Bayesian Modeling of Motion Perception Using Dynamical Stochastic Textures

Jonathan Vacher
jonathan.vacher@ens.fr
Département de Mathématique et Applications, École Normale Supérieure, Paris 75005, France; UNIC, Gif-sur-Yvette 91190, France; and CNRS, France

Andrew Isaac Meso
ameso@bournemouth.ac.uk
Institut des Neurosciences de la Timone, Marseille 13005, France, and Faculty of Science and Technology, Bournemouth University, Poole BH12 5BB, U.K.

Laurent U. Perrinet
Laurent.Perrinet@univ-amu.fr
Institut de Neurosciences de la Timone, Marseille 13005, France, and CNRS, France

Gabriel Peyré
gabriel.peyre@ens.fr
Département de Mathématique et Applications, École Normale Supérieure, Paris 75005, France, and CNRS, France

A common practice to account for psychophysical biases in vision is to frame them as consequences of a dynamic process relying on optimal inference with respect to a generative model. The study presented here details the complete formulation of such a generative model intended to probe visual motion perception with a dynamic texture model. It is derived in a set of axiomatic steps constrained by biological plausibility. We extend previous contributions by detailing three equivalent formulations of this texture model. First, the composite dynamic textures are constructed by the random aggregation of warped patterns, which can be viewed as three-dimensional gaussian fields. Second, these textures are cast as solutions to a stochastic partial differential equation (sPDE). This essential step enables real-time, on-the-fly texture synthesis using time-discretized autoregressive processes. It also allows for the derivation of a local motion-energy model, which corresponds to the log likelihood of the probability density. The log likelihoods are essential for the construction of a Bayesian inference framework. We use the dynamic texture model to psychophysically probe speed perception in humans using zoom-like changes in the spatial frequency content of the stimulus. The human data replicate previous findings showing perceived speed to be

Neural Computation **30, 3355–3392** (2018)
doi:10.1162/neco_a_01142

positively biased by spatial frequency increments. A Bayesian observer who combines a gaussian likelihood centered at the true speed and a spatial frequency dependent width with a "slow-speed prior" successfully accounts for the perceptual bias. More precisely, the bias arises from a decrease in the observer's likelihood width estimated from the experiments as the spatial frequency increases. Such a trend is compatible with the trend of the dynamic texture likelihood width.

1 Introduction

1.1 Modeling Visual Motion Perception. A normative explanation for the function of perception is to infer relevant unknown real-world parameters from the sensory input with respect to a generative model (Gregory, 1980). Equipped with some prior knowledge about both the nature of neural representations and the structure of the world, the modeling approach that emerges corresponds to the Bayesian brain hypothesis (Knill & Pouget, 2004; Doya, 2007; Colombo & Seriès, 2012; Kersten, Mamassian, & Yuille, 2004). This assumes that when given some sensory information S, the brain uses neural computations that ultimately conform with Bayes' theorem:

$$\mathbb{P}_{M|S}(m|s) = \frac{\mathbb{P}_{S|M}(s|m)\mathbb{P}_M(m)}{\mathbb{P}_S(s)}. \tag{1.1}$$

This computation yields an estimate of the parameters m where the probability distribution function $\mathbb{P}_{S|M}$ is given by the generative model and $\mathbb{P}_M$ represents prior knowledge. This hypothesis has been well illustrated with the case of motion perception (Weiss, Simoncelli, & Adelson, 2002). This framework uses a gaussian parameterization of the generative model and a unimodal (gaussian) prior in order to estimate perceived speed v when observing a visual input I.

However, gaussian likelihoods and priors do not always fit with psychophysical results (Wei & Stocker, 2012; Hassan & Hammett, 2015). Thus, a major challenge is to refine the construction of generative models so that they are consistent with the widest variety of empirical results.

In fact, the estimation problem inherent in perception is successfully solved in part through the definition of an adequate generative model. Probably the simplest generative model to describe visual motion is the luminance conservation equation (Adelson & Bergen, 1985). It states that luminance $I(x, t)$ for $(x, t) \in \mathbb{R}^2 \times \mathbb{R}$ is approximately conserved along trajectories defined as integral lines of a vector field $v(x, t) \in \mathbb{R}^2 \times \mathbb{R}$. The corresponding generative model defines random fields as solutions to the stochastic partial differential equation (sPDE),

$$\langle v,\ \nabla I\rangle + \frac{\partial I}{\partial t} = W, \tag{1.2}$$

where $\langle\cdot,\ \cdot\rangle$ denotes the Euclidean scalar product in $\mathbb{R}^2$ and ∇I is the spatial gradient of I. To match the distribution of spatial scale statistics of natural scenes (the 1/f amplitude fall-off of spatial frequencies) or some alternative category of textures, the driving term W is usually defined as a stationary colored gaussian noise corresponding to the average localized spatiotemporal correlation (which we refer to as the spatiotemporal coupling), and is parameterized by a covariance matrix Σ, while the field is usually a constant vector $v(x, t) = v_0$ accounting for a full-field translation with constant speed.

Ultimately, the application of this generative model is useful for probing the visual system with a probabilistic approach—for instance, for one seeking to understand how observers might detect motion in a scene. Indeed, as Nestares, Fleet, and Heeger (2000) and Weiss et al. (2002) showed, the negative log likelihood of the probability distribution of the solutions I to the luminance conservation equation 1.2 (on domain $\Omega \times [0, T]$ and for constant speed $v(x, t) = v_0$) is proportional to the value of the motion-energy model (Adelson & Bergen, 1985) given by

$$\int_\Omega \int_0^T |\langle v_0,\ \nabla(K \star I)(x, t)\rangle + \frac{\partial(K \star I)}{\partial t}(x, t)|^2 dt\, dx, \tag{1.3}$$

where K is the whitening filter corresponding to the inverse square root of Σ and $\star$ is the convolution operator. Using some prior knowledge about the expected distribution of motions—for instance, a preference for slow speeds—a Bayesian formalization can be applied to this inference problem (Weiss & Fleet, 2001; Weiss et al., 2002).

1.2 Previous Work in Context

1.2.1 Dynamic texture synthesis. The model defined in equation 1.2 is quite simplistic compared to the complexity of natural scenes. It is therefore useful here to discuss generative models associated with texture synthesis methods previously proposed in the computer vision and computer graphics community. Indeed, the literature on the subject of static textures synthesis is abundant (see, e.g., Wei, Lefebvre, Kwatara, & Turk, 2009). Of particular interest for us is the work by Galerne, Gousseau, and Morel (2011a) and Galerne (2011), which proposes a stationary gaussian model restricted to static textures. This provides an equivalent generative model based on Poisson shot noise. Realistic dynamic texture models have received less attention, and the most prominent method is the nonparametric gaussian autoregressive (AR) framework developed by

Doretto, Chiuso, Wu, and Soatto (2003), which has been thoroughly explored (Xia, Ferradans, Peyré, & Aujol, 2014; Yuan, Wen, Liu, & Shum, 2004; Costantini, Sbaiz, & Süsstrunk, 2008; Filip, Haindl, & Chetverikov, 2006; Hyndman, Jepson, & Fleet, 2007; Abraham, Camps, & Sznaier, 2005). These works generally consist in finding an appropriate low-dimensional feature space in which an AR process models the dynamics. Many of these approaches focus on the feature space where the decomposition is efficiently performed using singular value decomposition (SVD) or its higher-order version (HOSVD) (Doretto et al., 2003; Costantini et al., 2008). In Abraham et al. (2005), the feature space is the Fourier frequency domain, and the AR recursion is carried independently over each frequency, which defines the space-time stationary processes. A similar approach is used in Xia et al. (2014) to compute the average of several dynamic texture models. Properties of these AR models have been studied by Hyndman et al. (2007), who find that higher-order AR processes are able to capture perceptible temporal features. A different approach aims at learning the manifold structure of a given dynamic texture (Liu, Lin, Ahuja, & Yang, 2006), while yet another deals with motion statistics (Rahman & Murshed, 2008). What all these works have in common is the aim to reproduce the natural spatiotemporal behavior of dynamic textures with rigorous mathematical tools. Similarly, our concern is to design a dynamic texture model that is precisely parameterized for experimental purposes in visual neuroscience and psychophysics.

1.2.2 Stochastic Differential Equations. Stochastic ordinary differential equations (sODE) and their higher-dimensional counterparts, stochastic partial differential equations (sPDE), can be viewed as continuous-time versions of these one-dimensional or higher-dimensional AR models. Conversely, AR processes can therefore also be used to compute numerical solutions to these sPDE using finite difference approximations of time derivatives. Informally, these equations can be understood as partial differential equations perturbed by a random noise. The theoretical and numerical study of these sDEs is of fundamental interest in fields as diverse as physics and chemistry (Van & Nicolaas, 1992), finance (El Karoui, Peng, & Quenez, 1997), or neuroscience (Fox, 1997). They allow for the dynamic study of complex, irregular, and random phenomena such as particle interactions, stock or saving prices, or ensembles of neurons. In psychophysics, sODEs have been used to model decision-making tasks in which the stochastic variable represents the accumulation of knowledge until the decision is made, thus providing detailed information about predicted response times (Smith, 2000). In imaging sciences, sPDE with sparse nongaussian driving noise have been proposed as models of natural signals and images (Unser & Tafti, 2014). As described above, the simple motion energy model 1.3 can similarly be demonstrated to rely on the sPDE equation, 1.2, of a stochastic model of visual sensory input. This has not previously been presented in a formal way in the literature. One key goal of this letter is to comprehensively formulate a parametric family of gaussian sPDEs that describes the modeling of moving images (and the corresponding synthesis of visual stimulation) and thus allows for a finely grained systematic exploration of psychophysical behavior.

1.2.3 Inverse Bayesian Inference. Importantly, these dynamic stochastic models are closely related to the likelihood and prior models, which serve to infer motion estimates from the dynamic visual stimulation. In order to account for perceptual bias, a now well-accepted methodology in the field of psychophysics is to assume that observers are "ideal observers" and therefore make decisions using optimal statistical inference, typically a maximum a posteriori (MAP) estimator, using Bayes' formula to combine this likelihood with some internal prior (see equation 1.1). Several experimental studies have used this hypothesis as a justification for the observed perceptual biases by proposing some adjusted likelihood and prior models (Doya, 2007; Colombo & Seriès, 2012), and more recent work pushes these ideas even further. Observing some perceptual bias, is it possible to invert this forward Bayesian decision-making process, and infer the (unknown) internal prior that best fits a set of observed experimental choices made by observers? Following Stocker and Simoncelli (2006), we coined this promising methodology *inverse Bayesian inference*. This is, of course, an ill-posed and highly nonlinear inverse problem, making it necessary to add constraints on both the prior and the likelihood to make it tractable. For instance Sotiropoulos, Seitz, and Seriès (2014), Stocker and Simoncelli (2006), and Jogan and Stocker (2015) impose smoothness constraints in order to be able to locally fit the slope of the prior. Here, we propose to use visual stimulations generated by the (forward) generative model to test these inverse Bayesian models. To allow for a simple yet mathematically rigorous analysis of this approach within the context of speed discrimination, in this study, we use a restricted parametric set of descriptors for the likelihood and priors. This provides a self-consistent approach to test the visual system, from stimulation to behavior analysis.

1.3 Contributions. In this letter, we lay the foundations that we hope will enable a better understanding of human motion perception by improving generative models for dynamic texture synthesis. From that perspective, we motivate the generation of visual stimulation within a stationary gaussian dynamic texture model.

We develop our current model by extending, mathematically detailing, and testing in psychophysical experiments previously introduced dynamic noise textures (Sanz-Leon, Vanzetta, Masson, & Perrinet, 2012; Simoncini, Perrinet, Montagnini, Mamassian, & Masson, 2012; Vacher, Meso, Perrinet, & Peyré, 2015; Gekas, Masson, & Mamassian, 2017) coined motion clouds (MC). Our first contribution is a complete axiomatic

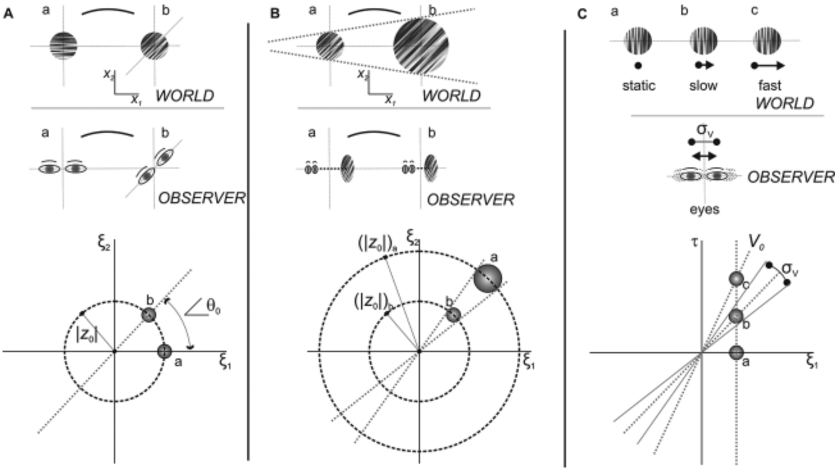


Figure 1: Parameterization of the class of motion clouds (MC) stimuli. The illustration relates the parametric changes in MC with real-world (top row) and observer (second row) movements. (A) Orientation changes resulting in scene rotation are parameterized through θ, as shown in the bottom row where (a) horizontal and (b) obliquely oriented MCs are compared. (B) Zoom movements, from either scene looming or observer movements in depth, are characterized by scale changes reflected by a frequency term z shown for (a) a more distant viewpoint compared to (b) a closer one. (C) Translational movements in the scene characterized by V using the same formulation for (a) static, (b) slow, and (c) fast-moving MC, with the variability in these speeds quantified by σ_V. The variables ξ and τ in the third row are the spatial and temporal frequency scale parameters. The development of this formulation is detailed in the text.

derivation of the model, seen as a shot noise aggregation of dynamically warped "textons." Within our generative model, the parameters correspond to average spatial and temporal transformations (zoom, orientation, and translation speed) and associated standard deviations of random fluctuations, as illustrated in Figure 1, with respect to external (objects) and internal (observer) movements. The second main contribution is the explicit demonstration of the equivalence between this model and a class of linear sPDEs. This shows that our model is a generalization of the well-known luminance conservation (see equation 1.2). This sPDE formulation has two chief advantages: it allows for a real-time synthesis using an AR recurrence (in the form of a GPU implementation) and allows one to recast the log likelihood of the model as a generalization of the classical motion energy model, which is crucial to allow for Bayesian modeling of perceptual biases. Our last contribution follows from the Bayesian approach and is an illustrative application of this model to the psychophysical study of motion perception in humans. This example of the model development constrains the likelihood, which enables a simple fitting procedure to be performed using both an empirical and a larger Monte Carlo–derived synthetic data set to determine the prior driving the perceptual biases. The code associated with this work is available at https://github.com/JonathanVacher/projects/tree/master/bayesian_observer.

1.4 Notations. In the following, we denote $(x, t) \in \mathbb{R}^2 \times \mathbb{R}$ the space-time variable and $(\xi, \tau) \in \mathbb{R}^2 \times \mathbb{R}$ the corresponding frequency variables. If $f(x, t)$ is a function defined on $\mathbb{R}^3$, then its Fourier transform is defined as

$$\hat{f}(\xi, \tau) \overset{\text{def.}}{=} \int_{\mathbb{R}^2} \int_{\mathbb{R}} f(x, t) e^{-\mathrm{i}(\langle x, \xi \rangle + \tau t)} \mathrm{d}t \mathrm{d}x.$$

For $\xi \in \mathbb{R}^2$, we denote $\xi = \|\xi\|(\cos(\angle\xi), \sin(\angle\xi)) \in \mathbb{R}^2$ its polar coordinates. For a function g defined on $\mathbb{R}^2$, we denote $\bar{g}(x) = g(-x)$. We denote a random variable with a capital letter such as A and a as a realization of A. We note as $\mathbb{P}_A(a)$ the corresponding probability distribution of A.

2 Axiomatic Construction of the Dynamic Textures

Dynamic textures that are efficient to probe visual perception should be generated from low-dimensional yet naturalistic parametric stochastic models. They should embed meaningful physical parameters (such as the effect of head rotations or whole-field scene movements; see Figure 1) into the local or global dependencies of the random field (e.g., the covariance). In the luminance conservation model, equation 1.2, the generative model is parameterized by a spatiotemporal coupling encoded in the covariance Σ of the driving noise and the motion flow v_0.

This localized space-time coupling (e.g., the covariance, if one restricts one's attention to gaussian fields) is essential, as it quantifies the extent of the spatial integration area, as well as the integration dynamics. This is an important issue in neuroscience when considering the implementation of spatiotemporal integration mechanisms from very small to very large scales, that is, going from local to global visual features (Rousselet, Thorpe, & Fabre-Thorpe, 2004; Born & Bradley, 2005; DiCarlo, Zoccolan, & Rust, 2012). In particular, this is crucial to understand the modular sensitivity within the different lower visual areas. In primates for instance, the primary visual cortex (V1) generally encodes small features in a given range of spatiotemporal scales. In contrast, ascending the processing hierarchy, the middle temporal (V5/MT) area exhibits selectivity for larger visual features. For instance, by varying the spatial frequency bandwidth of such dynamic textures, distinct mechanisms for perception and action have been identified in humans (Simoncini et al., 2012). Our goal here is to develop a principled axiomatic definition of these dynamic textures.

2.1 From Shot Noise to Motion Clouds. We propose a derivation of a general parametric model of dynamic textures. This model is defined by aggregation, through summation, of a basic spatial "texton" template $g(x)$. The summation reflects a transparency hypothesis, which has been adopted, for instance, by Galerne et al. (2011b). While one could argue that this hypothesis is overly simplistic and does not model occlusions or edges, it leads to a tractable framework of stationary gaussian textures, which has proved useful to model static microtextures (Galerne et al., 2001b) and dynamic natural phenomena (Xia, Ferradans, Peyré, & Aujol, 2014). The simplicity of this framework allows for a fine tuning of frequency-based (Fourier) parameterization, which is desirable for the interpretation of psychophysical experiments with respect to underlying spatiotemporal neural sensitivity.

We define a random field as

$$I_\lambda(x,t) \overset{\text{def.}}{=} \frac{1}{\sqrt{\lambda}} \sum_{p\in\mathbb{N}} g(\varphi_{A_p}(x - X_p - V_p t)), \tag{2.1}$$

where $\varphi_a : \mathbb{R}^2 \to \mathbb{R}^2$ is a planar deformation parameterized by a finite-dimensional vector a. The parameters $(X_p, V_p, A_p)_{p\in\mathbb{N}}$ are independent and identically distributed random vectors. They account for the variability in the position of objects or observers (φ_{A_p}) and their speed (V_p), thus mimicking natural motions in an ambient scene. The set of translations $(X_p)_{p\in\mathbb{N}}$ is a 2D Poisson point process of intensity $\lambda > 0$. This means that defining for any measurable A, $C(A) = \sharp\left\{p\ ;\ X_p \in A\right\}$, $C(A)$ has a Poisson distribution with mean $\lambda|A|$ (where $|A|$ is the measure of A) and $C(A)$ is independent of $C(B)$ if $A \cap B = \emptyset$.

Intuitively, this model equation 2.1, corresponds to a dense mixing of stereotyped static textons as in the work of Galerne et al. (2011b). In addition to the extension to the temporal domain, the originality of our approach is twofold. First, the components of this mixing are derived from the texton by visual transformations φ_{A_p}, which may correspond to arbitrary transformations such as zooms or rotations (in which case, A_p is a vector containing the scale factor and the rotation angle). (See the illustration in Figure 1.) Second, we explicitly model the motion (position X_p and speed V_p) of each individual texton.

In the following, we denote $\mathbb{P}_A$ the common distribution of the independent and identically distributed (i.i.d) $(A_p)_p$, and we denote $\mathbb{P}_V$ the distribution in $\mathbb{R}^2$ of the speed vectors $(V_p)_p$. Section 2.3 instantiates this model and proposes canonical choices for these variabilities.

The following result shows that the model, equation 2.1, converges for high point density $\lambda \to +\infty$ to a stationary gaussian field and gives the parameterization of the covariance. Its proof follows from a specialization of theorem 3.1 in Galerne (2011) to our setting.

Proposition 1. *I_λ is stationary with bounded second-order moments. Its covariance is $\Sigma(x, t, x', t') = \gamma(x - x', t - t')$ where γ satisfies*

$$\forall\,(x, t) \in \mathbb{R}^3, \quad \gamma(x, t) = \iiint_{\mathbb{R}^2} c_g(\varphi_a(x - \nu t))\mathbb{P}_V(\nu)\mathbb{P}_A(a)d\nu da, \tag{2.2}$$

where $c_g = g \star \bar{g}$ is the autocorrelation of g. When $\lambda \to +\infty$, it converges (in the sense of finite-dimensional distributions) toward a stationary gaussian field I of zero mean and covariance Σ.

This proposition enables us to give a precise definition of an MC:

Definition 1. *A motion cloud (MC) is a stationary gaussian field whose covariance is given by equation 2.2.*

Note that following Galerne et al. (2011a), the convergence result of proposition 1 could be used in practice to simulate a motion cloud I using a high but finite value of λ in order to generate a realization of I_λ. We do not use this approach and instead rely on the sPDE characterization proved in section 3, which is well tailored for an accurate and computationally efficient dynamic synthesis.

2.2 Toward Motion Clouds for Experimental Purposes. The previous section provides a theoretical definition of MC (see definition 1) that is characterized by c_g, φ_a, $\mathbb{P}_A$, and $\mathbb{P}_V$. In order to have better control of the covariance γ, one needs to resort to a low-dimensional representation of these parameters. We further study this model in the specific case where the warps φ_a are rotations and scalings (see Figure 1). They account for the characteristic orientations and sizes (or spatial scales) of a scene, in relation to the observer. We thus set

$$\forall\, a = (\theta, z) \in [-\pi, \pi) \times \mathbb{R}_+^*, \quad \varphi_a(x) \overset{\text{def.}}{=} zR_{-\theta}(x), \tag{2.3}$$

where R_θ is the planar rotation of angle θ. We now give some physical and biological motivation to account for our particular choices for the distributions of the parameters. We assume that the distributions $\mathbb{P}_Z$ and $\mathbb{P}_\Theta$ of spatial scales z and orientations θ, respectively (see Figure 1), are independent and have densities, thus considering

$$\forall\, a = (\theta, z) \in [-\pi, \pi) \times \mathbb{R}_+^*, \quad \mathbb{P}_A(a) = \mathbb{P}_Z(z)\,\mathbb{P}_\Theta(\theta). \tag{2.4}$$

The speed vector ν is assumed to be randomly fluctuating around a central speed $v_0 \in \mathbb{R}^2$, so that

$$\forall\, \nu \in \mathbb{R}^2, \quad \mathbb{P}_V(\nu) = \mathbb{P}_{\|V - v_0\|}(\|\nu - v_0\|). \tag{2.5}$$

In order to obtain "optimal" responses to the stimulation (as advocated by Young & Lesperance, 2001) and based on the structure of a standard receptive field of V1, it makes sense to define the texton so that it resembles an oriented Gabor (Fischer, Sroubek, Perrinet, Redondo, & Cristóbal, 2007). Such an elementary luminance feature acts as the generic atom

$$g_\sigma(x) = \frac{1}{2\pi} \cos\left(\langle x,\ \xi_0 \rangle\right) e^{-\frac{\sigma^2}{2}\|x\|^2}, \tag{2.6}$$

where σ is the inverse of the standard deviation and $\xi_0 \in \mathbb{R}^2$ is the spatial frequency. Since the orientation and scale of the texton are handled by the (θ, z) parameters, we can impose the normalization $\xi_0 = (1, 0)$ without loss of generality. In the special case where $\sigma \to 0$, g_σ is a grating of frequency ξ_0 and the image I is a dense mixture of drifting gratings, whose power spectrum has a closed-form expression detailed in proposition 2. It is fully parameterized by the distributions $(\mathbb{P}_Z, \mathbb{P}_\Theta, \mathbb{P}_V)$ and the central frequency and speed (ξ_0, v_0). Note that it is possible to consider any arbitrary textons g, which would give rise to more complicated parameterizations for the power spectrum $\hat{g}$, but here we decided to stick to the simple asymptotic case of gratings.

Proposition 2. *Consider the texton g_σ, when $\sigma \to 0$, the gaussian field $I_\sigma(x, t)$ defined in proposition 1 converges toward a stationary gaussian field of covariance having the power spectrum*

$$\begin{aligned} &\forall\, (\xi, \tau) \in \mathbb{R}^2 \times \mathbb{R},\ \hat{\gamma}(\xi, \tau) \\ &\quad = \frac{\mathbb{P}_Z\left(\|\xi\|\right)}{\|\xi\|^2} \mathbb{P}_\Theta\left(\angle \xi\right) \mathcal{L}(\mathbb{P}_{\|V - v_0\|}) \left(-\frac{\tau + \langle v_0,\ \xi \rangle}{\|\xi\|} \right), \end{aligned} \tag{2.7}$$

where the linear transform $\mathcal{L}$ is such that

$$\forall\, u \in \mathbb{R}, \quad \mathcal{L}(f)(u) \overset{def.}{=} \int_{-\pi}^{\pi} f(-u/cos(\varphi))d\varphi. \tag{2.8}$$

Proof. See appendix A for the proof.

Remark 1. Note that the envelope of $\hat{\gamma}$ as defined in equation 2.7 is constrained to lie within a cone in the spatiotemporal domain with the apex at zero (see the division by $\|\xi\|$ in the argument of $\mathcal{L}(\mathbb{P}_{\|V - v_0\|})$). This is an important and novel contribution when compared to a classical Gabor. Basing the generation of the textures on distributions of translations, rotations, and zooms, we provide a principled approach to show that speed bandwidth gets scaled with spatial frequency to provide a scale-invariant model of moving texture transformations.

2.3 Biologically Inspired Parameter Distributions. We now give meaningful specialization for the probability distributions $\mathbb{P}_Z$, $\mathbb{P}_\Theta$, and $\mathbb{P}_{\|V-v_0\|}$, which are inspired by some known scaling properties of the visual transformations relevant to dynamic scene perception.

2.3.1 Parameterization of $\mathbb{P}_Z$. First, the observer's small, centered linear movements along the axis of view (orthogonal to the plane of the scene) generate centered planar zooms of the image. From the linear modeling of the observer's displacement and the subsequent multiplicative nature of zoom, scaling should follow a Weber-Fechner law. This law states that subjective perceptual sensitivity when quantified is proportional to the logarithm of stimulus intensity. Thus, we choose the scaling z drawn from a log-normal distribution $\mathbb{P}_Z$, defined in equation 2.9. The parameter $\tilde{\sigma}_Z$ quantifies the variation in the amplitude of zooms of individual textons relative to the characteristic scale $\tilde{z}_0$. We thus define

$$\mathbb{P}_Z(z) \propto \frac{\tilde{z}_0}{z} \exp\left(- \frac{\ln\left(\frac{z}{\tilde{z}_0}\right)^2}{2\ln\left(1+\tilde{\sigma}_Z^2\right)} \right), \tag{2.9}$$

where $\propto$ means that we did not include the normalizing constant. In practice, we may prefer to parameterize this distribution by its mode and octave bandwidth (z_0, B_z) instead of $(\tilde{z}_0, \tilde{\sigma}_Z)$. (See appendix C, where we discuss two different parameterizations.)

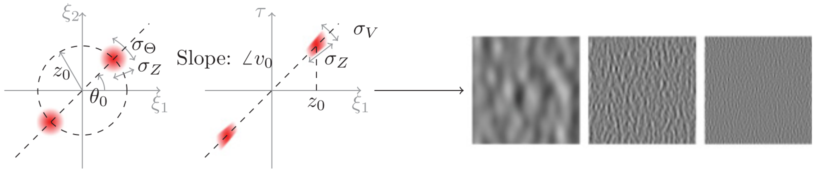
2.3.2 Parameterization of $\mathbb{P}_\Theta$. In our model, the texture is perturbed by variations in the global angle θ of the scene; for instance, the head of the observer may roll slightly around its normal position. The von Mises distribution, as a good approximation of the warped gaussian distribution around the unit circle, is an adapted choice for the distribution of θ with mean θ_0 and bandwidth σ_Θ,

$$\mathbb{P}_\Theta(\theta) \propto e^{\frac{\cos(2(\theta-\theta_0))}{4\sigma_\Theta^2}}. \tag{2.10}$$

2.3.3 Parameterization of $\mathbb{P}_{\|V-v_0\|}$. We may similarly consider that the position of the observer is variable in time. On the first-order approximation, movements perpendicular to the axis of view dominate, generating random perturbations to the global translation v_0 of the image at speed $v - v_0 \in \mathbb{R}^2$. These perturbations are, for instance, described by a gaussian random walk. Take, for instance, tremors, which are small, constant, and jittering movements of the eye ($\leqslant 1$ degree). This justifies the choice of a radial distribution (see equation 2.5) for $\mathbb{P}_V$. This radial distribution $\mathbb{P}_{\|V-v_0\|}$ is thus selected as

Table 1: Full Set of Six Parameters Characterizing the Motion Cloud Stimulus Model.

Parameter Name	Translation Speed	Orientation Angle	Spatial Frequency Modulus
(mean, dispersion)	(v_0, σ_V)	$(\theta_0, \sigma_\Theta)$	(z_0, B_Z)



Two different projections of $\|\xi\|^2\hat{\gamma}(\xi,\tau)$ in Fourier space.

MC of three different spatial frequencies z_0, $2z_0$ and $4z_0$.

Figure 2: Graphical representation of the covariance $\hat{\gamma}$ shown as a projection on the spatial frequency plane (left) and the spatiotemporal frequency plane (middle). Note the cone-like shape of the envelopes in both cases. The three luminance stimulus images on the right are an example of synthesized frames for three different spatial frequencies, respectively, from left to right, a low, a medium, and a high frequency.

a bell-shaped function of width σ_V, and we choose here a gaussian function for its generality:

$$\mathbb{P}_{\|V-v_0\|}(r) \propto e^{-\frac{r^2}{2\sigma_V{}^2}}. \tag{2.11}$$

Note that as detailed in section 3.2, a slightly different bell function (with a more complicated expression) should be used to obtain an exact equivalence with the sPDE discretization.

2.3.4 Bringing Everything Together. Plugging expressions 2.9 to 2.11 into the definition of the power spectrum of the defintion of MCs, equation 2.7, one obtains a parameterization that shares similarities with the one originally introduced in Simoncini et al. (2012).

Table 1 recaps the parameters of the biologically inpired MC models. It is composed of the central parameters v_0 for the speed, θ_0 for orientation, and z_0 for the central spatial frequency modulus, as well as corresponding dispersion parameters $(\sigma_V, \sigma_\Theta, B_Z)$ that account for the typical deviation around these central values. Figure 2 graphically shows the influence of these parameters on the shape of the MC power spectrum $\hat{\gamma}$.

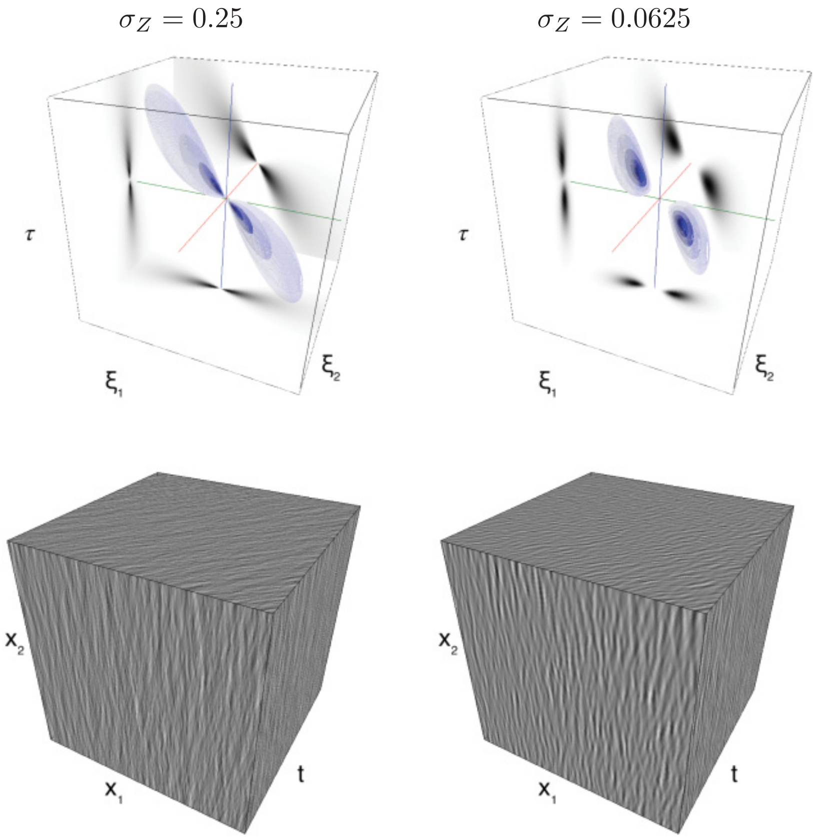


Figure 3: Comparison of a broadband (left) versus narrowband (right) stimulus. Two instances (left and right columns) of two motion clouds having the same parameters, except the frequency bandwidths σ_Z, which were different. The top column displays iso-surfaces of $\hat{\gamma}$, in the form of enclosing volumes at different energy values with respect to the peak amplitude of the Fourier spectrum. The bottom column shows an isometric view of the faces of a movie cube, which is a realization of the random field I. The first frame of the movie lies on the $(x_1, x_2, t = 0)$ spatial plane. The motion cloud with the broadest bandwidth is often thought to best represent stereotyped natural stimuli since it similarly contains a broad range of frequency components.

We show in Figure 3 two examples of such stimuli for two different spatial frequency bandwidths. This is particularly relevant as it is possible to dissociate the respective roles of broader or narrower spatial frequency bandwidths in action and perception (Simoncini et al., 2012). Using this formulation to extend the study of visual perception to other dimensions like orientation or speed bandwidths should provide a means to systematically titrate their respective role in motion integration and obtain a quantitative assessment of their respective contributions in experimental data.

3 sPDE Formulation and Synthesis Algorithm

In this section, we show that the MC model (see definition 1) can equally be described as the stationary solution of a stochastic partial differential equation (sPDE). This sPDE formulation is important since we aim to deal with dynamic stimulation, which should be described by a causal equation that is local in time. This is crucial for numerical simulations, since this allows us to perform real-time synthesis of stimuli using an autoregressive time discretization. This is a significant departure from previous Fourier-based implementation of dynamic stimulations (Sanz-Leon et al., 2012; Simoncini et al., 2012). Moreover, this is also important to simplify the application of MC inside a Bayesian model of psychophysical experiments (see section 4). In particular, the derivation of an equivalent sPDE model exploits a spectral formulation of MCs as gaussian random fields. The full proof along with the synthesis algorithm follows.

To be mathematically correct, all the sPDEs in this letter are written in the sense of generalized stochastic processes (GSP), which are to stochastic processes what generalized functions are to functions. This allows for the consideration of linear transformations of stochastic processes, like differentiation or Fourier transforms as for generalized functions. We refer to Unser, Tafti, and Sun (2014) for a recent use of GSP and to Gel'fand, Vilenkin, and Feinstein (1964) for the foundation of the theory. The connection between GSP and stochastic processes has been described in previous work (Meidan, 1980).

3.1 Dynamic Textures as Solutions of sPDE

3.1.1 Using a sPDE without Global Translation, $v_0 = 0$. We first give the definition of an sPDE cloud I making use of another cloud I_0 without translation speed. This allows us to restrict our attention to the case $v_0 = 0$ in order to define a simple sPDE and then to explicitly extend that result to the general case.

Definition 2. *For a given spatial covariance Σ_W, 2D spatial filters (α, β), and a translation speed $v_0 \in \mathbb{R}^2$, an sPDE cloud is defined as*

$$I(x,t) \overset{def.}{=} I_0(x - v_0 t, t). \tag{3.1}$$

where I_0 is a stationary gaussian field satisfying for all (x, t),

$$\mathcal{D}(I_0) = \frac{\partial W}{\partial t} \quad \textit{where} \quad \mathcal{D}(I_0) \overset{def.}{=} \frac{\partial^2 I_0}{\partial t^2} + \alpha \star \frac{\partial I_0}{\partial t} + \beta \star I_0, \tag{3.2}$$

where the driving noise $\frac{\partial W}{\partial t}$ is white in time (i.e., corresponds to the temporal derivative of a Brownian motion in time) and has a 2D stationary covariance σ_W in space and $\star$ is the spatial convolution operator.

The random field I_0 solving equation 3.2 thus corresponds to an sPDE cloud with no translation speed, $v_0 = 0$. The filters (α, β) parameterizing this sPDE cloud aim at enforcing an additional correlation of the model in time. Section 3.2 explains how to choose $(\alpha, \beta, \sigma_W)$ so that these sPDE clouds, which are stationary solutions of equation 3.2, have the power spectrum given in equation 2.7 (in the case that $v_0 = 0$), that is, are motion clouds. Defining a causal equation that is local in time is crucial for numerical simulation (as explained in section 3.3) but also to simplify the application of MC inside a Bayesian model of psychophysical experiments (see section 4.3.2).

The sPDE equation 3.2 corresponds to a set of independent stochastic ODEs over the spatial Fourier domain, which reads, for each frequency ξ,

$$\forall t \in \mathbb{R}, \qquad \frac{\partial^2 \hat{I}_0(\xi, t)}{\partial t^2} + \hat{\alpha}(\xi)\frac{\partial \hat{I}_0(\xi, t)}{\partial t} + \hat{\beta}(\xi)\hat{I}_0(\xi, t) = \hat{\sigma}_W(\xi)\hat{w}(\xi, t), \tag{3.3}$$

where $\hat{I}_0(\xi, t)$ denotes the Fourier transform with respect to the spatial variable x only and $\hat{\sigma}_W(\xi)^2$ is the spatial power spectrum of $\frac{\partial W}{\partial t}$, which means that

$$\Sigma_W(x, y) = c(x - y) \quad \text{where} \quad \hat{c}(\xi) = \hat{\sigma}_W^2(\xi). \tag{3.4}$$

Finally, $\hat{w}(\xi, t) \sim \mathcal{CN}(0, 1)$ where $\mathcal{CN}$ is the complex-normal distribution.

While equation 3.3 should hold for all time $t \in \mathbb{R}$, the construction of stationary solutions (hence, sPDE clouds) of this equation is obtained by solving the sODE, equation 3.3, forward for time $t > t_0$ with arbitrary boundary conditions at time $t = t_0$, and letting $t_0 \to -\infty$. This is consistent with the numerical scheme detailed in section 3.3.

The theoretical study of equation 3.2 is beyond the scope of this letter; however, one can show the existence and uniqueness of stationary solutions for this class of sPDE under stability conditions on the filters (α, β) (see, e.g., Unser & Tafti, 2014, and Brockwell & Lindner, 2009, and appendix theorem 1). These conditions are automatically satisfied for the particular case of section 3.2.

3.1.2 sPDE with Global Translation. The easiest way to define and synthesize an sPDE cloud I with nonzero translation speed v_0 is to first define

I_0 solving equation 3.3 and then translating it with constant speed using equation 3.1. An alternative way is to derive the sPDE satisfied by I, as detailed in the following proposition. This is useful to define motion energy in section 4.3.2.

Proposition 3. *The MCs noted I with $(\alpha, \beta, \Sigma_W)$ the speed parameters, and v_0 the translation speed are the stationary solutions of the sPDE,*

$$\mathcal{D}(I) + \langle \mathcal{G}(I), v_0 \rangle + \langle \mathcal{H}(I) v_0, v_0 \rangle = \frac{\partial W}{\partial t}, \tag{3.5}$$

where $\mathcal{D}$ is defined in equation 3.2, $\nabla_x^2 I$ is the Hessian of I (second-order spatial derivative), and

$$\mathcal{G}(I) \overset{def.}{=} \alpha \star \nabla_x I + 2\partial_t \nabla_x I \quad \textit{and} \quad \mathcal{H}(I) \overset{def.}{=} \nabla_x^2 I. \tag{3.6}$$

Proof. See appendix A for the proof.

3.2 Equivalence between the Spectral and sPDE Formulations. Since both MCs and sPDE clouds are obtained by a uniform translation with speed v_0 of a motionless cloud, we can restrict our analysis to the case $v_0 = 0$ without loss of generality.

In order to relate MCs to sPDE clouds, equation 3.3 makes explicit that the functions $(\hat{\alpha}(\xi), \hat{\beta}(\xi))$ should be chosen in order for the temporal covariance of the resulting process to be equal to (or at least to approximate well) the temporal covariance appearing in equation 2.7. This covariance should be localized around 0 and be nonoscillating. It thus makes sense to constrain $(\hat{\alpha}(\xi), \hat{\beta}(\xi))$ so that the corresponding ODE, equation 3.3, be critically damped, which corresponds to imposing the following relationship,

$$\forall \xi, \quad \hat{\alpha}(\xi) = \frac{2}{\hat{\nu}(\xi)} \quad \text{and} \quad \hat{\beta}(\xi) = \frac{1}{\hat{\nu}^2(\xi)},$$

for some relaxation step size $\hat{\nu}(\xi)$. The model is thus solely parameterized by the noise variance $\hat{\sigma}_W(\xi)$ and the characteristic time $\hat{\nu}(\xi)$.

The following proposition shows that the sPDE cloud model, equation 3.2, and the motion cloud model, equation 2.7, are identical for an appropriate choice of function $\mathbb{P}_{\|V-v_0\|}$.

Proposition 4. *When considering*

$$\forall r > 0, \quad \mathbb{P}_{\|V-v_0\|}(r) = \mathcal{L}^{-1}(h)(r/\sigma_V) \quad \textit{where} \quad h(u) = (1+u^2)^{-2} \tag{3.7}$$

where $\mathcal{L}$ is defined in equation 2.7, equation 3.2 admits a solution I, which is a stationary gaussian field with power spectrum defined in equation 2.7, when setting

$$\hat{\sigma}_W^2(\xi) = \frac{4}{\hat{\nu}(\xi)^3\|\xi\|^2}\mathbb{P}_Z(\|\xi\|)\mathbb{P}_\Theta(\angle\xi), \quad \textit{and} \quad \hat{\nu}(\xi) = \frac{1}{\sigma_V\|\xi\|}. \tag{3.8}$$

Proof. See appendix A for the proof.

3.2.1 Expression for $\mathbb{P}_{\|V-v_0\|}$. Equation 3.7 states that in order to obtain a perfect equivalence between the MC defined by equations 2.7 and 3.2, the function $\mathcal{L}^{-1}(h)$ has to be well defined. Therefore, we need to compute the inverse transform of the linear operator $\mathcal{L}$:

$$\forall u \in \mathbb{R}, \quad \mathcal{L}(f)(u) = 2\int_0^{\pi/2} f(-u/\cos(\varphi))d\varphi.$$

This is done for the function h in appendix B, proposition 7.

3.3 AR(2) Discretization of the sPDE. Most previous work on gaussian texture synthesis (such as Galerne et al., 2011b, for static and Sanz-Leon et al., 2012, Simoncini et al., 2012, for dynamic textures) has used a global Fourier-based approach and the explicit power spectrum expression, equation 2.7. The main drawbacks of such an approach are that (1) it introduces an artificial periodicity in time and thus can only be used to synthesize a finite number of frames; (2) these frames must be synthesized at once, before the stimulation, which prevents real-time synthesis; and (3) the discrete computational grid may introduce artifacts—in particular, when one of the included frequencies is of the order of the discretization step or a bandwidth is too small.

To address these issues, we follow the previous works of Doretto et al. (2003) and Xia et al. (2014) and make use of an autoregressive (AR) discretization of the sPDE, equation 3.2. In contrast with this previous work, we use a second-order AR(2) regression (instead of a first-order AR(1) model). Using higher-order recursions is crucial to make the output consistent with the continuous formulation equation, 3.2. Indeed, numerical simulations show that AR(1) iterations lead to unacceptable temporal artifacts. In particular, the time correlation of AR(1) random fields typically decays too fast in time.

3.3.1 AR(2) Synthesis without Global Translation, $v_0 = 0$. The discretization computes a (possibly infinite) discrete set of 2D frames $(I_0^{(\ell)})_{\ell\geqslant\ell_0}$ separated by a time-step Δ, and we approach the derivatives at time $t = \ell\Delta$ as

$$\frac{\partial I_0(\cdot,t)}{\partial t} \approx \Delta^{-1}(I_0^{(\ell)} - I_0^{(\ell-1)}) \quad \text{and}$$

$$\frac{\partial^2 I_0(\cdot,t)}{\partial t^2} \approx \Delta^{-2}(I_0^{(\ell+1)} + I_0^{(\ell-1)} - 2I_0^{(\ell)}),$$

which leads to the following explicit recursion,

$$\forall\, \ell \geqslant \ell_0, \quad I_0^{(\ell+1)} = (2\delta - \Delta\alpha - \Delta^2\beta) \star I_0^{(\ell)} \\ + (-\delta + \Delta\alpha) \star I_0^{(\ell-1)} + \Delta^2 W^{(\ell)}, \tag{3.9}$$

where δ is the 2D Dirac distribution and where $(W^{(\ell)})_\ell$ are i.i.d. 2D gaussian field with distribution $\mathcal{N}(0, \Sigma_W)$, and $(I_0^{(\ell_0-1)}, I_0^{(\ell_0-1)})$ can be arbitrary initialized.

One can show that when $\ell_0 \to -\infty$ (to allow for a long enough warmup phase to reach approximate time stationarity) and $\Delta \to 0$, then I_0^Δ defined by interpolating $I_0^\Delta(\cdot, \Delta\ell) = I^{(\ell)}$ converges (in the sense of finite-dimensional distributions) toward a solution I_0 of the sPDE, equation 3.2. Here we choose to use the standard finite difference. However, we refer to Unser, Tafti, Amini, and Kirshner (2014) and Brockwell, Davis, and Yang (2007) for more advanced discretization schemes. We implement the recursion equation 3.9 by computing the 2D convolutions with FFTs on a GPU, which allows us to generate high-resolution videos in real time, without the need to explicitly store the synthesized video.

3.3.2 AR(2) Synthesis with Global Translation. The easiest way to approximate an sPDE cloud using an AR(2) recursion is to simply apply formula 3.1 to $(I_0^{(\ell)})_\ell$ as defined in equation 3.9, that is, to define

$$I^{(\ell)}(x) \overset{\text{def.}}{=} I_0^{(\ell)}(x - v_0\Delta\ell).$$

An alternative approach would consist in directly discretizing the sPDE, equation 3.5. We did not use this approach because it requires the discretization of spatial differential operators $\mathcal{G}$ and $\mathcal{H}$ and is, hence, less stable. A third, somehow hybrid, approach, is to apply the spatial translations to the AR(2) recursion and define the following recursion,

$$I^{(\ell+1)} = \mathcal{U}_{v_0} \star I^{(\ell)} + \mathcal{V}_{v_0} \star I^{(\ell-1)} + \Delta^2 W^{(\ell)}, \tag{3.10}$$

$$\text{where} \quad \begin{cases} \mathcal{U}_{v_0} \overset{\text{def.}}{=} (2\delta - \Delta\alpha - \Delta^2\beta) \star \delta_{-\Delta v_0}, \\ \mathcal{V}_{v_0} \overset{\text{def.}}{=} (-\delta + \Delta\alpha) \star \delta_{-2\Delta v_0}, \end{cases} \tag{3.11}$$

where δ_s indicates the Dirac at location s, so that $(\delta_s \star I)(x) = I(x - s)$ implements the translation by s. Numerically, it is possible to implement equation 3.10 over the Fourier domain,

$$\hat{I}^{(\ell+1)}(\xi) = \hat{\mathcal{U}}_{v_0}(\xi)\hat{I}^{(\ell)}(\xi) + \hat{\mathcal{V}}_{v_0}(\xi)\hat{I}^{(\ell-1)}(\xi) + \Delta^2\hat{\sigma}_W(\xi)\hat{w}^{(\ell)}(\xi),$$

$$\text{where} \quad \begin{cases} \hat{\mathcal{U}}_{v_0}(\xi) \overset{\text{def.}}{=} (2 - \Delta\hat{\alpha}(\xi) - \Delta^2\hat{\beta}(\xi))e^{-\mathrm{i}\Delta v_0 \xi}, \\ \hat{\mathcal{Q}}_{v_0}(\xi) \overset{\text{def.}}{=} (-1 + \Delta\hat{\alpha}(\xi))e^{-2\mathrm{i}\Delta v_0 \xi}, \end{cases}$$

and where $w^{(\ell)}$ is a 2D white noise.

4 An Empirical Study of Visual Speed Discrimination

To exploit the useful parametric transformation features of our MC model and provide a generalizable proof of concept based on motion perception, we consider here the problem of judging the relative speed of moving dynamical textures. The overall aim is to characterize the impact of both average spatial frequency and average duration of temporal correlations on perceptual speed estimation, based on empirical evidence.

4.1 Methods. The task was to discriminate the speed $v \in \mathbb{R}$ of an MC stimulus moving with a horizontal central speed $\mathbf{v} = (v, 0)$. We assign as the independent experimental variable the most represented spatial frequency z_0, denoted z in the rest of the letter for easier reading. The other parameters are set to the following values,

$$\sigma_V = \frac{1}{t^\star z}, \quad \theta_0 = \frac{\pi}{2}, \quad \sigma_\Theta = \frac{\pi}{12}.$$

Note that σ_V is thus dependent on the value of z to ensure that $t^\star = \frac{1}{\sigma_V z}$ stays constant. This parameter $t^\star$ controls the temporal frequency bandwidth, as illustrated in the middle of Figure 2. We used a two-alternative forced-choice (2AFC) paradigm. In each trial, a gray fixation screen with a small dark fixation spot was followed by two stimulus intervals of 250 ms each, separated by an uniformly gray 250 ms interstimulus interval. The first stimulus had parameters (v_1, z_1), and the second had parameters (v_2, z_2). At the end of the trial, a gray screen appeared asking the participant to report which one of the two intervals was perceived as moving faster by pressing one of two buttons—that is, whether $v_1 > v_2$ or $v_2 > v_1$.

Given reference values $(v^\star, z^\star)$, for each trial, (v_1, z_1) and (v_2, z_2) are selected such that

$$\begin{cases} v_i = v^\star, \ z_i \in z^\star + \Delta_Z \\ v_j \in v^\star + \Delta_V, \ z_j = z^\star \end{cases} \quad \text{where} \quad \Delta_V = \{-2, -1, 0, 1, 2\},$$

where $(i, j) = (1, 2)$ or $(i, j) = (2, 1)$ (i.e., the ordering is randomized across trials) and where z values are expressed in cycles per degree (c/°) and v values in °/s. The range Δ_Z is defined in Table 2. Ten repetitions of each of the 25 possible combinations of these parameters are made per block of 250 trials and at least 4 of such blocks were collected per condition tested.

Table 2: Stimulus Parameters for the Range of Tested Experimental Conditions.

Case	$t^\star$	σ_Z	B_Z	$v^\star$	$z^\star$	Δ_Z
A1	200 ms	1.0 c/°	×	5 °/s	0.78 c/°	$\{-0.31, -0.16, 0, 0.16, 0.47\}$
A2	200 ms	1.0 c/°	×	5 °/s	1.25 c/°	$\{-0.47, -0.31, 0, 0.31, 0.63\}$
A3	200 ms	×	1.28	5 °/s	1.25 c/°	$\{-0.47, -0.31, 0, 0.31, 0.63\}$
A4	100 ms	×	1.28	5 °/s	1.25 c/°	$\{-0.47, -0.31, 0, 0.31, 0.63\}$
A5	200 ms	×	1.28	10 °/s	1.25 c/°	$\{-0.47, -0.31, 0, 0.31, 0.63\}$

Note: A1 and A2 in the first two rows are both bandwidth controlled in c/° and A3 to A5 are bandwidth controlled in octaves with high (A3 and A5) and low (A4) $t^\star$.

The outcome of these experiments are summarized by psychometric curve samples $\hat{\varphi}_{v^\star,z^\star}$, where for all $(v - v^\star, z - z^\star) \in \Delta_V \times \Delta_Z$, the value $\hat{\varphi}_{v^\star,z^\star}(v, z)$ is modeled as a Bernoulli random variable with parameter $\varphi_{v^\star,z^\star}(v, z)$ that a stimulus generated with parameters $(v^\star, z)$ is moving faster than a stimulus with parameters $(v, z^\star)$.

We tested different scenarios summarized in Table 2. Each row corresponds to approximately 35 minutes of testing per participant and was always performed by at least three of the participants. Stimuli were generated using Matlab 7.10.0 on a Mac running OS 10.6.8 and displayed on a 20″ Viewsonic p227f monitor with resolution 1024 × 768 at 100 Hz. Psychophysics routines were written using Matlab, and Psychtoolbox 3.0.9 controlled the stimulus display. Observers sat 57 cm from the screen in a dark room. Five male observers with normal or corrected-to-normal vision took part in these experiments. They gave their informed consent, and the experiments received ethical approval from the Aix-Marseille Ethics Committee in accordance with the declaration of Helsinki.

4.2 Psychometric Results. Estimating speed in dynamic visual scenes is undoubtedly a crucial skill for successful interaction with the visual environment. Human judgments of perceived speed have therefore generated much interest and been studied with a range of psychophysics paradigms. The different results obtained in these studies suggest that rather than computing a veridical estimate, the visual system generates speed judgments influenced by contrast (Thompson, 1982), speed range (Thompson, Brooks, & Hammett, 2006), luminance (Hassan & Hammett, 2015), spatial frequency (Brooks, Morris, & Thompson, 2011; Simoncini et al., 2012; Smith, Majaj, & Movshon, 2010), and retinal eccentricity (Hassan, Thompson, & Hammett, 2016). There are currently no theoretical models of the underlying mechanisms serving speed estimation that capture this dependence on such a broad range of image characteristics. One of the reasons for this might be that the simplified grating stimuli used in most of the previous studies do not allow experimenters to shed light on the possible elaborations in neural processing that arise when more complex natural or naturalistic stimulation is used. Such elaborations, like nonlinearities in spatiotemporal frequency space, can be seen in their simplest form even with a superposition of a pair of gratings (Priebe, Cassanello, & Lisberger, 2003). In the current work, we use our formulation of motion cloud stimuli, which allows for separate parametric manipulation of peak spatial frequency (z), spatial frequency bandwidth (B_z, σ_z), and stimulus lifetime ($t^\star$), which is inversely related to the temporal variability. The stimuli are all broadband, closely resembling the frequency properties under natural stimulation. Our approach is to test five participants under several parametric conditions given in Table 2 and using a large number of trials.

4.2.1 Psychometric Function Estimation. The psychometric function is estimated by the following sigmoidal template function,

$$\varphi^{\mu,\Sigma}_{v^\star,z^\star}(v,z) = \psi\left(\frac{v - v^\star - \mu_{z,z^\star}}{\Sigma_{z,z^\star}}\right), \tag{4.1}$$

where $\psi(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-s^2/2}\mathrm{d}s$ is the cumulative normal function and $(\mu_{z,z^\star}, \Sigma_{z,z^\star})$ denotes, respectively, bias and inverse sensitivity. The collected data are used to fit the two parameters using maximum likelihood estimation (see Wichmann & Hill, 2001),

$$(\hat{\mu}, \hat{\Sigma}) = \underset{\mu,\Sigma}{\operatorname{argmin}} \sum_{v} \mathrm{KL}(\hat{\varphi}_{v^\star,z^\star}|\varphi^{\mu,\Sigma}_{v^\star,z^\star}),$$

where $\mathrm{KL}(\hat{p}|p)$ is the Kullback-Leibler divergence between samples $\hat{p}$ and model p under a Bernouilli distribution:

$$\mathrm{KL}(\hat{p}|p) = \hat{p}\log\left(\frac{\hat{p}}{p}\right) + (1-\hat{p})\log\left(\frac{1-\hat{p}}{1-p}\right).$$

Results of these estimations are shown in Figure 8 for both nonparametric and linear $\Sigma_{z,z^\star}$ fits.

Remark 2. In practice we perform the fit in the log-speed domain, that is, we consider $\varphi_{\tilde{v}^\star,z^\star}(\tilde{v},z)$ where $\tilde{v} = \ln(1 + v/v_0)$ with $v_0 = 0.3^\circ/\mathrm{s}$ following Stocker and Simoncelli (2006). As the estimated bias $\tilde{\mu}$ is obtained in the log-speed domain, we convert it back to the speed domain by computing μ, which solves the following equation:

$$\log(1 + (v^\star + \mu)/v_0) = \log(1 + v^\star/v_0) + \tilde{\mu}.$$

Then the speed bias is $\mu = (v_0 + v^\star)(\exp(\tilde{\mu}) - 1)$.

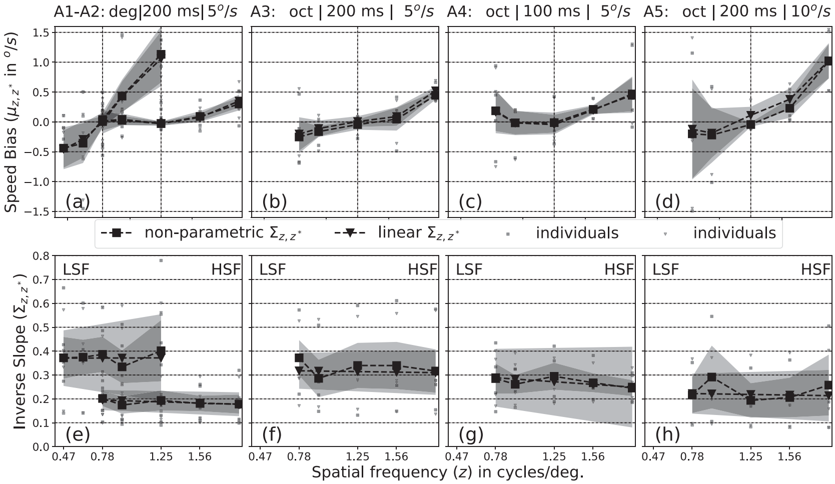


Figure 4: Averaged results over participants for the perceptual biases (top row) and inverse slope (bottom row) plotted against the tested central spatial frequency z. The specific parameters for each column are indicated above: bandwidth in degrees (deg) or octaves (oct.), value of stimulus lifetime $t^\star$, and reference speed $v^\star$. Small markers represent individual results, and large markers represent population average. (a–d) Speed biases, which generally show an increase at higher frequencies, but with individual differences. (e–h) Inverse psychometric slopes that generally appear to be constant or decreasing across frequency. From left to right: conditions A1–A2, A3, A4, and A5 (see Table 2 for details).

4.2.2 Cycle-Controlled Bandwidth Conditions. The main manipulation in each case is the direct comparison of the speed of a range of five stimuli in which the central spatial frequency varies between five values, but all other parameters are equated under the different conditions. In a first manipulation in which bandwidth is controlled by fixing it at a value of 1 c/° for all stimuli (conditions A1 and A2 in Table 2), we find that lower frequencies are consistently perceived to be moving slower than higher frequencies (see Figure 4a). This trend is the same for both the lower and the higher spatial frequency ranges used in the tasks, yet the biases are smaller for the higher frequency range (see A1 and A2 in Table 2 for details). This suggests that the effect generalizes across the two scales used, but that shifting the central spatial frequency value of the stimulus, which forms the reference scale, results in a change in sensitivity during speed discrimination. For example, comparing A1 and A2 performance in Figure 4, when the five stimuli of different speed that make up the reference scale are changed from $z^\star = 0.78$ (A1) to $z^\star = 1.25$ (A2), speed estimates seem to become less reliable. The same comparison is using a different psychometric measurement scale in each case. The sensitivity to the discrimination of stimuli measured in the inverse of the psychometric slope is found to remain approximately constant across the range of frequency tested for each of the tested spatial frequencies (see Figure 4e). However, the sensitivity increases significantly ($\Sigma_{z^\star,z}$ decreases) from condition A1 to condition A2. Such an effect suggests that an increasing trend in sensitivity may exist (see section 4.2.4).

4.2.3 Octave-Controlled Bandwidth Conditions. The octave-bandwidth-controlled stimuli of conditions A3 to A5 (see Table 2) allow us to vary the spatial frequency manipulations (z) in a way that generates scale-invariant bandwidths exactly as would be expected from zooming movements toward or away from scene objects (see Figure 1). Thus, if trends seen in Figure 4e were purely the result of ecologically invalid fixing of bandwidths at 1 c/° in the manipulations, this would be corrected in the current manipulation. Only the higher-frequency comparison range from conditions A2 is used because trends are seen to be consistent across conditions A1 and A2. We find that the trends are generally the same as in Figure 4a. Indeed higher spatial frequencies are consistently perceived as faster than lower ones, as shown in Figures 4b to 4d. Interestingly, for the degree bandwidth-controlled stimuli, the biases are lower than those for the equivalent octave-controlled stimuli (e.g., compare Figure 4a with 4b). This can also be seen in Figure 10 (conditions A2 and A3). A change in the bias is also seen with the manipulation of $t^\star$, as increasing temporal frequency variability when going from biases in Figure 4b to those in Figure 4c entails a reduction in measured biases, with an effect of about 25%, which is also visible in Figure 10 (conditions A3 and A4 for M1).

4.2.4 Is Sensitivity Dependent on Stimulus Spatial Frequency? To explore further the sensitivity trend, we fit the data with a psychometric function by assuming a linear model for $\Sigma_{z,z^\star}$ and test for a significant negative slope. None of the slopes are significantly different from 0 at the population level. At the individuals' level, among all conditions and subjects, we find that 13 out of 21 slopes were significantly decreasing. Therefore, we interpret this as a possible decrease in sensitivity at higher z seen in 13 out of 21 of the cases, but one that shows large individual differences in sensitivity trends.

4.2.5 Qualitative Results Summary

- Spatial frequency has a positive effect on perceived speed ($\mu_{z,z^\star}$ increases as z increases).
- The inverse sensitivity remains constant or is decreasing with spatial frequency (resp. $\Sigma_{z,z^\star}$ does not depend on z or decreases as z increases), but there are large individual differences in this sensitivity change.

In the next section, we detail a Bayesian observer model to account for these observed effects.

4.3 Observer Model. We list here the general assumptions underlying our model:

1. The observer performs abstract measurement of the stimulus denoted by a real random variable M.
2. The observer estimates speed using an estimator based on the posterior speed distribution $\mathbb{P}_{V|M}$.
3. The posterior distribution is implicit; Bayes' rule states that $\mathbb{P}_{V|M} \propto \mathbb{P}_{M|V}\mathbb{P}_V$.
4. The observer knows all other stimulus parameters (in particular, the spatial frequency z).
5. The observer takes a decision without noise.

These asssumptions correspond to the ideal Bayesian observer model. We detail below the relation between this model and the psychometric bias and inverse sensitivity $(\mu_{z,z^\star}, \Sigma_{z,z^\star})$. We also give details to derive the likelihood directly from the MC model and discuss the expected consequences.

4.3.1 Ideal Bayesian Observer. Assumptions 1 to 5 correspond to the methodology of the Bayesian observer used for instance in Stocker and Simoncelli (2006), Sotiropoulos et al. (2014), and Jogan and Stocker (2015). This previous work provides the foundation for the work on Bayesian observer models in perception on which we build our modifications accounting for our naturalistic dynamic stimulus case. We assume that the posterior speed distribution may depend on spatial frequency because any observed effects must come from the change in spatial frequency and the effect it may have on the likelihood. This assumption is also motivated by a body of empirical evidence showing consistent effects of spatial frequency changes on speed estimation (Brooks et al., 2011; Vacher et al., 2015). Findings from primate neurophysiology probing extrastriate cortical neurons with compound gratings also show that speed is estimated by neural units whose speed response (i.e., not just response variance associated with likelihood widths) is highly dependent on spatiotemporal frequency structure (Priebe et al., 2003; Perrone & Thiele, 2001). Finally, we also assume that the observer measures speed using a maximum a posteriori (MAP) estimator,

$$\begin{aligned}\hat{v}(m) &= \underset{v}{\operatorname{argmax}}\, \mathbb{P}_{V|M,Z}(v|m,z)\\ &= \underset{v}{\operatorname{argmin}}[-\log(\mathbb{P}_{M|V,Z}(m|v,z)) - \log(\mathbb{P}_{V|Z}(v|z))],\end{aligned} \tag{4.2}$$

computed from the internal representation $m \in \mathbb{R}$ of the observed stimulus. Note that the distribution of measurements (the likelihood) and the prior

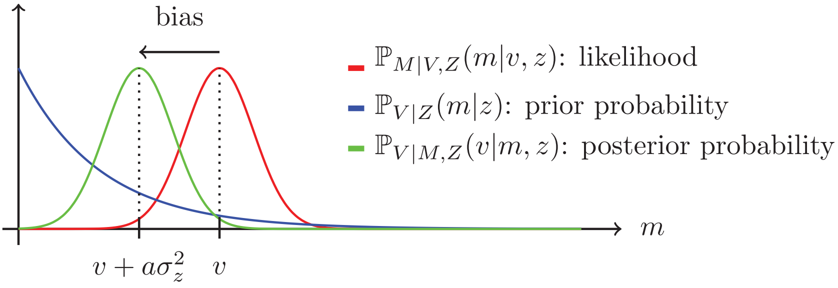


Figure 5: Multiplying a gaussian likelihood by a Laplacian prior gives a gaussian posterior that is similar to a shifted version of the likelihood.

are both conditioned on spatial frequency z. As the likelihood is also obviously conditioned on speed, we denote measurement as $M_{v,z}$. To simplify the numerical analysis, we assume a gaussian likelihood (in log-speed domain), with a variance independent of v consistently with the previous literature (Stocker & Simoncelli, 2006; Sotiropoulos et al., 2014; Jogan & Stocker, 2015). Furthermore, we assume that the prior is Laplacian (in log-speed domain) as this gives a good description of the a priori statistics of speeds in natural images (Dong, 2010),

$$\mathbb{P}_{M|V,Z}(m|v,z) = \frac{1}{\sqrt{2\pi}\sigma_z} e^{-\frac{|m-v|^2}{2\sigma_z^2}} \quad \text{and} \quad \mathbb{P}_{V|Z}(v|z) = \mathbb{P}_V(v) \propto e^{av}, \tag{4.3}$$

where $a < 0$.

Remark 3. We initially assume that the posterior speed distribution is conditioned on spatial frequency; thus, the likelihood and prior distributions also depend on spatial frequency. However, there is currently no conclusive support in favor of a spatial frequency-dependent speed prior in the literature, but evidence of spatial frequency influencing speed estimation is discussed in the previous paragraph. Therefore, only the likelihood width σ_z depends on spatial frequency z, and the log-prior slope a does not. We discuss in more detail the choice of the likelihood and its dependence on spatial frequency in section 4.3.2.

Figure 5 shows an example of how the likelihood and prior described in equation 4.3 combine into a posterior distribution that resembles a shifted version of the likelihood. In practice, we are able to compute the distribution of the estimates $\hat{v}(M_{v,z})$ as stated in the following proposition:

Proposition 5. *In the special case of the MAP estimator, equation 4.2 with a parameterization defined in equation 4.3, one has*

$$\hat{v}(M_{v,z}) \sim \mathcal{N}(v + a\sigma_z^2, \sigma_z^2). \tag{4.4}$$

Once the observer has estimated the speed of two presented stimuli, he must take a decision to judge which stimulus was faster. Following assumption 5, the decision is ideal in the sense that it is performed without noise. In other words, the observer compares the two speeds and decides whether $(\hat{v}(m_{v,z^\star}), \hat{v}(m_{v^\star,z}))$ belongs to the decision set $E = \{(v_1, v_2) \in \mathbb{R}^2 | v_1 > v_2\}$. Thus, we define the theoretical psychometric curve of an ideal Bayesian observer as

$$\varphi_{v^\star,z^\star}(v, z) \overset{\text{def.}}{=} \mathbb{E}(\hat{v}(M_{v,z^\star}) > \hat{v}(M_{v^\star,z})).$$

Following proposition 5, in our special case of gaussian likelihood and Laplacian prior, the psychometric curve can be computed in closed form.

Proposition 6. *In the special case of the MAP estimator, equation 4.2, with a parameterization defined in equation 4.3, one has*

$$\varphi_{v^\star,z^\star}(v, z) = \varphi^{a,\sigma}_{v^\star,z^\star}(v, z) \overset{def.}{=} \psi\left(\frac{v - v^\star + a(\sigma_{z^\star}^2 - \sigma_z^2)}{\sqrt{\sigma_{z^\star}^2 + \sigma_z^2}}\right), \tag{4.5}$$

where ψ is defined in equation 4.1.

Proof. See appendix A for the proof.

Proposition 6 provides the connection between the Bayesian model parameters and the classical psychometric measures of bias and sensitivity. In particular, it explains the heuristic sigmoidal templates commonly used in psychophysics (see section 4.2). An example of two psychometric curves is shown in Figure 6. We have the following relations:

$$\mu_{z,z^\star} = a(\sigma_z^2 - \sigma_{z^\star}^2), \tag{4.6}$$

$$\Sigma_{z,z^\star}^2 = \sigma_{z^\star}^2 + \sigma_z^2. \tag{4.7}$$

Remark 4. The experiment allows us to estimate bias and inverse sensitivity $(\mu_{z,z^\star}, \Sigma_{z,z^\star})$. Knowing these parameters, it is possible to recover parameters of the ideal Bayesian observer model. Equation 4.7 has a unique solution, and equation 4.6 can be solved using the least square estimator.

Remark 5. Under this model, a positive bias comes from a decrease in likelihood width and a negative log-prior slope. As concluded in section 4.2, we observe a significant decrease in inverse sensitivity in 13 of 21 subjects and conditions. Therefore, the model, when fitted to the data, will force the likelihood width to decrease. Further experiments will be necessary to verify the significance of this observation. Yet, the model is well supported by the

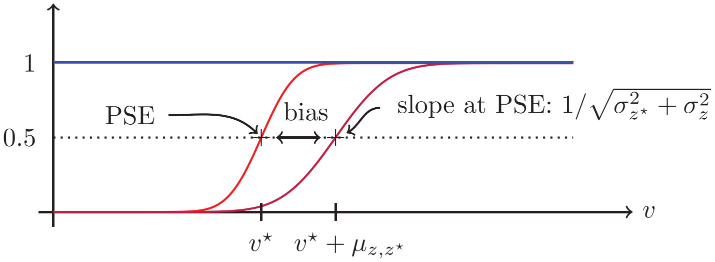


Figure 6: The shape of the psychometric function follows the estimation of the two speeds by the Bayesian inference described in Figure 5. This figure illustrates proposition 6. The bias ensues from the difference between the bias on the two estimated speeds.

literature (see Stocker & Simoncelli, 2006; Jogan & Stocker, 2015; Sotiropoulos et al., 2014) and is compatible with the properties of the stimuli (see section 4.3.2).

4.3.2 Discussion: Likelihood. An MC $I_{v,z}$ is a random gaussian field of power spectrum defined by equation 2.7, with central speeds $v_0 = (v, 0)$ and central spatial frequency z (the other parameters being fixed, as explained in section 4.1). Assuming that the abstract measurements correspond to the presented frames, $M_{v,z} = I_{v,z}$, it is possible to use the MC generative model as a likelihood. In the absence of a prior, the MAP estimator is equal to the maximum likelihood estimator (MLE):

$$\hat{v}(m) = \hat{v}^{\mathrm{MLE}}(i) = \underset{v}{\operatorname{argmin}} - \log(\mathbb{P}_{I|V,Z}(i|v,z)). \tag{4.8}$$

Thanks to the sPDE formulation, it is possible to give a simple rigorous expression for $-\log(\mathbb{P}_{I|V,Z}(i|v,z))$ in the case of discretized clouds satisfying the AR(2) recursion equation, 3.10. In this case, for some input video $I_{v,z} = (I^{(\ell)})_{\ell=1}^{L}$, the log likelihood reads

$$-\log(\mathbb{P}_{I|V,Z}(I_{v,z}|v,z)) = \tilde{Z}_I + K_{v_0}(I_{v,z}) \quad \text{where}$$

$$K_{v_0}(I_{v,z}) \overset{\text{def.}}{=} \frac{1}{\Delta^4} \sum_{\ell=1}^{L} \int_{\Omega} |K_W \star I^{(\ell+1)}(x)$$

$$-\mathcal{U}_{v_0} \star K_W \star I^{(\ell)}(x) - \mathcal{V}_{v_0} \star K_W \star I^{(\ell-1)}(x)|^2 dx,$$

where $\mathcal{U}_{v_0}$ and $\mathcal{V}_{v_0}$ are defined in equation 3.11 and K_W is the spatial filter corresponding to the square root inverse of the covariance Σ_W, that is,

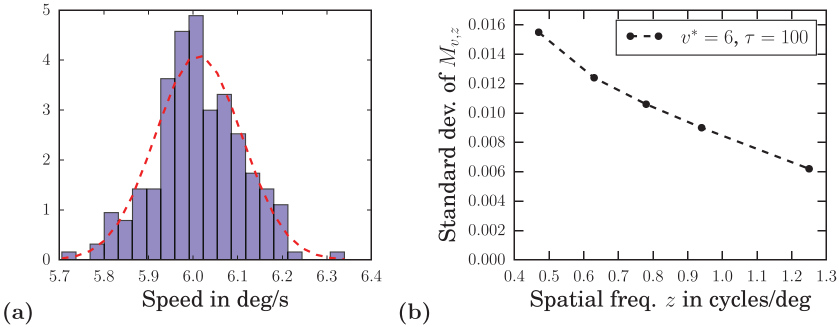


Figure 7: Simulation of the speed distributions of a set of motion clouds with the experimentally tested parameters. (a) Histogram of estimates of $\hat{v}^{\text{MLE}}(I_{v,z})$ for $z = 0.8$ c/° defined by equation 4.8. These estimates are well approximated by a gaussian distribution (red dashed line). (b) Standard deviation of estimates of $\hat{v}^{\text{MLE}}(I_{v,z})$ as a function of z. The standard deviation of these estimates is inversely proportional to the spatial frequency z.

which satisfies $\hat{K}_W(\xi) \overset{\text{def.}}{=} \hat{\sigma}_W(\xi)^{-1}$. This convenient formulation can be used to rewrite the MLE estimator of the horizontal speed v parameter of a MC as

$$\hat{v}^{\text{MLE}}(i) = \underset{v}{\operatorname{argmin}}\, K_{v_0}(i) \quad \text{where} \quad v_0 = (v, 0) \in \mathbb{R}^2, \tag{4.9}$$

where we used the fact that $\tilde{Z}_I$ is independent from v_0.

The solution to this optimization problem with respect to v is computed using the Newton-CG optimization method implemented in the Python library `scipy`. In Figure 7a, we show a histogram of speed estimates $\hat{v}^{\text{MLE}}(I_{v,z})$ performed over 200 motion clouds generated with speed $v = 6$ °/s and spatial frequency $z = 0.78$ c/°. In Figure 7b, we show the evolution of the standard deviation of speed estimates $\hat{v}^{\text{MLE}}(I_{v,z})$ as a function of spatial frequencies $z \in \{0.47$ c/°, 0.62 c/°, 0.78 c/°, 0.94 c/°, 1.28 c/°$\}$. For each spatial frequency, estimates are again similarly obtained over a set of 200 motion clouds generated with speed $v = 6$ °/s. First, we observed that $\hat{v}^{\text{MLE}}(I_{v,z})$ is well approximated by a gaussian random variable with mean v. Second, the standard deviation of these estimates decreases when the spatial frequency increases. The two conclusions follow the fact that our model is gaussian and that we impose the relation $\sigma_V = 1/(t^\star z)$—that standard deviation of speed is inversely proportional to spatial frequency. The decreasing trend combined with a prior for slow speed $a < 0$ would reproduce the positive bias of spatial frequency over speed perception observed in section 4.2. If a human subject were estimating speed in such an optimal way, equation 4.7 indicates that inverse sensitivity $\Sigma_{z,z^\star}$ would also be inversely proportional to spatial frequency. Yet the primary analysis conducted in section 4.2 does not give a clear trend for the inverse sensitivity. As a consequence, that analysis is ambiguous and does not allow us to make definitive conclusions about the compatibility of the MC model and the existing literature with the observed human performances.

4.4 Likelihood and Prior Estimation. In order to fit this model to our data, we use an iterative two-step method, each minimizing the Kullback-Leibler divergence between the model and its samples. This process is the equivalent of a maximum likelihood estimate. The first step is to fit each psychometric curve individually and the second step to use the results as a starting point to fit all the psychometric curves together. Numerically, we used the Nelder-Mead simplex method as implemented in the Python library `scipy`.

Step 1: For all z, $z^\star$, initialized at a random point, compute

$$(\bar{\mu}, \bar{\Sigma}) = \underset{\mu, \Sigma}{\operatorname{argmin}} \sum_{v} \mathrm{KL}(\hat{\varphi}_{v^\star, z^\star} | \varphi^{\mu, \Sigma}_{v^\star, z^\star}),$$

where $\varphi^{\mu, \Sigma}_{v^\star, z^\star}$ is defined in equation 4.1.

Step 2: Solve equations 4.6 and 4.7 between $(\bar{\mu}, \bar{\Sigma})$ and $(\bar{a}, \bar{\sigma})$, initialize at $(\bar{a}, \bar{\sigma})$, and compute

$$(\hat{a}, \hat{\sigma}) = \underset{a, \sigma}{\operatorname{argmin}} \sum_{z, z^\star} \sum_{v} \mathrm{KL}(\hat{\varphi}_{v^\star, z^\star} | \varphi^{a, \sigma}_{v^\star, z^\star}),$$

where $\varphi^{a, \sigma}_{v^\star, z^\star}$ is defined in equation 4.5.

We use a repeated stochastic initialization in the first step in order to overcome the presence of local minima encountered during the fitting process. The approach was found to exhibit better results than a direct and global fit (third point).

4.5 Modeling Results. We use the Bayesian formulation detailed in section 4.3.1 and the fitting process described in section 4.4 to estimate, for each subject, the likelihood widths and the corresponding log-prior slopes under the tested experimental conditions. We plot in Figure 8 the fit of bias and inverse sensitivity for the sigmoid model (see section 4.2) and Bayesian model (see section 4.3.1) averaged over subjects. Figure 9 displays the corresponding likelihood widths and log-prior slopes for the Bayesian model also averaged over subjects. Finally, Figure 10 summarizes the total bias between extremal tested spatial frequencies for each experimental condition and for both models. Error bars correspond to the standard deviation of the mean.

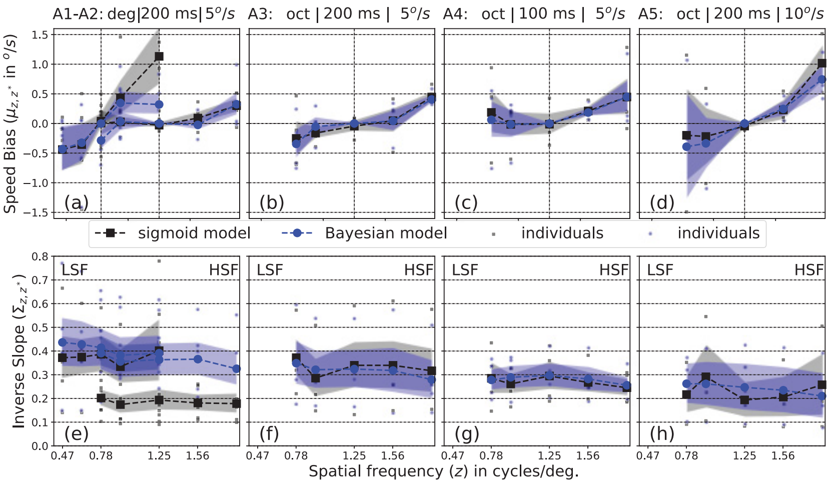


Figure 8: The model fitted speed biases (top row) and inverse sensitivity (bottom row) for the different conditions for the Bayesian model (blue) and the sigmoid model (black). (a–d) Speed biases generally increase with increasing spatial frequency. (e–h) Inverse sensitivity tends to decrease for the Bayesian model but is configured not to do so for the sigmoid model. The parameters are indicated above, respectively: bandwidth in octave (oct.) or degree (deg.), value of stimulus lifetime $t^\star$, and reference speed $v^\star$. Small markers represent individual results, and large markers represent population average. From left to right: conditions A1–A2, A3, A4, and A5.

4.5.1 Measured Biases and Inverse Sensitivity. As shown in Figure 8, both models M2 and M3 correctly account for the biases and inverse sensitivity estimated with model M1 (see section 4.2) except for conditions A1 and A2. For condition A1 (see Figure 8a), the bias is underestimated by models M2 and M3 compared to model M1. For condition A2 (see Figure 8e), the inverse sensitivity is overestimated by models M2 and M3 compared to model M1. The observed differences come from the fact that in models M2 and M3, the overlapping spatial frequencies of conditions A1 and A2 are pulled together. As a consequence, the fit is more constrained than for model M1 and is therefore smoother. While that discrepancy does not affect our conclusion, it raises the question of pulling different overlapping conditions together. The overlapping tested spatial frequencies are together, whereas they were collected with different reference spatial frequencies such that the sensitivity of each of the psychometric speed measurement scales appears to have been different. Despite averaging over subjects, the Bayesian estimates of inverse sensitivity appear smoother than the sigmoid estimates (see Figures 8f and 8h). Finally, a clearer decreasing trend is visible in the Bayesian estimates of inverse sensitivity.

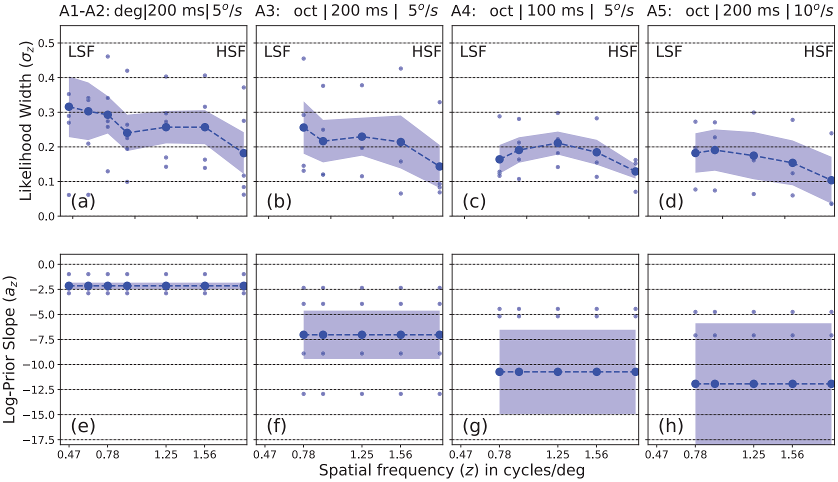


Figure 9: The model-fitted likelihood widths (top row) and log-prior slopes (bottom row) for the different conditions for the Bayesian model. (a–d) Likelihood widths tend to decrease with increasing spatial frequency. (e–h) Log-prior slopes are negative and higly variable between subjects. The parameters are indicated above, respectively: bandwidth in octave (oct.) or degree (deg.), value of stimulus lifetime $t^\star$ and reference speed $v^\star$. Small markers represent individual results, and large markers represent population average. From left to right: conditions A1–A2, A3, A4, and A5.

4.5.2 Corresponding Sensory Likelihood Widths. There is a systematic decreasing trend within the likelihood width fits in Figures 9a, 9b, and 9d, which shows an inverted U-shape. The fact that all subjects did not run all experimental conditions explains this difference (two subjects out of four show a U-shape bias; see Figure 8c). Subject-to-subject variability is similar for all conditions except for the least temporal variability for which it is smaller (see Figure 9c).

4.5.3 Corresponding Log-Prior Slopes. The log-prior slope estimates have a high subject-to-subject variability for conditions A3 to A5 (see Figures 9f to 9h) compared to conditions A1–A2 (see Figure 9e). The high intersubject variability is expected in speed discrimination tasks, and in the case of conditions A4 and A5, this is particularly magnified by two subjects that have an extremely low value for a_z (their small markers are not visible in Figures 9e and 9f).

4.6 Insights into Human Speed Perception. We exploited the principled and ecologically motivated parameterization of MC to study biases in human speed judgements under a range of parametric conditions.

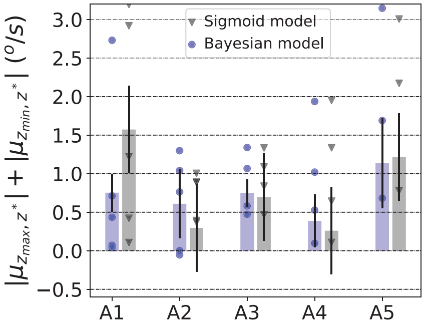


Figure 10: *Biases amplitude.* Sum of the absolute biases at lower and higher spatial frequency averaged over participants. Small markers represent individual results, bars represent population average, and error bars represent one standard error of the mean.

Primarily, we consider the effect of scene scaling on perceived speed, manipulated via central spatial frequencies in a similar way to previous experiments that have shown spatial frequency-induced perceived speed biases (Brooks et al., 2011; Smith & Edgar, 1990). In general, our experimental result confirms that higher spatial frequencies are consistently perceived to be moving faster than compared lower frequencies, which is the same result as reported in a previous study using both simple gratings and compounds of paired gratings, the second of which can be considered as a relatively broadband bandwidth stimulus (Brooks et al., 2011) compared to single grating stimuli, without considering the inhibitive interactions we know to occur when multiple gratings are superimposed (Priebe et al., 2003). In that work, they noted that biases were present but slightly reduced in the compound (broadband) stimuli. That conclusion is consistent with a more recent psychophysics manipulation in which up to four distinct composite gratings were used in relative speed judgments. Estimates were found to be closer to veridical as bandwidth was increased by adding components from the set of four, but increasing spatial frequencies generally biased toward faster perceived speed, even if individual participants showed different trends (Jogan & Stocker, 2015). Indeed, findings from primate neurophysiology studies have also noted that while responses are biased by spatial frequency, the tendency toward true speed sensitivity (measured as the proportion of individual neurons showing speed sensitivity) increases when broadband stimulation is used (Priebe et al., 2003; Perrone & Thiele, 2001). A model of visual motion sensitivity with a hierarchical framework, selectively reading from and optimally decoding V1 inputs in an MT layer, has also been tested. It was found to be consistent with human speed sensitivity to natural images (Burge & Geisler, 2015).

It is increasingly being recognized that linear systems approaches to interrogating visual processing with single sinusoidal luminance grating inputs represent a powerful but limited approach to study speed perception, as they fail to capture the fact that naturalistic broadband frequency distributions may support speed estimation (Brooks et al., 2011; Meso & Simoncini, 2014; Meso & Zanker, 2009; Gekas et al., 2017). A linear consideration, for example, may not fully account for the fact that estimation in the presence of multiple sinusoidal components results in linear optimal combination performing best among alternatives (Jogan & Stocker, 2015). In that case, the simple monotonic increase in perceived speed predicted by the optimal model when components were added to the compound is not seen in the data, particularly in the difference between three and four components. This may be due to interaction between components that are not fully captured by this optimal linear model. Our work seeks to extend the body of previous studies by looking at spatial-frequency-induced biases, using a parametric configuration in the form of motion clouds, which allow a manipulation across a continuous scale of frequency and bandwidth parameters. The effect of frequency interactions across the broadband stimulus defined along the two-dimensional spatiotemporal luminance plane allows us to measure the perceptual effect of the projection of different areas (e.g., see Figure 2) onto the same speed line. The measurement would be the result of proposed inhibitory interactions, which occur during spatiotemporal frequency integration for speed perception (Simoncini et al., 2012; Gekas et al., 2017), which cannot be observed with component stimuli separated by several octaves (Jogan & Stocker, 2015).

We use a slower and a faster speed because previous work using sinusoidal grating stimuli has shown that below the slower range (<8 °/s), uncertainty manipulated through lower contrasts causes an underestimation of speeds, while at faster speeds (>16 °/s), it causes an overestimation, an effect that is not fully explained by Bayesian models with a prior encouraging slow speeds. (Thompson et al., 2006; Hassan & Hammett, 2015). Our findings show that biases are larger at the faster speed than the slower one. Biases are also generally lower for the octave-controlled than for the cycle-controlled stimuli, indicating that the underlying system was better at processing the former.

The Bayesian fitting identifies a decrease in the likelihood width estimates, which could explain the biases in over half of our fitted psychometric functions. For cases of the same frequency range where comparable likelihoods are estimated, some conditions—like the low and high $t^\star$ cases—have very different prior estimates. This result can be interpreted in light of recent work (Gekas et al., 2017): biases might act along the speed line and an orthogonal scale line within the spatiotemporal space, depending on the spread or bandwidth of the stimulus. While the current work does not resolve some of the ongoing gaps in our understanding of speed perception mechanisms, particularly as it does not tackle contrast-related biases, it shows that known frequency biases in speed perception also arise from orthogonal spatial and temporal uncertainties when RMS contrast is controlled—as it is within the MC stimuli. Bayesian models such as the one we apply, which effectively project distributions in the spatiotemporal plane onto a given speed line in which a linear low speed prior applies (Stocker & Simoncelli, 2006), may be insufficient to capture the effect of spatiotemporal priors, which may underlie some of the broad set of empirical results. Individual differences, which are pervasive in these experiments, may also be associated with internal assumptions that can be considered as priors. Indeed for Bayesian models to fully predict speed perception with more complex or composite stimuli, they often require various elaborations away from the simplistic combination of likelihood and low speed prior (Hassan & Hammett, 2015; Gekas et al., 2017; Jogan & Stocker, 2015; Sotiropoulos et al., 2014). Indeed even imaging studies considering the underlying mechanisms fail to find definitive evidence for the encoding of a slow speed prior (Vintch & Gardner, 2014).

5 Conclusion

In this work, we have proposed and detailed a generative model for the estimation of the motion of dynamic images based on a formalization of small perturbations from the observer's point of view and parameterized by rotations, zooms, and translations. We connected these transformations to descriptions of ecologically motivated movements of both observers and the dynamic world. The fast synthesis of naturalistic textures optimized to probe motion perception was then demonstrated through fast GPU implementations applying autoregression techniques with much potential for future experimentation. This extends previous work from Sanz-Leon et al. (2012) by providing an axiomatic formulation. Finally, we used the stimuli in a psychophysical task and showed that these textures allow one to further understand the processes underlying speed estimation. We used broadband stimulation to study frequency-induced biases in visual perception, using various stimulus configurations, including octave bandwidth and RMS contrast-controlled manipulations, which allowed us to manipulate central frequencies as scale-invariant stimulus zooms. We showed that measured biases under these controlled conditions were qualitatively the same at both a faster and a slower tested speed. By linking the stimulation directly to the standard Bayesian formalism, we demonstrated that the sensory representation of the stimulus (the likelihoods) in such models is compatible with the generative MC model in over half of the collected empirical data cases. Together with a slow speed prior, the inference framework correctly accounts for the observed bias. We foresee that more experiments with naturalistic stimuli such as MCs and a consideration of more generally applicable priors will be needed in the future.

Acknowledgments

We thank Guillaume Masson for fruitful discussions during the project. We also acknowledge Manon Bouyé and Élise Amfreville for proofreading. Finally, we strongly thank the reviewers for their useful comments that helped us improve the readability and quality of this letter. L.U.P. was supported by EC FP7-269921, BrainScaleS, and BalaV1 ANR-13-BSV4-0014-02. The work of J.V. and G.P. was supported by the European Research Council (ERC project SIGMA-Vision). A.I.M. and L.U.P. were supported by SPEED ANR-13-SHS2-0006.

References

Abraham, B., Camps, O. I., & Sznaier, M. (2005). Dynamic texture with Fourier descriptors. In *Proceedings of the 4th International Workshop on Texture Analysis and Synthesis* (vol. 1, pp. 53–58). Edinburgh: Heriot-Watt University.

Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of Optical Society of America, A*, *2*(2), 284–99.

Born, R. T., & Bradley, D. C. (2005). Structure and function of visual area MT. *Annual Review of Neuroscience*, *28*(1), 157–189.

Brockwell, P., Davis, R., & Yang, Y. (2007). Continuous-time gaussian autoregression. *Statistica Sinica*, *17*(1), 63.

Brockwell, P. J., & Lindner, A. (2009). Existence and uniqueness of stationary Lévy-driven CARMA processes. *Stochastic Processes and Their Applications*, *119*(8), 2660–2681.

Brooks, K. R., Morris, T., & Thompson, P. (2011). Contrast and stimulus complexity moderate the relationship between spatial frequency and perceived speed: Implications for MT models of speed perception. *Journal of Vision*, *11*(14), 19.

Burge, J., & Geisler, W. S. (2015). Optimal speed estimation in natural image movies predicts human performance. *Nature Communications*, *6*, 7900.

Colombo, M., & Seriès, P. (2012). Bayes in the brain: On bayesian modelling in neuroscience. *British Journal for the Philosophy of Science*, *63*(3), 697–723.

Costantini, R., Sbaiz, L., & Süsstrunk, S. (2008). Higher order SVD analysis for dynamic texture synthesis. *IEEE Transactions on Image Processing*, *17*(1), 42–52.

DiCarlo, J. J. J., Zoccolan, D., & Rust, N. C. C. (2012). How does the brain solve visual object recognition? *Neuron*, *73*(3), 415–434.

Dong, D. (2010). Maximizing causal information of natural scenes in motion. In G. S. Masson, S. Guillaume, & U. J. Ilg (Eds.), *Dynamics of visual motion processing* (pp. 261–282). Berlin: Springer.

Doretto, G., Chiuso, A., Wu, Y. N., & Soatto, S. (2003). Dynamic textures. *International Journal of Computer Vision*, *51*(2), 91–109.

Doya, K. (2007). *Bayesian brain: Probabilistic approaches to neural coding*. Cambridge, MA: MIT Press.

El Karoui, N., Peng, S., & Quenez, M. C. (1997). Backward stochastic differential equations in finance. *Mathematical Finance*, *7*(1), 1–71.

Filip, J., Haindl, M., & Chetverikov, D. (2006). Fast synthesis of dynamic colour textures. In *Proceedings of the 18th International Conference on Pattern Recognition* (vol. 4, pp. 25–28). Piscataway, NJ: IEEE.

Fischer, S., Sroubek, F., Perrinet, L. U., Redondo, R., & Cristóbal, G. (2007). Self-invertible 2D log-Gabor wavelets. *International Journal of Computer Vision*, *75*(2), 231–246.

Fox, R. F. (1997). Stochastic versions of the Hodgkin-Huxley equations. *Biophysical Journal*, *72*(5), 2068–2074.

Galerne, B. (2011). *Stochastic image models and texture synthesis*. Unpublished doctoral diss., ENS de Cachan.

Galerne, B., Gousseau, Y., & Morel, J. M. (2011a). Random phase textures: Theory and synthesis. *IEEE Transactions on Image Processing*, *20*, 257–267.

Galerne, B., Gousseau, Y., & Morel, J. M. (2011b). Micro-Texture synthesis by phase randomization. *Image Processing On Line*, *1*.

Gekas, N., Meso, A. I., Masson, G. S., & Mamassian, P. (2017). A normalization mechanism for estimating visual motion across speeds and scales. *Current Biology*, *27*(10), 1514–1520.

Gel’fand, I. M., Vilenkin, N. Y., & Feinstein, A. (1964). *Generalized functions: Applications of harmonic analysis.* Providence, RI: American Mathematical Society.

Gregory, R. L. (1980). Perceptions as hypotheses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *290*(1038), 181–197.

Hassan, O., & Hammett, S. T. (2015). Perceptual biases are inconsistent with Bayesian encoding of speed in the human visual system. *Journal of Vision*, *15*(2), 1–9.

Hassan, O., Thompson, P., & Hammett, S. T. (2016). Perceived speed in peripheral vision can go up or down. *Journal of Vision*, *16*(6), 1–7.

Hyndman, M., Jepson, A. D., & Fleet, D. J. (2007). Higher-order autoregressive models for dynamic textures. In *Proceedings of the British Machine Vision Conference*. Surrey: BMVA Press.

Jogan, M., & Stocker, A. A. (2015). Signal integration in human visual speed perception. *Journal of Neuroscience*, *35*(25), 9381–9390.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Rev. Psychol.*, *55*, 271–304.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences*, *27*(12), 712–719.

Liu, C.-B., Lin, R.-S., Ahuja, N., & Yang, M.-H. (2006). Dynamic textures synthesis as nonlinear manifold learning and traversing. In *Proceedings of the British Machine Vision Conference*. Surrey: BMVA Press.

Meidan, R. (1980). On the connection between ordinary and generalized stochastic processes. *Journal of Mathematical Analysis and Applications*, *76*(1), 124–133.

Meso, A. I., & Simoncini, C. (2014). Towards an understanding of the roles of visual areas MT and MST in computing speed. *Frontiers in Computational Neuroscience*, *8*.

Meso, A. I., & Zanker, J. M. (2009). Speed encoding in correlation motion detectors as a consequence of spatial structure. *Biological Cybernetics*, *100*(5), 361–370.

Nestares, O., Fleet, D., & Heeger, D. (2000). Likelihood functions and confidence bounds for total-least-squares problems. In *Proceedings of the IEEE Conference on*

Computer Vision and Pattern Recognition (pp. 523–530). Hoboken, NJ: Wiley–IEEE Computer Society Press.

Perrone, J. A., & Thiele, A. (2001). Speed skills: Measuring the visual speed analyzing properties of primate MT neurons. *Nat. Neurosci.*, *4*(5), 526–532.

Priebe, N., Cassanello, C., & Lisberger, S. (2003). The neural representation of speed in macaque area MT/V5. *J. Neurosci.*, *23*, 5650–5661.

Rahman, A., & Murshed, M. (2008). Dynamic texture synthesis using motion distribution statistics. *Journal of Research and Practice in Information Technology*, *40*(2), 129.

Rousselet, G. A., Thorpe, S. J., & Fabre-Thorpe, M. (2004). How parallel is visual processing in the ventral pathway? *Trends in Cognitive Sciences*, *8*(8), 363–370.

Sanz-Leon, P., Vanzetta, I., Masson, G. S., & Perrinet, L. U. (2012). Motion clouds: Model-based stimulus synthesis of natural-like random textures for the study of motion perception. *Journal of Neurophysiology*, *107*(11), 3217–3226.

Simoncini, C., Perrinet, L. U., Montagnini, A., Mamassian, P., & Masson, G. S. (2012). More is not always better: Adaptive gain control explains dissociation between perception and action. *Nature Neurosci.*, *15*(11), 1596–1603.

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

Smith, M. A., Majaj, N., & Movshon, J. A. (2010). Dynamics of pattern motion computation. In G. S. Masson & U. J. Ilg (Eds.), *Dynamics of visual motion processing: Neuronal, behavioral and computational approaches* (pp. 55–72). Heidelberg: Springer.

Smith, P. L. (2000). Stochastic dynamic models of response time and accuracy: A foundational primer. *Journal of Mathematical Psychology*, *44*(3), 408–463.

Sotiropoulos, G., Seitz, A. R., & Seriès, P. (2014). Contrast dependency and prior expectations in human speed perception. *Vision Research*, *97*, 16–23.

Stocker, A. A., & Simoncelli, E. P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience*, *9*(4), 578–585.

Thompson, P. (1982). Perceived rate of movement depends on contrast. *Vision Research*, *22*(3), 377–380.

Thompson, P., Brooks, K., & Hammett, S. T. (2006). Speed can go up as well as down at low contrast: Implications for models of motion perception. *Vision Research*, *46*(6), 782–786.

Unser, M., & Tafti, P. (2014). *An introduction to sparse stochastic processes*. Cambridge: Cambridge University Press.

Unser, M., Tafti, P. D., Amini, A., & Kirshner, H. (2014). A unified formulation of gaussian versus sparse stochastic processes—Part II: Discrete-domain theory. *IEEE Transactions on Information Theory*, *60*(5), 3036–3051.

Unser, M., Tafti, P. D., & Sun, Q. (2014). A unified formulation of gaussian versus sparse stochastic processes—Part I: Continuous-domain theory. *IEEE Transactions on Information Theory*, *60*(3), 1945–1962.

Vacher, J., Meso, A. I., Perrinet, L. U., & Peyré, G. (2015). Biologically inspired dynamic textures for probing motion perception. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, & R. Garnett (Eds.), *Advances in neural information processing systems, 28* (pp. 1918–1926). Red Hook, NY: Curran.

Van, K., & Nicolaas, G. (1992). *Stochastic processes in physics and chemistry*. Amsterdam: Elsevier.

Vintch, B., & Gardner, J. L. (2014). Cortical correlates of human motion perception biases. *Journal of Neuroscience*, *34*(7), 2592–2604.

Wei, L. Y., Lefebvre, S., Kwatra, V., & Turk, G. (2009). State of the art in example-based texture synthesis. In *Eurographics 2009, State of the Art Report*. Eurographics Association.

Wei, X.-X., & Stocker, A. A. (2012). Efficient coding provides a direct link between prior and likelihood in perceptual Bayesian inference. F. Pereira, C. J. C. Burges, L. Bottou, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems*, *25* (pp. 1313–1321). Red Hook, NY: Curran.

Weiss, Y., & Fleet, D. J. (2001). Velocity likelihoods in biological and machine vision. In R. P. N. Rao, B. A. Olshausen, A. Bruno, & M. S. Lewicki (Eds.), *Probabilistic models of the brain: Perception and neural function* (pp. 81–100). Cambridge, MA: MIT Press.

Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience*, *5*(6), 598–604.

Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, & goodness of fit. *Attention, Perception, and Psychophysics*, *63*(8), 1293–1313.

Xia, G. S., Ferradans, S., Peyré, G., & Aujol, J. F. (2014). Synthesizing and mixing stationary gaussian texture models. *SIAM Journal on Imaging Sciences*, *7*(1), 476–508.

Young, R. A., & Lesperance, R. M. (2001). The gaussian derivative model for spatial-temporal vision: II. Cortical data. *Spatial Vision*, *14*(3), 321–390.

Yuan, L., Wen, F., Liu, C., & Shum, H.-Y. (2004). Synthesizing dynamic texture with closed-loop linear dynamic system. In *Proceedings of Computer Vision-ECCV* (pp. 603–616). Berlin: Springer.

Received July 25, 2017; accepted August 18, 2018.