of MCs for studying several long-lasting problems of visual motion processing such as two-dimensional motion integration, motion segmentation and transparency. For each, we will compare the usefulness of MCs relative to existing low-dimension stimuli. Finally, we discuss how this approach can be generalized to different aspects of visual system identification.

## METHODS

**RPTs and natural retinal motion.** First, RPTs are defined as generic random motion textures that are optimal for probing luminance-based visual processing. Most of the information present in a given dynamical image can be divided into its geometry (that is the outline of the objects it represents) and its distribution of luminance in space and time (Jasinski et al. 1992; Neri et al. 1998; Perrone and Thiele 2001,

2002). In the spatiotemporal Fourier space, this is well separated between the phase and the absolute amplitude spectra, respectively (Oppenheim and Lim 1981). This can be easily seen by gradually perturbing the phase spectrum of a natural scene: while form is progressively lost, its global motion information remains essentially unchanged (see Figure 1). This invariance with respect to phase shuffling in the Fourier domain is generally considered to be characteristic of the first-order motion stage (Derrington et al. 2004; Lu and Sperling 2001). We next formally define a linear generative model for the synthesis of such natural-like moving textures. Most generally, we can describe luminance at position  $(x, y)$  and time  $t$  as the scalar  $I(x, y, t)$  that is the sum of the contribution of a set of basis functions:

$$I(x, y, t) = \sum_k a_k \cdot G(x, y, t; \beta_k) \quad (1)$$

The function  $G$  defines the family of basis functions where each basis function is defined by parameters  $\beta_k$ . Scalars  $a_k$  give the relative amplitude for each basis function and therefore will change for each individual image  $I$ , while the parameters  $\beta_k$  are fixed for a set of stimuli. The advantage of this generative model is to separate the temporal scale of coding a specific moving stimulus (represented by the scalars  $a_k$ ) from the temporal scale of a whole set of stimuli (as represented by the  $\beta_k$ ). Efficient coding strategies use such generative models by optimizing scalars  $a_k$  knowing a fixed set of basis functions  $\beta_k$ . Note that finding the optimal set  $a_k$  knowing this linear generative model and the image  $I$  is in general a nonlinear problem (it is the coding problem). When the set given by  $\beta_k$  is large, this problem becomes difficult. In that context, divisive normalization gives a fair account for this problem using for its solution second-order correlations across basis functions (Schwartz and Simoncelli 2001). On a slower temporal scale, such model is used in neural modeling for studying the emergence of receptive fields by optimizing  $\beta_k$ , such as by using a Bayesian framework (Perrinet 2010).

In Fourier space, by linearity:

$$\mathcal{F}(I)(f_x, f_y, f_t) = \sum_k a_k \cdot \mathcal{F}(G)(f_x, f_y, f_t; \beta_k) \quad (2)$$

where  $\mathcal{F}$  is the Fourier transform. Here, we will use a fixed set of spatiotemporal Gabor kernels to implement localized, moving grating-like textures as they are known to efficiently code for natural images. For this particular set, parameters are defined as the peak's spatiotemporal position  $\{x_k, y_k, t_k\}$ , velocity (direction and speed), orientation, and scale. We may write the translation of each component using the shift operator in the Fourier domain:

$$\mathcal{F}(I)(f_x, f_y, f_t) = \sum_k a_k \cdot \mathcal{E}_{\tilde{\beta}_k}(f_x, f_y, f_t) \cdot e^{-2i\pi(f_x x_k + f_y y_k + f_t t_k)} \quad (3)$$

where  $\tilde{\beta}_k$  denotes the parameters without positions  $\{x_k, y_k, t_k\}$ . In general, parameters  $\beta_k$  have some statistical regularities in Fourier space: for instance, velocity, orientation and scale parameters are correlated in space and time (Lagae et al. 2009; Lewis 1984). This defines an average spectral density envelope that we denote as  $\mathcal{E}_{\tilde{\beta}}$  and that is characteristic of the particular class of natural images that is coded (Torralba and Oliva 2003). We use this generative model to define RPTs and their MCs derivative that can be seen as first-order motion textures. If we shift randomly and independently the central position of edges (see Fig. 1) and if this perturbation is stochastically independent from the distribution of the others parameters, one can describe the image by the following mean-field equation on its Fourier transform:

$$\mathcal{F}(I)(f_x, f_y, f_t) = \mathcal{E}_{\tilde{\beta}}(f_x, f_y, f_t) \cdot \sum_k a_k \cdot e^{-2i\pi(f_x x_k + f_y y_k + f_t t_k)} \quad (4)$$

By consequence, the envelope is modulated by a stochastic spectrum that is defined at any point in Fourier space as the sum of random independent variables with the same distribution and variance. By virtue of the central limit theorem, we may define the set of stimuli  $I$  as the random sequences generated by  $I$ ) an average envelope  $\mathcal{E}_{\tilde{\beta}}$ , 2)

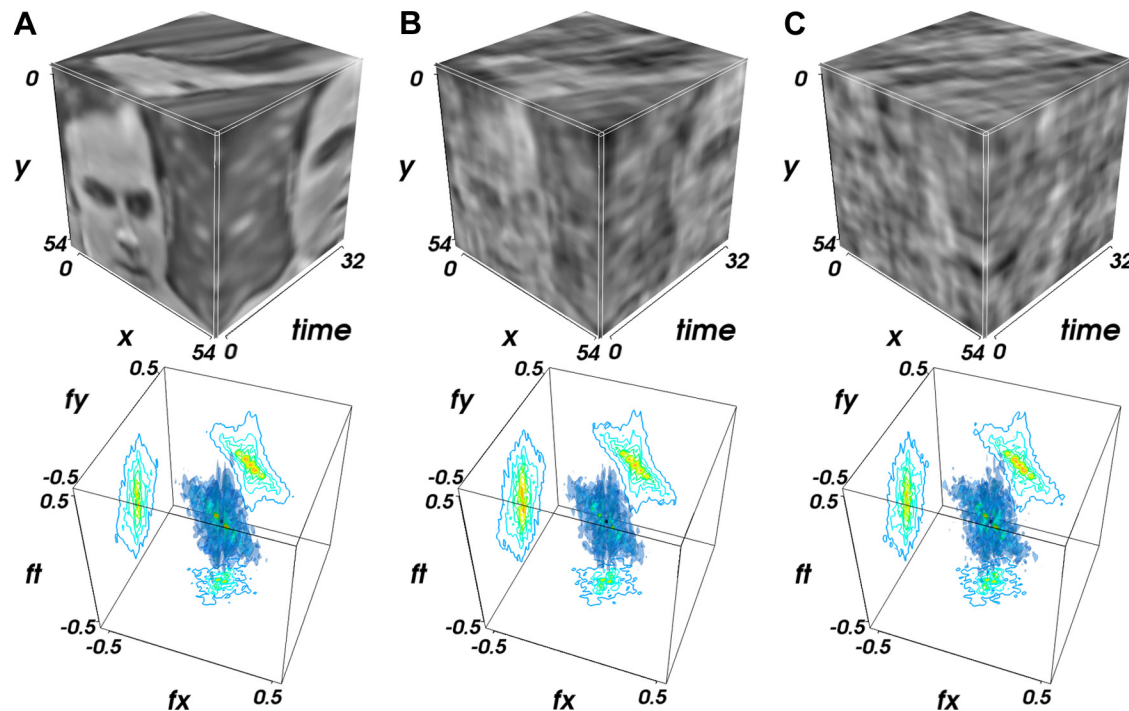


Fig. 1. *A, top*: a natural movie with the main motion component consisting of a horizontal, rightward full-field translation. Such a movie would be produced by an eye movement with constant mean velocity to the left (negative  $\times$  direction), plus some residual, centered jitter noise in the motion-compensated natural scene. We represent the movie as a cube, the  $(x, y, t = 0)$  face of which corresponds to the first frame, and the  $(x, y = 0, t)$  face shows the rightward translation motion as diagonal stripes. As a result of the horizontal motion direction, the  $(x = 54; y, t)$  face is a reflected image of the  $(x, y, t = 0)$  face, contracted or dilated depending on the amplitude of motion. *Bottom*: corresponding Fourier energy spectrum, as well as its projections onto 3 orthogonal planes. For any given point in frequency space, the energy value with respect to the maximum is coded by 6 discrete color iso-surfaces (i.e., 90, 75, 50, 25, 11, and 6% of peak). Amplitude of the Fourier energy spectrum has been normalized to 1 in all A–C, and the same conventions used here apply to all following figures. *B and C*: image is progressively morphed (A through B to C) into a random phase texture (RPT) by perturbing independently the phase of each Fourier component. *Top row*: form is gradually lost in this process; *bottom row*: most motion energy information is preserved, as it is concentrated around the same speed plane in all 3 cases (the spectral envelopes are nearly identical).

a normally distributed, independent and identically distributed amplitude spectrum  $A$ , and 3) a uniformly distributed phase spectrum  $\Phi$  in  $[0; 2\pi]$  that is to RPT (Galerne et al. 2010) trivially extended to the spatiotemporal domain:

$$\mathcal{F}(I) = \mathcal{E}_{\beta} \cdot A \cdot e^{i\Phi} \quad (5)$$

As noted by (Galerne et al. 2010), the main visual ingredients of RPTs are the envelope spectrum and the random phase spectrum, while  $A$  has little perceptual effect. Indeed, removing the random amplitude spectrum, we still have a random fluctuation of the sign of each Fourier coefficient. From the central limit theorem, under the condition that the number of mixed components is large enough, the coefficient spectrum resulting from the mixing described by

$$\mathcal{F}(I) = \mathcal{E}_{\beta} \cdot e^{i\Phi} \quad (6)$$

will still be given by a normal distribution and correspond to a correct generative model for RPTs as in Eq. 5. Stimuli corresponding to such equations correspond to band-limited filtering of white noise, that is, to a white noise (in space and time) linearly filtered by the kernel  $K = \mathcal{F}^{-1}(\mathcal{E}_{\beta})$  corresponds to the average impulse response of the texture (see Fig. 2).

This class of random, textured, dynamical stimuli have several advantages over classical narrow bandwidth, low entropy stimuli, such as gratings or combinations of gratings. First, by varying the weight of each Fourier coefficient, we can vary its content and probe different types of motion integration models. Second, we can generate several different series of stimuli with different randomization seeds, while keeping all other parameters constant. Third, we can play with the bandwidth along each dimension to titrate the role of distributions

of frequencies onto neuronal or behavioral responses. Fourth, we can reproduce the statistics of natural images by controlling the global envelope in Fourier space. Fifth, stochastic properties are generated only by varying the phase spectrum, without the need for adding noise component to the motion stimulus or controlling lifetime of individual features. Below we shall discuss several examples of the experimental usability of such stimuli. Stimuli similar to RPT have been already used. This was first formalized for the generation of natural-like static textures in computer vision (Lewis 1984) such as procedural or Perlin textures and is still largely used (Lagae et al. 2009). Mathematically, the resulting patterns are related to the morphogenesis studies pioneered by (Turing 1952). Such static textures were used in psychophysics (Essock et al. 2009), in neurophysiology, for instance to study sensitivity of V1 neurons to dynamical expansion (Wang and Yao 2011) or nonlinear properties of nonclassical receptive fields of primate MT neurons (Solomon et al. 2010). It is worth noting that a similar stimulus design was proposed for investigating another sensory system, i.e., audition (Klein et al. 2000; Rieke et al. 1996).

*MCs as one particular type of RPT.* Defining optimal motion stimuli to probe the first-order, luminance-based motion system is a nontrivial problem. A straightforward approach is to generate static textures and to generate sequences as an exact, full-field translation of this static texture (Drewes et al. 2008). However, this approach is not generic enough. In particular, it lacks the possibility to vary the distribution of speeds being present in a given movie, a parameter that might be crucial to study precision and robustness of motion processing for perception or eye movements. MCs can be defined as the subset of RPTs that results from a generative model inspired by a rigid translation at central velocity of a large texture filling the whole visual field. This generative model will be specified by a central velocity



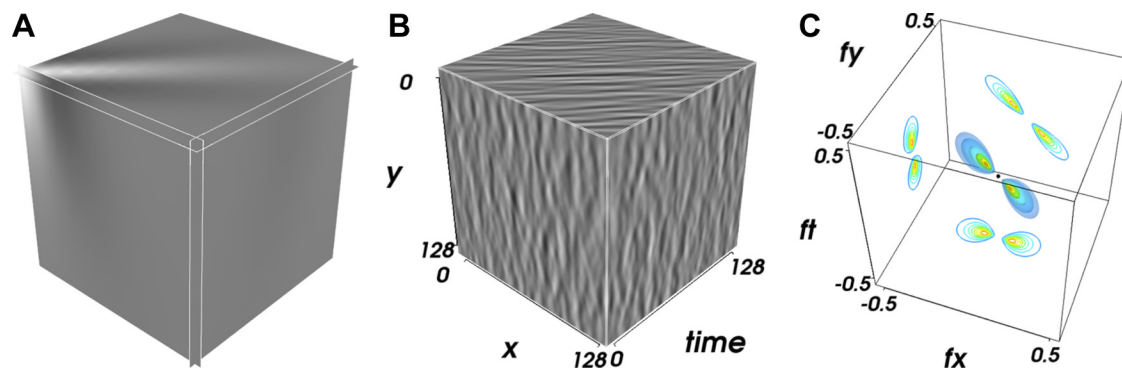


Fig. 2. From an impulse to a Motion Cloud (MC). *A*: movie corresponding to a typical “edge”, i.e., a moving Gabor patch that corresponds to a localized grating. The Gabor patch being relatively small, for clarity, we zoomed 8 times into the nonzeros values of the image. *B*: by densely mixing multiple copies of the kernel shown in *A* at random positions, we obtain a RPT (see Supplemental Movie S1). *C*: we show here the envelope of the Fourier transform of kernel *K*: inversely, *K* is the impulse response in image space of the filter defined by this envelope. Due to the linearity of the Fourier transform, apart from a multiplicative constant that vanishes by normalizing the energy of the RPT to 1, the spectral envelope of the RPT in *B* is the same as the one of the kernel *K* shown in *A*:  $\mathcal{E}_{\tilde{\beta}} = \mathcal{F}(K)$ . Note that, the spectral energy envelope of a “classical” grating would result in a pair of Dirac delta functions centered on the peak of the patches in *C* (the orange “hot-spots”). MCs are defined as the subset of such RPTs the main motion component of which is a full-field translations and thus characterized by spectral envelopes concentrated on a plane.

$\vec{V}$  for the full-field translation, plus random independent perturbations of velocities around the central velocity, given by a bandwidth  $B_V$ . As a consequence, the spectral distribution of energy of such a sequence is centered on and squeezed onto a plane defined by the normal vector  $\vec{V}$ . The L-NL models of direction selectivity match the spatiotemporal properties of V1 or MT neuron receptive field with this plane. Phase information is concealed using the squared sum from the activity of receptive fields of odd and even phase (Adelson and Bergen 1985). By definition, MCs using a similar envelope as given by the spatiotemporal filtering properties of V1 or MT neurons are thus equivalently defined as the set of stimuli that are optimally detected by these energy detectors (Nishimoto and Gallant 2011). Moreover, they are also optimal for motion coding in the information-theoretic sense, since they maximize entropy (Field 1994) compared with the presentation of a simple kernel *K* as in Watson and Turano (1995). Similar random textures as MCs have been generated by displaying a rectangular grid of Gabor patches with random orientations and directions (Scarfe and Johnston 2010). However, such a regular grid introduces some geometrical information that may interfere with the processing of motion, as opposed to RPTs. Our MCs are more similar to the texture stimuli introduced by Schrater et al. (2000) or to the dynamical displays designed by Tsuchiya and Braun (2007). Below we propose one well-defined mathematical formalization for our MCs before presenting a solution for their implementation in psychophysical toolboxes.

#### MATHEMATICAL DEFINITION OF MCS

We define MCs as RPTs that are characterized by several key features. First, first-order motion information is independent on changes in the phase of the Fourier coefficients of image sequences since it is contained in the amplitude of the spectral coefficients (Eq. 5).

$$I(x, y, t) = F^{-1} \left\{ \mathcal{E}_{\tilde{\beta}}(f_x, f_y, f_t) \cdot e^{i\phi(f_x, f_y, f_t)} \right\} \quad (7)$$

Second, in Fourier space, full-field, constant, translational motions correspond to an envelope  $\mathcal{E}_{\tilde{\beta}}$  the distribution of which is concentrated on a speed plane (a plane in Fourier space that contains the origin). Third, the distribution of the spectral envelope  $\mathcal{E}_{\tilde{\beta}}$  is defined as a Gaussian. This is explained by the fact that Gabor filters have a Gaussian envelope and thus an optimal spread in Fourier space (Marcelja 1980). As such they are well known models of simple cells in the primary visual

cortex (Daugman 1980) that can describe the most salient properties of receptive fields and their tuning for spatial localization, orientation, and spatial frequency selectivity (e.g., Jones and Palmer 1987). Moreover, Lee (1996) derived the conditions under which a set of two-dimensional Gabor wavelets are a suitable image basis for a complete representation of any image. This was further extended to the case of sequences with a known motion (Watson and Turano 1995) and therefore constitutes an accurate set for studying first-order motion. In summary, envelopes of MCs are essentially Gaussian distributions that are concentrated close to a speed plane (see Fig. 2). Equivalently, these characteristics define MCs as dense random mixing of spatiotemporal Gabor filters with similar speeds. The implementation presented herein is based on a simplified parametrization of the envelope of the amplitude spectrum. Given that speed, radial frequency, and orientation spread are independent, we can parametrize different types of MCs based on a factorization of each component.

$$\mathcal{E}_{\tilde{\beta}} = V_{(V_x, V_y, B_V)} \times G_{(f_0, B_f)} \times O_{(\theta, B_\theta)} \times C_{(\alpha)} \quad (8)$$

where all envelope parameters are given in their sublabel and envelopes correspond respectively to the speed plane, the frequency, and the orientation tuning along this plane: 1) For the speed envelope  $V$  (and thus the speed plane) while one parameter defines the bandwidth  $B_V$  of this plane as we jitter the mean motion  $\vec{V}$ . Varying these parameters allows to study the response of motion detectors to different speeds and amounts of velocity noise. 2) Projected onto the speed plane, we can define *a*) the radial frequency envelope  $G$  with two parameters that set its mean value  $f_0$  and bandwidth  $B_f$ ; also, *b*) the orientation envelope  $O$  is defined by two parameters: mean orientation  $\theta$  and bandwidth  $B_\theta$ . In both cases, the two parameters can be thought as defining the nominal value and the uncertainty of each respective component of motion information. 3) An additional envelope  $C$  is parameterized by  $\alpha$ . It tunes the overall shape of the envelope similarly to what is observed in natural images. Note that one can modify the parameters of each envelope independently and moreover, by the commutativity of the product operation, the order of the envelopes is arbitrary. It shall further be noticed that the actual

values of each of these parameters can be set based on known properties of the biological system to be investigated for each level of observation. We will now detail each of them.

**Speed envelope.** The first axis of a MC is its speed component. Let us first recall that the Fourier transform of a static image with a global translation motion is the Fourier transform of the static image (that lies in the  $f_t = 0$  plane) tilted on a plane perpendicular to  $\vec{v} = (V_x, V_y)$  defined as:

$$V_x f_x + V_y f_y + f_t = 0 \quad (9)$$

The orientation and tilt of the plane are determined by the direction and speed of motion, respectively. For larger  $v = \|\vec{v}\| = \sqrt{V_x^2 + V_y^2}$  the tilt becomes greater. To model speed variability (jitters) within a MC, we shall assume that motion varies slightly in both speed axes (i.e., direction and amplitude). Such envelope is given for instance by:

$$V_{(V_x, V_y, B_f)}(f_x, f_y, f_t) = \exp\left[-\frac{1}{2}\left(\frac{f_x \cdot V_x + f_y \cdot V_y + f_t}{B_f \cdot f_r}\right)^2\right] \quad (10)$$

where  $f_r = \sqrt{f_x^2 + f_y^2 + f_t^2}$  is the radial frequency.

**Radial frequency envelope.** The second characteristic of a MC is its radial frequency envelope. This is defined as the one-dimensional distribution of radial frequency using spherical coordinates in the Fourier domain. Indeed, by spherical symmetry, this radial frequency envelope is then independent to motion and orientation tuning. An intuitive description of this envelope is a Gaussian distribution along this radial dimension, as it is often encountered to describe the frequency component of Gabor filters. An inconvenient of Gabor functions is the fact that their sum is not perfectly null. This shows up in Fourier space as a nonzero value at the origin. To overcome this issue we use the log-Gabor filters (Fischer et al. 2007). A second advantage of using log-Gabor filters is that they better encode natural images (Field 1987). We thus build a spatial frequency band Gaussian filter that depends on the logarithm of the spatial radial frequency. We define  $f_0$  as the mean radial frequency and  $B_f$  as the filter's bandwidth.

$$G_{(f_0, B_f)}(f_x, f_y, f_t) = \frac{1}{f_r} \exp\left\{-\frac{1}{2}\left[\frac{\ln\left(\frac{f_r}{f_0}\right)}{\ln\left(\frac{f_0 + B_f}{f_0}\right)}\right]^2\right\} \quad (11)$$

**Orientation envelope.** The third property of a MC is its orientation envelope. Oriented structures in space-time yield oriented structures in the Fourier domain. Thus, the orientation component of the spectrum is given by the function  $v = \|\vec{v}\| = \sqrt{V_x^2 + V_y^2}$ . It is defined by a density function located at a mean orientation  $\theta$  and the spread of which is modeled using a Von-Mises distribution with parameter  $B_\theta$  that represents its bandwidth centered on the symmetric with respect to the origin:

$$G(f_x, f_y) = \exp\left(\frac{\cos(\theta_f - \theta)}{B_\theta}\right) + \exp\left(\frac{\cos(\theta_f - \theta - \pi)}{B_\theta}\right) \quad (12)$$

where  $\theta_f = \arctan(f_y, f_x)$  is the angle in the Fourier domain. Note again that this envelope is independent upon both speed and radial tuning. We define its bandwidth using the standard deviation  $B_\theta$ . If  $B_\theta$  has a small value, a highly coherent orientation pattern is generated.

**Spectral color.** An important property of a MCs is their global statistics. Therefore, the average shape of its power spectrum must be controlled. It has been shown that the average power spectrum of natural scene follows a power law (Field 1987):

$$C_\alpha(f_x, f_y, f_t) = \frac{1}{f_R^\alpha} \quad (13)$$

where  $\alpha$  is usually set within the range  $0 < \alpha < 2$ . By analogy with the color terminology used to characterize noise patterns, we call this function color envelope. The simplest stimulus that can be built with our model is filtered spatial noise as a function of the power (exponent) factor  $\alpha$ . In our model, we assume that the spectrum shape is independent of orientation and varies solely as a function of a radial frequency ( $f_R$ ), defined as in Schrater et al. (2000) by:

$$f_R = \sqrt{f_x^2 + f_y^2 + \frac{f_t^2}{f_{t_0}^2}} \quad (14)$$

The factor  $f_{t_0}$  is a normalization factor and is associated to a normalized stimulus velocity.

The color envelope weights the different frequency channels according to the statistics of natural images and therefore is optimal regarding the sensitivity of the primate visual system to the different spatiotemporal frequencies (Atick 1992). In the examples given below, we choose  $\alpha = 1$ , corresponding to a pink noise distribution. Note that this particular value allows for the marginal distribution integrated over all orientations to coincide with the speed and frequency envelope. Qualitatively, this global envelope does not change neither motion nor texture appearance of a MC since it has no preferred speed, frequency, or orientation. This is true in particular for MCs with a relatively narrow envelope in Fourier space. When using larger bandwidth values of the radial frequency distribution  $B_f$ , the shape of the global envelope becomes more important.

**Implementation.** Since our objective is to provide a new set of stimuli for conducting neurophysiological or psychophysical experiments, we must propose a framework for generating and displaying MCs under well controlled parameter settings. With the use of standard computer libraries, the theoretical framework described above can be implemented while taking into account technical constraints such as discretization and videographic displays. In the Supplementary Materials, we provide with the source code used to generate our calibrated MCs using Python libraries.

## RESULTS

To illustrate how MCs can be used to investigate different aspects of motion processing, we now describe some of their applications. We emphasize how classical stimuli, such as gratings or plaid patterns, can be conveniently represented as MCs. This last aspect is important: MCs can be seen as a single class of motion stimuli encompassing both low-dimension and complex dynamical stimuli. It becomes then possible to parametrically investigate the effects of spatiotemporal frequency content upon different stages of motion processing. It shall be noticed that all the following examples are chosen such as to fit the characteristics of visual motion systems; yet, the same logic

applies to other aspects of visual processing, such as texture or shape perception.

**MCs equivalents of classical stimuli.** Sinusoidal luminance gratings are defined by a small set of parameters (orientation, direction, frequency). This translates naturally into a set of MCs with the parameters that we defined: speed, orientation, frequency. In addition, we now have the choice of three extra parameters,  $B_v$ ,  $B_\theta$ , and  $B_f$ , that tune the bandwidth along each of these components, respectively (see Fig. 3A). It thus becomes possible to investigate spatial or temporal frequency, orientation, or direction selectivity, as well as the role of their respective tuning bandwidths.

With drifting gratings, perceived motion direction is necessarily defined perpendicular to its orientation. This is related to the aperture problem: translation of a one-dimensional elongated edge is ambiguous and its visual motion is compatible with an infinite number of velocity vectors (Movshon et al. 1985). A novel formulation of this problem can be designed by creating a MC the direction of which is not perpendicular to the main orientation and the orientation of which bandwidth is very narrow. Indeed, a classical motion detector would then be incapable of determining nonambiguously the speed plane that corresponds to such an envelope (see Fig. 3B).

MCs also encompass textures similar to random-dot kinematograms (RDKs). Usually, RDKs consist in a set of small dots drifting in a given direction and speed, each dot having a limited life time. This is similar to our original definition of MCs. Such pattern is defined in Eq. 1 with a kernel that would correspond to a Dirac delta function in space, a square ON and OFF function in time and a sparse set of coefficients  $a_i$ . Note that this kernel would

correspond to  $a$  at envelope on the speed plane with a bandwidth proportional to the inverse of the lifetime of dots. This is therefore controlled in MCs by the parameter  $B_v$ , and indeed, we observe that shorter values induced “features” that last longer. We stress, however, that MCs are necessarily equivalent to dense, not sparse, noise patterns.

Moreover, each MC is generated by a fully known, computer-generated noise. It is therefore possible to regenerate exactly the same stimulus by using the same seed in the random number generator. This property allows to investigate intertrial variability and thus the relative importance, for the system at hand, of external noise (measurement noise) and internal noise (uncertainty due to ambiguities and mixtures in the signal representation). This approach corresponds to the use of frozen noise stimuli (Mainen and Sejnowski 1995), that is, with a set of inputs for the visual system that is randomly generated but can be presented many times in a strictly identical manner.

**Comparing broad band and narrow band motion stimuli.** Varying the spatiotemporal frequency distribution from a grating-like stimulus to complex RPTs should be a powerful method for investigating neuronal selectivity and cortical maps of extra-striate areas. MCs (and other types of RPT patterns) shall be able to drive cortical neurons known for receiving converging inputs from several spatiotemporal frequency channels (Rust and Movshon 2005). We have considered the idea of generating stimuli to explore the effects of varying a single bandwidth parameter:  $B_f$ , while setting  $B_v$  and  $B_\theta$  to some fixed values (with relatively low values to get some precision along these components). We use the name broad band stimuli for the MCs with a large  $B_f$ , whereas narrow band stimuli are MCs

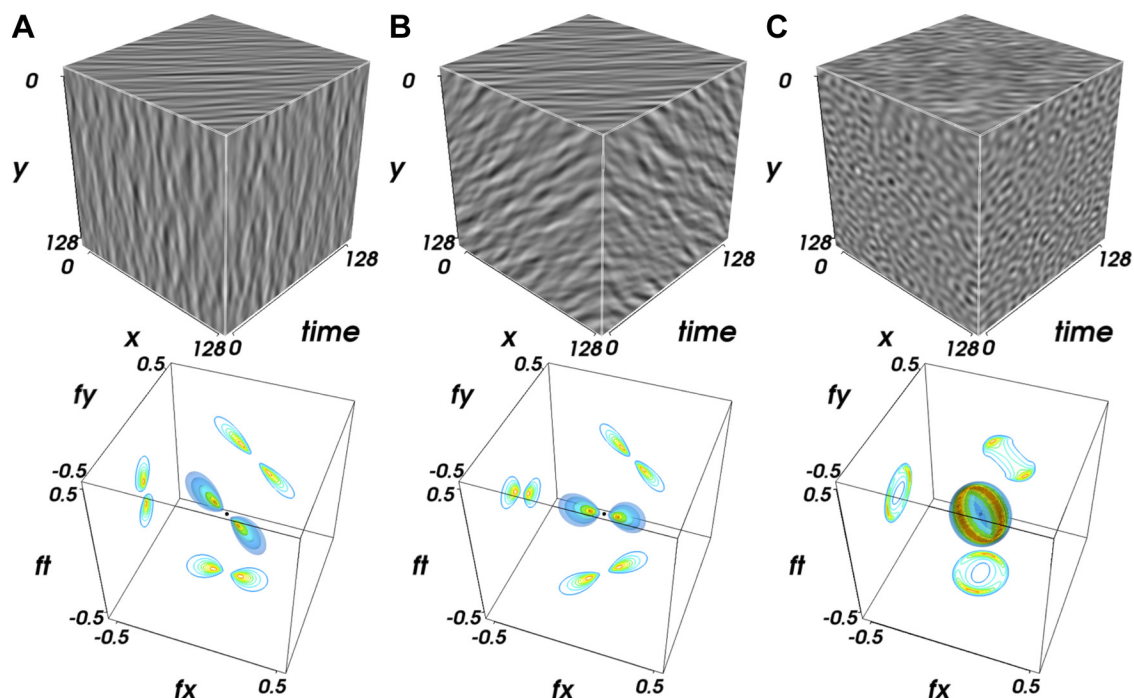


Fig. 3. Equivalent MC representations of some classical stimuli. A, *top*: a narrow-orientation-bandwidth MC produced only with vertically oriented kernels and a horizontal mean motion to the right (Supplemental Movie S1). A, *bottom*: spectral envelopes concentrated on a pair of patches centered on a constant speed surface. Note that this “speed plane” is thin [as seen by the projection onto the  $(f_x, f_t)$  face], yet it has a finite thickness, resulting in small, local, jittering motion components. B: a MC illustrating the aperture problem. *Top*: stimulus, having oblique preferred orientation ( $\theta = \pi/4$  and narrow bandwidth  $B_\theta = \pi/36$ ) is moving horizontally and rightwards. However, the perceived speed direction in such a case is biased towards the oblique downwards, i.e., orthogonal to the orientation, consistently with the fact that the best speed plane is ambiguous to detect (Supplemental Movie S2). C: a low-coherence random-dot kinematogram-like MC: its orientation and speed bandwidths,  $B_\theta$  and  $B_v$ , respectively, are large, yielding a low-coherence stimulus in which no edges can be identified (Supplemental Movie S3).



characterized by small  $B_f$  values. As illustrated in Fig. 3, *middle*, broad band and narrow band clouds are both symmetric, air foil-shaped volumes. However, the broadband envelope contains more frequency information than the narrow band one (Fig. 4). Therefore, it is thought to better represent natural images. Recently, we have used such MC stimuli to investigate how the visual system integrates different spatial frequency information levels, by varying  $B_f$  across a large range of spatial frequencies (Simoncini et al. 2012). The stimuli were displayed using Psychtoolbox v3 (Brainard 1997; Pelli 1997) for MATLAB (<http://psychtoolbox.org>) on a CRT monitor ( $1,280 \times 1,024 @ 100$  Hz). They covered  $47^\circ$  of visual angle at a viewing distance of 57 cm. We have used these stimuli to understand how two different visual behaviors, perceptual speed discrimination and reflexive tracking, would take advantage of presenting a single speed at different spatial scales. We found that the visual system pools motion information adaptively, as a function of constraints raised by different tasks. MCs were found to be useful to resolve problems associated with the integration of multiple spatial frequencies as they allow a precise control all variables related to speed and frequency content. In particular, previous studies have failed to understand how speed information is reconstructed across different spatial frequencies because the mixing of two, or more, gratings poses several perceptual problems. For instance, depending on the phase relationship between spatial frequency components, different interference patterns would appear, generating second-order motion in the same or opposite motion direction (Smith and Edgar 1990). Second, mixing sparse RDKs moving at the same speed but band-pass filtered at different spatial frequency results in complex patterns that can be perceived as being either coherent or transparent (Watson and Eckert 1994). The same difficulties have been encountered by neurophysiological studies trying to understand the origin of

speed selectivity in V1 complex cells (Priebe et al. 2006) or MT neurons (Priebe et al. 2003).

*Clouds with competing motions.* Low and midlevel visual integration and segmentation mechanisms have been extensively investigated with either combinations of gratings (i.e., plaid patterns) or random dot patterns with different directions, speed, and/or spatiotemporal components. Such plaid stimuli have been extensively studied and constitute an important pillar in motion detection theories, such as the separation between component and pattern cells in area MT (see Born and Bradley 2005; Burr and Thompson 2011; Movshon et al. 1985 for reviews). However, there has been a long-standing controversy about which information can be used in plaid patterns, such as component gratings, their product, or two-dimensional features called blobs that are generated at the intersection between component gratings (Derrington et al. 2004). It is also unclear how different direction and spatial frequency channels are mixed to create pattern direction selectivity (Rust et al. 2006). As explained above, MCs stimuli are by definition less susceptible to create interference patterns (or Moiré patterns) when mixed together. This is a striking difference with respect to classical low-entropy stimuli, such as gratings. Being able to mix together two textures with the same motion but different characteristic spatial frequencies is also critical to further study motion integration (e.g., single neuron selectivity: Rust and Movshon 2005; ocular tracking behavior: Masson and Castet 2002; motion perception: Smith and Edgar 1990). By contrast, it must be also possible to mix two textures with different motions to study the competition between integration and segmentation, leading to different percepts such as coherent or transparent motions. With the use of MCs, there is a further number of combinations that may be of interest for studying motion detection. We illustrate several possibilities in Fig. 5. Figure 5A shows a standard MC with added explicit noise,

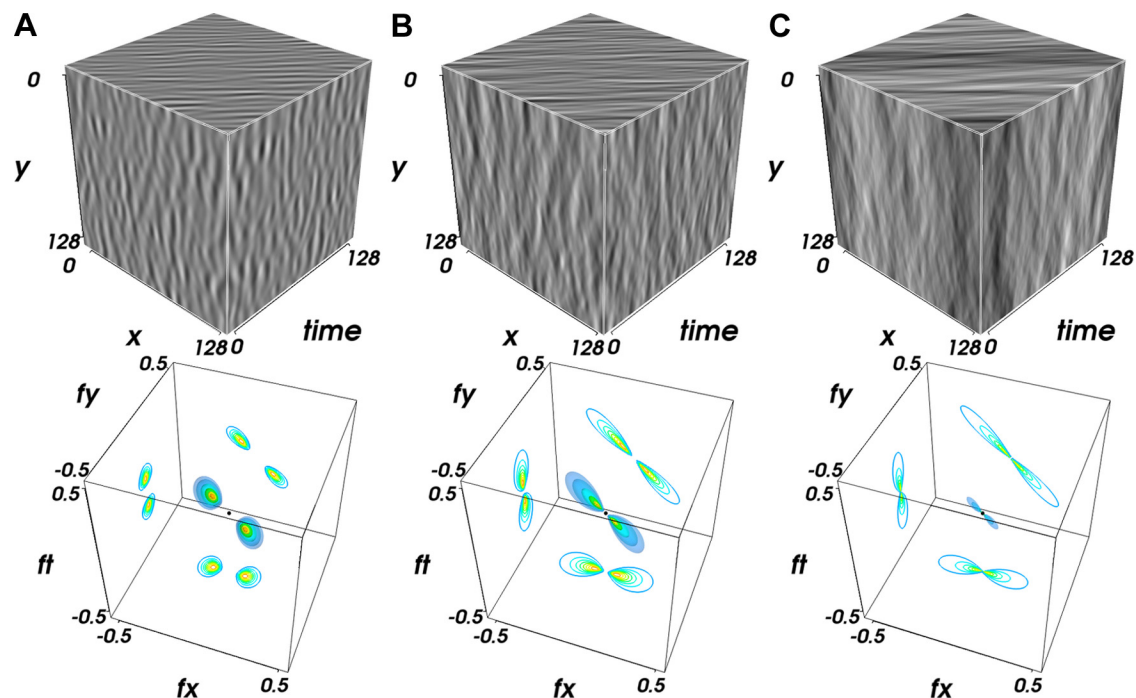


Fig. 4. Broadband vs. narrowband stimuli. From A through B to C, the frequency bandwidth  $B_f$  increases, while all other parameters (such as  $f_0$ ) are kept constant. The MC with the broadest bandwidth is thought to best represent natural stimuli, since, as those, it contains many frequency components. A:  $B_f = 0:05$  (Supplemental Movie S4). B:  $B_f = 0:15$  (Supplemental Movie S5). C:  $B_f = 0:4$  (Supplemental Movie S6).

corresponding to an envelope broadly centered around  $V = 0$ . Figure 5B illustrates the plaid-equivalent MC obtained by adding two MCs of same velocity but different orientations, similarly to a plaid stimulus. In the Fig. 5C, the two components have different velocities (here opposite ones) while all other parameters are identical. With standard gratings, such two gratings would interfere and create a counter-phase, flickering stimulus. With MCs, there is no such interference and the resulting stimulus has all desired energy distributed on both speed planes. By varying the relative direction of two, or more, components, it becomes possible to produce several transparent patterns and therefore to overcome a limit of classical motion stimuli such as gratings.

## DISCUSSION

In this article, we described the mathematical framework and provide the computer implementation of a set of complex stimuli that we call MCs. Those are an instantiation of a more generic class of stimuli called RPTs. These stimuli, presented herein in the context of visual motion perception, represent an attempt to fill the gap between simple stimuli (such as spots of light or sinusoidal gratings), stimulus ensembles consisting of simple stimuli (for instance, white noise patterns), and natural stimuli (Felsen and Dan 2005, Rust and Movshon 2005). Similar approaches have been used before in the case of motion detection (Schrater et al. 2000), but stimuli have been described in a somewhat incomplete and nonaccessible way. Here, our goal was to provide a complete and rigorous mathematical description of those stimuli, as well as tools for generating them. We have also given a few examples of different subset of MCs that could be used for probing detec-

tion, integration and segmentation stages at both psychophysical and neurophysiological levels. To conclude, we indicate a few future extensions and possible uses.

**Embedding spectral properties of natural images.** Both sensory and motor functions are natural tasks, and therefore it is essential to understand how they deal with natural stimuli. Following the core principles of Natural Systems Analysis (Geisler and Ringach 2009), we think it is possible to extend our model of natural stimulation to carry out different and more complex experiments in the visual system. Seminal work from Zeki (1983) showed that there exists a selectivity for color in higher visual areas such as V4 of the macaque monkey. Moreover, the work by Conway et al. (2007) shows that in the extra-striate cortex (V3, V4, and inferior temporal cortex) color is processed in terms of the full range of hues found in color space. The spatial structure is represented by “globs” that are clustered by color preference and organized as color columns. It is therefore important to develop an extension of MCs to be able to probe color vision. A first approach would consist in creating a simple colored MC using an RGB scheme (Galerne et al. 2010). In this case, we should add the same uniform random phase to each color channel. More realistically, a short medium long-cone scheme will have to be used, taking into account the cone fundamentals. Such color texture stimuli would permit to create a wide variety of new psychophysics experiments related to color perception.

**Exploiting phase parameters: towards a systematic exploration of the role of geometry.** The amplitude spectra of natural images are characterized by their  $1/f$  shape; in consequence, the global power spectrum cannot provide much information on any natural image that can be used, for instance, for fine pattern recognition or classification (e.g., Victor and Conte

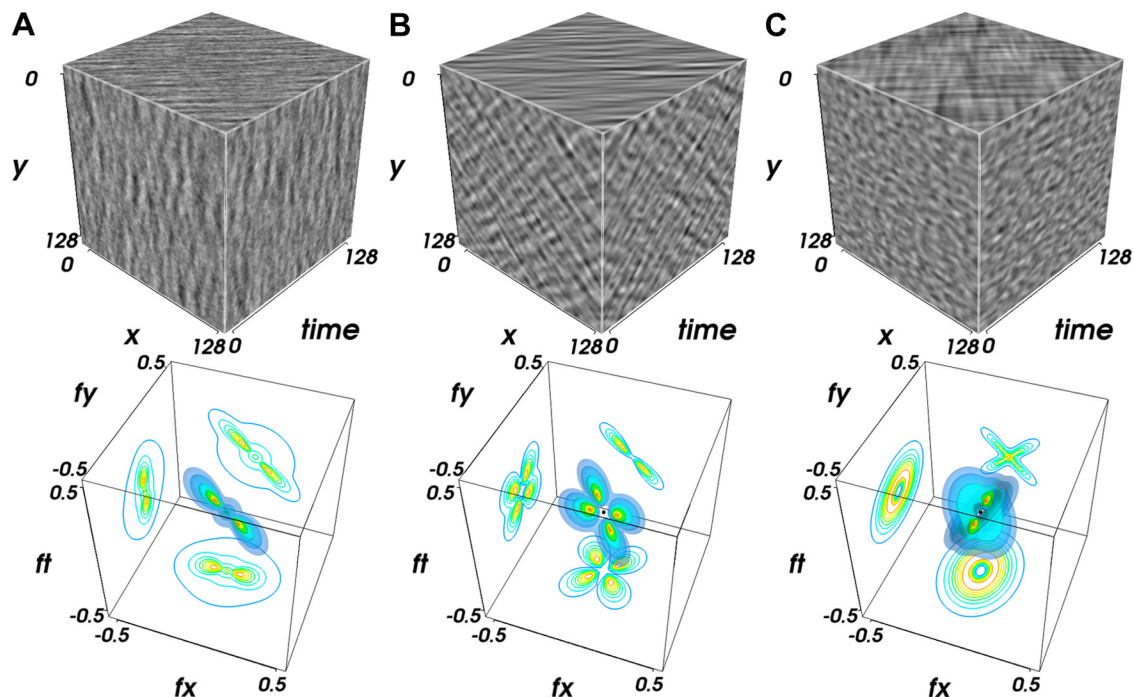


Fig. 5. Competing MCs. A: narrow-orientation-bandwidth MC with explicit noise. A red noise envelope was added to the global envelope of a MC with a bandwidth in the orientation domain (Supplemental Movie S7). B: 2 MCs with same motion but different preferred orientation were added together, yielding a plaid-like MC texture (Supplemental Movie S8). C: 2 MCs with opposite velocity directions were added, yielding a texture similar to a “counter-phase” grating (Supplemental Movie S9). Note that the crossed shape in the  $f_x - f_y$  plane is a signature of the opposite velocity directions, while two gratings with the same spatial frequency and in opposite directions would generate a flickering stimulus with energy concentrated on the  $f_t$  plane.



1996; see also Oliva and Torralba 2001; Torralba and Oliva 2003). Information contained within the phase spectrum is therefore the key to identifying the contents in the images, i.e., how shape is coded in natural images. This implies that the visual system must be sensitive to the phase structure of artificial stimuli or natural images at least at some spatial scales (Hansen and Hess 2006; Phillips and Todd 2010). This could be related to the rich representation of phase provided by the receptive field structure of visual neurons, from primary visual cortex up to extra-striate areas and further. We believe that MCs, and to a larger extent RPTs, are powerful tools to probe the properties of phase-sensitive mechanisms in neural populations. In the cases presented above, patterns have parameterized random phases: phase values are drawn from a uniform probability distribution. However, one can evidently draw these phases according to some structure known a priori, for instance, by correlating the phase of edges with similar orientations. This would progressively introduce collinearities in the set of stimuli, as needed to trigger short-range properties within the so-called association fields (Hess et al. 2003). By manipulating these parameters, we shall be able to control for the detailed information content in the different axes of the corresponding associative field, for instance the role of collinearity vs. cocircularity (Perrinet et al. 2011).

**Extension to increased complexity.** Random textured dynamical stimuli are generated as instances of few random variables defined by a generative model of synthesis. As a consequence, one may control the structural complexity of these synthetic textures by tuning the structure of the generative equations. In fact, the geometry of the visual world can be handled by using models to deal directly with the statistics of concurrent parameters, for instance edges or textures. For example, within each texture and/or edge class, low-dimensional models control the complexity of the stimuli using few meaningful inputs (regularity of edges, number of crossings, curvature of the texture flow, etc.). This complexity parameterization gives access to both the local geometry of the image (for instance its local orientation, frequency, scale, and granularity) and to more global integration properties (good continuation of edges and approximate periodicity).

These models can be assembled, thus leading to a rich content that mixes edge and texture patterns. It is believed that this hierarchical structure of generative models maps on a one-to-one basis with the structure of the visual system, from the detection of moving contrasts in the retina through edges in the primary visual cortex up to higher order attributes like motion and shape. By designing such models with increasing scales of complexity, it shall therefore be possible to specifically target structures in the low-level visual system, such as, respectively, V1, V2, V4, and MT. The generative framework underlying MCs can make an important contribution to this long-term goal.

**Exporting MCs to other sensory modalities.** Strong parallels have been drawn between visual and haptic processing for low level encoding of motion information for instance (Pei et al. 2011). Simple stimuli like drifting relief gratings, dynamic noise patterns or single elements such as lines and spots have already been used to investigate the properties of the somatosensory system. There is also a strong need to develop more sophisticated stimuli that can reproduce, in a controllable way, the statistics of natural somesthetic inputs. The theoretical

framework described in the present article may also be used to design somesthetic inputs using mechanical actuators to excite the vibrissal array of rats' whiskers (Jacob et al. 2008). This is another potential application of a set of stimuli bridging the gap between artificial and natural sensory input.

## ACKNOWLEDGMENTS

We thank Anna Montagnini, Frédéric Chavane, and Gabriel Peyré for comments on an earlier version of the manuscript.

## GRANTS

This work is supported by the European Union project Number FP7-269921, "BrainScaleS," and by VISAFIX Grant ANR-10-Blan-1432 from the Agence Nationale de la Recherche. P. Sanz Leon is supported by a Ministère de la Recherche Doctoral Fellowship.

## DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

## ENDNOTE

At the request of the author(s), readers are herein alerted to the fact that additional materials related to this manuscript may be found at the institutional website of one of the authors, which at the time of publication they indicate is: [http://invibe.net/LaurentPerrinet/MotionClouds]. These materials are not a part of this manuscript and have not undergone peer review by the American Physiological Society (APS). APS and the journal editors take no responsibility for these materials, for the website address, or for any links to or from it.

## REFERENCES

- Adelson EH, Bergen JR. Spatiotemporal energy models for the perception of motion. *J Opt Soc Am A*: 284–299, 1985.
- Atick JJ. Could information theory provide an ecological theory of sensory processing? *Network* 3: 213–252, 1992.
- Blasdel GG, Salama G. Voltage-sensitive dyes reveal a modular organization in monkey striate cortex. *Nature* 321: 579–585, 1986.
- Born RT, Bradley DC. Structure and function of visual area MT. *Ann Rev Neurosci* 28: 157–189, 2005.
- Braddick O. Segmentation versus integration in visual motion processing. *Trends Neurosci* 16: 263–268, 1993.
- Bradley DC, Goyal MS. Velocity computation in the primate visual system. *Nat Neurosci* 9: 686–695, 2008.
- Brainard DH. The psychophysics toolbox. *Spat Vis* 10: 433–436, 1997.
- Burr D, Thompson P. Motion psychophysics: 1985–2010. *Vision Res* 51: 1431–1456, 2011.
- Carandini M, Demb JB, Mante V, Tolhurst DJ, Dan Y, Olshausen BA, Gallant JL, Rust NC. Do we know what the early visual system does? *J Neurosci* 25: 10577–10597, 2005.
- Conway BR, Moeller S, Tsao DY. Specialized color modules in macaque extrastriate cortex. *Neuron* 56: 560–573, 2007.
- Dan Y, Atick JJ, Reid RC. Efficient coding of natural scenes in the lateral geniculate nucleus: experimental test of a computational theory. *J Neurosci* 16: 3351–3362, 1996.
- Daugman JG. Two-dimensional spectral analysis of cortical receptive field profiles. *Vision Res* 20: 847–856, 1980.
- Derrington AM, Allen HA, Delicato LS. Visual mechanisms of motion analysis and motion perception. *Annu Rev Psychol* 55: 181–205, 2004.
- DeValois RL, DeValois KK. *Spatial Vision*. Oxford, UK: Oxford Univ. Press, 1988.
- Drewes J, Barthelemy F, Masson GS. Human ocular following and natural scene statistics. *J Vis* 8: 383, 2008.
- Essock EA, Haun AM, Kim YJ. An anisotropy of orientation-tuned suppression that matches the anisotropy of typical natural scenes. *J Vis* 9: 35.1–15, 2009.
- Felsen G, Dan Y. A natural approach to studying vision. *Nat Neurosci* 8: 1643–1646, 2005.
- Field DJ. Relations between the statistics of natural images and the response properties of cortical cells. *J Opt Soc Am A* 4: 2379–2394, 1987.

- Field DJ. What is the goal of sensory coding? *Neural Comput* 6: 559–601, 1994.
- Field DJ. Wavelets, vision and the statistics of natural scenes. *Philos T Roy Soc A* 357: 2527–2542, 1999.
- Fischer S, Sroubek F, Perrinet LU, Redondo R, Cristóbal G. Self-invertible 2D log-gabor wavelets. *Int J Comput Vis* 75: 231–246, 2007.
- Galerne B, Gousseau Y, Morel JM. Random phase textures: theory and synthesis. *IEEE Trans Image Process* 20: 257–267, 2010.
- Geisler WS, Ringach D. Natural systems analysis. *Vis Neurosci* 26: 1–3, 2009.
- Graham N. Does the brain perform a fourier analysis of the visual scene? *Trends Neurosci* 2: 207–208, 1979.
- Hansen B, Hess R. The role of spatial phase in texture segmentation and contour integration. *J Vis* : 594–615, 2006.
- Heeger DJ, Simoncelli EP, Movshon JA. Computational models of cortical visual processing. *Proc Natl Acad Sci USA* 93: 623–627, 1996.
- Henriksson L, Nurminen L, Hyvärinen A, Vanni S. Spatial frequency tuning in human retinotopic visual areas. *J Vis* 8: 5.1–13, 2008.
- Hess R, Hayes A, Field DJ. Contour integration and cortical processing. *J Physiol (Paris)* 97: 105–119, 2003.
- Jacob V, Le Cam J, Ego-Stengel V, Shulz DE. Emergent properties of tactile scenes selectively activate barrel cortex neurons. *Neuron* 60: 1112–1125, 2008.
- Jasinschi R, Rosenfeld A, Sumit K. Perceptual motion transparency: the role of geometrical information. *J Opt Soc Am A* 9: 1865–1879, 1992.
- Jones JP, Palmer LA. An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. *J Neurophysiol* 58: 1233–1258, 1987.
- Klein DJ, Depireux DA, Simon JZ, Shamma SA. Robust spectrotemporal reverse correlation for the auditory system: optimizing stimulus design. *J Comput Neurosci* 9: 85–111, 2000.
- Lagae A, Lefebvre S, Drettakis G, Dutré P. Procedural noise using sparse gabor convolution. *Technical Report Report CW 545*. Leuven, Belgium: Dept. of Computer Science, K. U. Leuven, 2009.
- Lee TS. Image representation using 2D gabor wavelets. *IEEE T Pattern Anal* 18: 959–971, 1996.
- Lewis JP. Texture synthesis for digital painting. *ACM SIG-GRAPH* 18: 245–252, 1984.
- Lu ZL, Sperling G. Three-systems theory of human visual motion perception: review and update. *J Opt Soc Am A* 18: 2331–2270, 2001.
- Mainen ZF, Sejnowski TJ. Reliability of spike timing in neocortical neurons. *Science* 268: 1503–1506, 1995.
- Marcelja S. Mathematical description of the responses of simple cortical cells. *J Opt Soc Am A* 70: 1297–1300, 1980.
- Masson GS, Castet E. Parallel motion processing for the initiation of short-latency ocular following in humans. *J Neurosci* 22: 5149–5163, 2002.
- Masson GS, Ilg UJ. (editors). *Dynamics of Visual Motion Processing: Neuronal, Behavioral and Computational Approaches* (1st ed.). Berlin, Germany: Springer, 2010.
- Movshon JA, Adelson EH, Gizzi MS, Newsome, (1985) WT. The analysis of moving visual patterns. In: *Pattern Recognition Mechanisms*, edited by Chagas C, Gattass R, Gross C. Rome: Vatican Press, 1985, vol 554, p. 117–151.
- Neri P, Morrone MC, Burr DC. Seeing biological motion. *Nature* 395: 894–896, 1988.
- Newsome WT. Deciding about motion: linking perception to action. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 181: 5–12, 1997.
- Nishimoto S, Gallant JL. A three-dimensional spatiotemporal receptive field model explains responses of area MT neurons to naturalistic movies. *J Neurosci* 31: 14551–14564, 2011.
- Oliva A, Torralba A. Modeling the shape of the scene: a holistic representation of the spatial envelope. *Int J Comput Vision* 42: 145–175, 2001.
- Oppenheim AV, Lim JS. The importance of phase in signals. *Proc IEEE* 69: 529–541, 1981.
- Pei YC, Hsiao SS, Craig JC, Bensmaia SJ. Neural mechanisms of tactile motion integration in somatosensory cortex. *Neuron* 69: 536–547, 2011.
- Pelli DG. The videotoolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision* 10: 437–442, 1997.
- Perrinet L, Fitzpatrick D, Bednar JA. Edge statistics in natural images versus laboratory animal environments: implications for understanding lateral connectivity in V1. In: *Society for Neuroscience Abstracts*, program no. 530.04, 2011.
- Perrinet LU. Role of homeostasis in learning sparse representations. *Neural Comput* 22: 1812–1836, 2010.
- Perrone JA, Thiele A. Speed skills: measuring the visual speed analyzing properties of primate MT neurons. *Nat Neurosci* 4: 526–532, 2001.
- Perrone JA, Thiele A. A model of speed tuning in MT neurons. *Vision Res* 42: 1035–1051, 2002.
- Phillips F, Todd J. Texture discrimination based on global feature alignments. *J Vis* 10: 6, 2010.
- Priebe N, Cassanella C, Lisberger S. The neural representation of speed in macaque area MT/V5. *J Neurosci* 23: 5650–5661, 2003.
- Priebe N, Lisberger S, Movshon JA. Tuning for spatiotemporal frequency and speed in directionally selective neurons of macaque striate cortex. *J Neurosci* 26: 2941–2950, 2006.
- Rieke F, Bodnar DA, Bialek W. Naturalistic stimuli increase the rate and efficiency of information transmission by primary auditory afferents. *Proc Biol Sci* 262: 259–265, 1996.
- Rust NC, Mante V, Simoncelli EP, Movshon JA. How MT cells analyze the motion of visual patterns. *Nat Neurosci* 9: 1421–1431, 2006.
- Rust NC, Movshon JA. In praise of artifice. *Nat Neurosci* 8: 1647–1650, 2005.
- Rust NC, Schwartz O, Movshon JA, Simoncelli EP. Spatiotemporal elements of macaque V1 receptive fields. *Neuron* 46: 945–956, 2005.
- Scarfe P, Johnston A. Motion drag induced by global motion gabor arrays. *J Vis* 10: 14, 2010.
- Schrater PR, Knill DC, Simoncelli EP. Mechanisms of visual motion detection. *Nat Neurosci* 3: 64–68, 2000.
- Schwartz O, Simoncelli EP. Natural signal statistics and sensory gain control. *Nat Neurosci* 4: 819–825, 2001.
- Simoncelli EP, Heeger DJ. A model of neuronal responses in visual area MT. *Vision Res* 38: 743–761, 1998.
- Simoncini C, Perrinet LU, Montagnini A, Mamassian P, Masson G. Measuring speed of moving textures: different pooling of motion information for human ocular following and perception. *Front. Neurosci. Conference Abstract: Neural Coding, Decision-Making & Integration in Time*. First published January 16, 2012; doi:10.3389/conf.fnins.2012.86.00016.
- Singh KD, Smith AT, Greenlee MW. Spatiotemporal frequency and direction sensitivities of human visual areas measured using fMRI. *NeuroImage* 12: 550–564, 2000.
- Smith A, Edgar G. The influence of spatial frequency on perceived temporal frequency and perceived speed. *Vision Res* 30: 1467–1474, 1990.
- Solomon SG, Tailby C, Cheong SK, Camp AJ. Linear and nonlinear contributions to the visual sensitivity of neurons in primate lateral geniculate nucleus. *J Neurophysiol* 104: 1884–1898, 2010.
- Torralba A, Oliva, (2003) A. Statistics of natural image categories. In: *Network: Computation in Neural Systems*, 2003, p. 391–412.
- Touryan J. Analysis of sensory coding with complex stimuli. *Curr Opin Neurobiol* 11: 443–448, 2001.
- Ts'o D, Frostig R, Lieke E, Grinvald A. Functional organization of primate visual cortex revealed by high resolution optical imaging. *Science* 249: 417–420, 1990.
- Tsuchiya N, Braun J. Contrast thresholds for component motion with full and poor attention. *J Vis* 7: 1, 2007.
- Turing A. The chemical basis of morphogenesis. *Philos. Trans. Roy. Soc. London Ser. B*, 1952, vol. 327, p. 37–72.
- Victor J, Conte M. The role of high-order phase correlations in texture processing. *Vision Res* 36: 1615–1631, 1996.
- Vinje WE, Gallant JL. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science* 287: 1273–1276, 2000.
- Wang C, Yao H. Sensitivity of V1 neurons to direction of spectral motion. *Cereb Cortex* 21: 964–973, 2011.
- Watson A, Eckert M. Motion-contrast sensitivity: visibility of motion gradients of various spatial frequencies. *J Opt Soc Am A* 11: 486–505, 1994.
- Watson A, Turano K. The optimal motion stimulus. *Vision Res* 35: 325–336, 1995.
- Watson AB, Barlow HB, Robson JG. What does the eye see best? *Nature* 302 : 419–422, 1983.
- Wu MCK, David SV, Gallant JL. Complete functional characterization of sensory neurons by system identification. *Annu Rev Neurosci* 29: 477–505, 2006.
- Zeki S. Colour coding in the cerebral cortex: the reaction of cells in monkey visual cortex to wavelengths and colours. *Neuroscience* 9: 741–765, 1983.