

Dynamic updating of distributed neural representations using forward models

Eric L. Sauser · Aude G. Billard

Received: 19 September 2006 / Accepted: 21 October 2006
© Springer-Verlag 2006

Abstract In this paper, we present a continuous attractor network model that we hypothesize will give some suggestion of the mechanisms underlying several neural processes such as velocity tuning to visual stimulus, sensory discrimination, sensorimotor transformations, motor control, motor imagery, and imitation. All of these processes share the fundamental characteristic of having to deal with the dynamic integration of motor and sensory variables in order to achieve accurate sensory prediction and/or discrimination. Such principles have already been described in the literature by other high-level modeling studies (Decety and Sommerville in *Trends Cogn Sci* 7:527–533, 2003; Oztop et al. in *Neural Netw* 19(3):254–271, 2006; Wolpert et al. in *Philos Trans R Soc* 358:593–602, 2003). With respect to these studies, our work is more concerned with biologically plausible neural dynamics at a population level. Indeed, we show that a relatively simple extension of the classical neural field models can endow these networks with additional dynamic properties for updating their internal representation using external commands. Moreover, an analysis of the interactions between our model and external inputs also shows interesting properties, which we argue are relevant for a better understanding of the neural processes of the brain.

1 Introduction

For several decades, a large number of neural network architectures have been studied in order to understand the computational properties of the brain and its underlying neural mechanisms. A very influential series of models that have been proposed are known as continuous attractor neural networks or neural fields (Amari 1977; Wilson and Cowan 1973). Indeed, in addition to their biologically plausible structural relationship with real cortical neural ensembles, their numerous computational properties make them very attractive to the computational neuroscience community. For instance, these models were applied to research areas related to visual processing (Ben-Yishai et al. 1995; Giese 2000; Mineiro and Zipser 1998), visual attention (Rougier 2006), spatial navigation (Xie et al. 2002; Zhang 1996), decision making (Erlhagen and Schöner 2002; Sauser and Billard 2006; Schöner 2002), sensorimotor transformations (Burnod et al. 1999; Salinas and Thier 2000; Sauser and Billard 2005), stimulus binding (Wersing et al. 2001), and parameter estimation (Deneve et al. 1999). These networks are primarily based on a center-surround recurrent connectivity, where neurons sharing similar firing properties cooperate by exciting themselves and where neurons distant in their preferential tuning inhibit each other. This interneuron relationship is the basis of their fundamental ability to allow one single localized activity packet to emerge from this neural implementation of a winner-take-all operation.

The first point we would like to address here is concerned with the *dynamics of stimulus–network interactions* and, more precisely, with the temporal variation of a stimulus input in neural space. Indeed, apart from purely abstract theoretical works and a few applied to

E. L. Sauser (✉) · A. G. Billard
Learning Algorithms and Systems Laboratory (LASA),
Ecole Polytechnique Fédérale de Lausanne (EPFL),
1015 Lausanne, Switzerland
e-mail: eric.sauser@a3.epfl.ch

the modeling of biological systems (Giese 2000; Mineiro and Zipser 1998; Xie et al. 2002; Zhang 1996), the use of such networks in practical neurobiological modeling assumes quasistatic neural dynamics. By quasistatic we mean that the time scale of external input changes in neural space is much larger than that of the network. Since the dynamics of neural fields are, by definition, strongly influenced by the recurrent connectivity, the reaction time to a changing stimulus is thus higher than that of single neurons (Panzeri et al. 2001). Hence, in quasistatic network dynamics, the time scale of the input updating is set to a large value. When comparing these network implementations with real biological systems, this issue is not really a problem since their performance is considered in relatively slow tasks. However, when dealing, for example, with precise and fast movements like catching a ball or smooth eye pursuit, the internal representations should be updated very quickly, or even in advance, in order to predict accurately the outcome of self-generated movements (Miall and Wolpert 1996; Wolpert and Kawato 1998). An influential theory related to motor control suggests that the brain may use forward models in order to better control movements, arguing that, in humans, a closed-loop control system alone would be relatively inaccurate, given the long time needed for sending motor commands and receiving the resulting sensory feedback (Miall and Reckess 2002; Miall and Wolpert 1996; Wolpert and Kawato 1998). For instance, in an eye-tracking of a self-moved target experiment, it has been shown that, when the subjects actively move a target, the presence of the movement sensory feedback is not necessary to achieve almost zero latency, in contrast to the motor efference copies (Vercher et al. 1991). Furthermore, when considering neurophysiological data, predictive neural responses were found in the monkey visual, parietal, and frontal cortices and in the cerebellum (Nakamura and Colby 2002; Roitman et al. 2005; Unema and Goldberg 1997), highly suggesting the presence of forward control.

A second, more practical, motivation concerns the *neural mechanisms of self-awareness and recognition*. A current theoretical framework related to this research topic mainly considers two forms of cues that may be at the origin of such a human capacity (Decety and Sommerville 2003; Haggard and Clarke 2003; Jeannerod 2003). When one has to recognize one's own limb, body cues such as the spatial and visual attributes of the limb are primarily used. Since it is not within the scope of this paper to address the problem of self-limb recognition from these visual attributes, we are rather more interested in the second form of cues. Indeed, when the former attributes are ambiguous, it has been shown that one relies more on action or movement cues, e.g., the

time course of the movement velocity and its acceleration (Jeannerod 2003). A plausible mechanism has been suggested whereby an internal prediction of the consequences of a motor act is compared with the real sensory outcome. Then, depending on their similarity or discrepancy, the brain can determine the ownership of the observed movement (Decety and Sommerville 2003). In addition, another problem arises when considering the possible neural substrates concerned with the perception of self action and that of others. Indeed, the current body of evidence suggests that a common neural substrate devoted to the recognition and production of movements exists in both humans and monkeys (Iacoboni et al. 1999; Rizzolatti et al. 2001). A behavioral correlate of such a discovery, described in several psychophysics experiments (Chaminade et al. 2005; Kilner et al. 2003), is that observing movements of others influences the quality of one's own performance. The observation of such an interference effect, while supporting the view of a common pathway for the transfer of visuo-motor information, calls for an explanation as to how the same substrate can both integrate multisensory information and determine the ownership of the observed movement.

Despite the apparent differences between the two previously introduced topics, they both share a common fundamental property: their need for a *predictive forward model*, which would allow, respectively, (a) an almost instantaneous updating of the internal representation of the sensory states and (b) the computation of sensory predictions to be compared with movement outcome. Indeed, the timing of neural updating for internal sensory information is crucial in motor control for accurate movement generation (Miall and Reckess 2002; Vercher and Gauthier 1988). Similarly, a short time between movement execution and the perception of sensory feedback is also crucial for perceiving the agency of an action (Haggard and Clarke 2003). Our interest here is effectively to show *how a neural field*, a neural substrate for representing information, *can integrate the efferent commands from a forward model in order to update its internal representation*. As internal representations, we restrict our modeling by considering neural ensembles encoding simultaneously a variable value and its first-order time derivative. For instance, the position and the velocity of either a visual stimulus in retinal space or the hand location in cartesian or joint space may be considered. Indeed, real populations of neurons with such a tuning property have been found in several brain areas, such as the motor, parietal, visual, and temporal cortices and in the cerebellum (Ben Hamed et al. 2003; Hubel and Wiesel 1977; Jellema et al. 2004; Kettner et al. 1988; Roitman et al. 2005). The

modeling novelties that we introduce in this paper are threefold. First, we provide a generalized framework for the dynamic integration of velocity commands within continuous attractor networks by selectively adding an asymmetric recurrent connectivity on neural ensembles sharing similar movement direction preferences. Second, we also consider the need for a detailed stimulus encoding function that can compensate for the internal dynamics of both neurons and the network. Third, we describe and analyze properties and applications resulting from the proposed integration mechanism, such as (a) dynamic or input driven velocity tuning, (b) instantaneous and predictive-like information transfer, and (c) abilities for stimuli discrimination based on their dynamic properties.

This paper is organized as follows. Section 2 introduces the model core architecture and then describes the mechanisms underlying the integration of velocity commands. Next, the form of stimulus encoding and the synaptic projections for transferring information across neural populations are defined. Note that detailed mathematical calculations were left in the Appendix. Further, in Sect. 3, we show some analysis of the network dynamic properties with and without external stimulus inputs, such as velocity tuning. We also address the timing of information transfer across neural populations and then tackle the problem of stimulus discrimination. Moreover, we also compare our model dynamics with that of a more commonly used neural field architecture and then show its weaknesses. Finally, we discuss in Sect. 4 the relationship between our model properties and biological data and present some implications and predictions from our work regarding several of the brain neural mechanisms.

2 Model

We consider a continuous attractor neural network composed of a set of neurons preferentially tuned to a primary variable \mathbf{r} and a secondary variable \mathbf{s} following a uniform distribution such that $\mathbf{r} \in \mathcal{D}_{\mathbf{r}}$ and $\mathbf{s} \in \mathcal{D}_{\mathbf{s}}$. As mentioned previously, these neural spaces are assumed to encode, respectively, a variable value and its variation in time. For example, if we consider that a stimulus location is encoded in the space $\mathcal{D}_{\mathbf{r}}$, a neuron tuned to a specific \mathbf{s} will fire preferentially when the stimulus is moving in the direction given by that \mathbf{s} . The present hypothesis regarding a combined preferential tuning to both a variable value and its direction of variation is mainly motivated by several neurophysiological studies showing neurons having similar firing properties (Fu et al. 1997; Jellema et al. 2004; Kettner et al. 1988; Roit-

Table 1 Definition domains of neural preferential tuning considered in this paper

	Ring	Torus
$\mathcal{D}_{\mathbf{r}}$	$\{\mathbb{R}^2 \mid \ \mathbf{r}\ = 1\}$	$[-\pi, \pi] \times [-\pi, \pi]$
$\mathcal{D}_{\mathbf{s}}$	$\{-1, 1\}$	$\{\mathbb{R}^2 \mid \ \mathbf{s}\ = 1\}$

man et al. 2005). In addition to the computational power that such a neural representation may provide to other connected brain networks, we argue and will show that it may also help in updating itself.

Further, the cases of both a unidimensional and a two-dimensional (2D) attractor are addressed (Table 1). To avoid boundary effects, periodic spaces are assumed such that these domains form, respectively, a ring and its 2D analog, a torus. Despite the discrete nature of \mathbf{s} in the case of the ring attractor, the neural ensemble follows continuous attractor network dynamics where the time evolution of the neurons' membrane potential $u(\mathbf{r}, \mathbf{s}, t)$ satisfies

$$\begin{aligned} \tau \dot{u}(\mathbf{r}, \mathbf{s}, t) = & -u(\mathbf{r}, \mathbf{s}, t) + h(\mathbf{s}, t) + x(\mathbf{r}, \mathbf{s}, t) \\ & + \iint [W(\mathbf{r} - \mathbf{r}') - \lambda \nabla W(\mathbf{r} - \mathbf{r}') \cdot \mathbf{s}'] \\ & \times f(u(\mathbf{r}', \mathbf{s}', t)) d\mathbf{r}' d\mathbf{s}'. \end{aligned} \tag{1}$$

The transfer function $f(u)$ is the linear threshold function $\max(0, u)$, and $\tau \in \mathbb{R}_+^*$ is the time constant of the neurons. The network receives external inputs that were separated into two distinct forms. $x(\mathbf{r}, \mathbf{s}, t)$ is defined as the *stimulus input*, whereas $h(\mathbf{s}, t)$ is the *background input*. Note that $h(\mathbf{s}, t)$ is, by definition, homogeneous across the subpopulations of neurons sharing the same preferential tuning to the variable \mathbf{s} . Henceforth, these subpopulations will be designated as sublayers.

The network is fully connected by means of a set of synaptic weights composed of two parts: a center-surround Gaussian-like,¹ translation-invariant, and symmetric component $W(\mathbf{r} - \mathbf{r}')$ and an asymmetric term $-\lambda \nabla W(\mathbf{r} - \mathbf{r}') \cdot \mathbf{s}'$, where ∇ corresponds to the gradient operator along \mathbf{r} . $\lambda \in \mathbb{R}_+^*$ is a constant scaling factor. The first term is a convolution kernel that links the neurons along the neural fundamental variable \mathbf{r} , while the second term links them along \mathbf{s} . These convolution weights and the network architecture are illustrated in Figs. 1 and 2. As will be described in Sect. 2.1, this connectivity is the basis of the system's ability to integrate velocity commands.

¹ The exact shape of W is not crucial for our argument, as long as it allows the network to sustain an activity packet on its neural surface (Amari 1977). Nevertheless, in Appendix C, we define the exact shape of W that we will use further in our experiments.

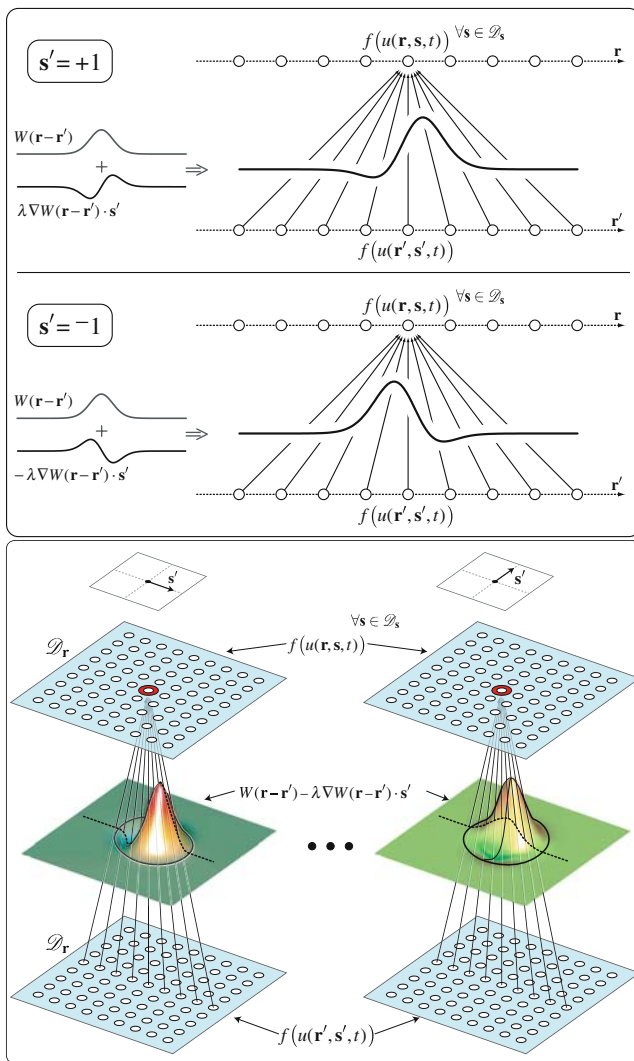


Fig. 1 Illustration of model architecture and weight kernels for both the ring (top) and the torus attractor spaces (bottom). Each sublayer encoding a variable in the neural space defined on \mathbf{r} has a preferred movement direction \mathbf{s} that is the result of the asymmetric self-connectivity. This weight kernel is shown for different values of \mathbf{s} . It is superimposed on the arrows denoting the synaptic projections across the sublayers

As illustrated in Fig. 2, this type of neural dynamics are known to form an activity packet or attractor bump on the surface of the neural field. It is suggested that this class of networks conveys information by such a compact blob of excitation. A typical read-out mechanism, the *population vector* $\mathbf{p}(t)$, has been suggested to measure the macroscopic effect of the joint activities of large sets of neurons (Georgopoulos 1996). It consists of a weighted summation of the firing activity of each neuron with its preferred direction. The estimate $\mathbf{p}(t) \in \mathcal{D}_r$ of the variable value encoded along \mathbf{r} is given by

$$\mathbf{p}(t) = \frac{\iint f(u(\mathbf{r}, \mathbf{s}, t)) \mathbf{r} \, d\mathbf{r} \, d\mathbf{s}}{\iint f(u(\mathbf{r}, \mathbf{s}, t)) \, d\mathbf{r} \, d\mathbf{s}} \quad (2)$$

An estimate of the encoded variable in \mathcal{D}_s might be envisaged similarly. However, since that space is an extension of the primary space \mathcal{D}_r used for integration purposes, its explicit definition is not necessary. The next section addresses the use of that dimension for updating the network neural representation using external commands provided through the background input.

2.1 Forward model commands and intrinsic dynamics

In this section, we are interested in how forward model commands, by acting on the background input, may drive the network internal dynamics so that the neural representation may be updated accordingly. Afterwards, we also determine the boundaries in which this velocity integration remains valid. Let us start by defining how the population vector $\mathbf{p}(t)$ encoded by the neural representation may be varied by an external command $\mathbf{v}^*(t)$ such that

$$\dot{\mathbf{p}}(t) = \mathbf{v}^*(t). \quad (3)$$

Furthermore, we consider here that the network is already representing and sustaining information and that the stimulus input is absent, i.e., $x(\mathbf{r}, \mathbf{s}, t) = 0$. The background input $h(\mathbf{s}, t)$ is then defined as

$$h(\mathbf{s}, t) = h_0 [1 + \hat{\mathbf{h}}_0(t) \cdot \mathbf{s}], \quad (4)$$

where $h_0 > 0$ is a constant excitation and $\hat{\mathbf{h}}_0$ is a directional input component that can break the network symmetry when different from zero (Fig. 3a). Indeed, when referring to Eq. (1), it can be noticed that a strictly homogeneous input ($\hat{\mathbf{h}}_0 = 0$) leads to a zero value of the integral along \mathbf{s} of the asymmetric weights convolution, i.e.,

$$\iint \lambda \nabla W(\mathbf{r} - \mathbf{r}') \cdot \mathbf{s}' f(u(\mathbf{r}', \mathbf{s}', t)) \, d\mathbf{r}' \, d\mathbf{s}' = 0. \quad (5)$$

By symmetry, the system thus settles into a constant state along \mathbf{s} . In this case, this network is equivalent to a typical continuous attractor network without an asymmetric connectivity, which is known to exhibit marginally stable bump solutions.

Further, an input $\hat{\mathbf{h}}_0 \neq 0$ breaks this symmetry. It consequently favors a higher excitation of the network sublayers with a direction preference \mathbf{s} close to that given by $\hat{\mathbf{h}}_0$. Now, if we first consider a single sublayer, it has been shown that, as an attractor state, it develops a traveling activity bump symmetric in shape and with constant velocity. That speed depends strictly on the ratio between the strength of the asymmetric weights and the neuron's time constant and is given by λ/τ (Zhang 1996). Coming back, then, to the full architecture, the coupling across the sublayers introduces a competitive

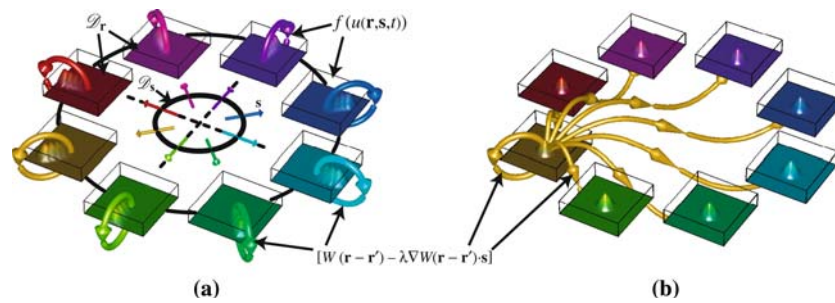


Fig. 2 Illustration of model architecture. Each sublayer encoding a variable in the neural space defined on $\mathbf{r} \in \mathcal{D}_r$ has a preferred movement direction $\mathbf{s} \in \mathcal{D}_s$ that is the result of the asymmetric self-connectivity. **a** The center arrows indicate the preferred movement direction of the sublayers. If considered alone, each

of them would display a traveling activity blob in its preferred movement direction, as shown by the trace of the activity blobs on the neural surface. **b** Each sublayer projects its activity onto all sublayers using the same weight profile

interaction between those having an opposite direction preference. This produces a push–pull mechanism regulated by the respective balance of excitation among the sublayers. In our model, this effect is controlled by the strength of the asymmetry of the background input $\hat{\mathbf{h}}_0$. Therefore, by balancing this input factor, one may drive the network representation $\mathbf{p}(t)$ in direction and velocity. The detailed calculations given in Appendix A show that near the equilibrium, the relationship between $\mathbf{p}(t)$ and $\hat{\mathbf{h}}_0(t)$ can be approximated linearly such that

$$\dot{\mathbf{p}}(t) \approx \frac{\lambda}{\tau} \gamma \hat{\mathbf{h}}_0(t), \tag{6}$$

where $\hat{\mathbf{h}}_0(t)$ is small. γ is the slope of the linear approximation that depends on the recurrent weight parameters.² Since this updating of the neural representation is performed by modulating the respective influence of opposite sublayers, the maximal velocity that our model may attain is constrained by that of a single sublayer, i.e., λ/τ . Thus, the network velocity is bounded such that

$$\dot{\mathbf{p}}(t) \approx \frac{\lambda}{\tau} \min \left(\gamma \hat{\mathbf{h}}_0(t), \frac{\hat{\mathbf{h}}_0(t)}{\|\hat{\mathbf{h}}_0(t)\|} \right). \tag{7}$$

Then, rewriting Eq. (4) using Eqs. (3) and (7) gives

$$h(\mathbf{s}, t) = h_0 \left[1 + \frac{\tau}{\lambda \gamma} \mathbf{v}^*(t) \cdot \mathbf{s} \right] \quad \|\mathbf{v}^*(t)\| \ll \frac{\lambda}{\tau}, \tag{8}$$

which corresponds to the background homogeneous input to be applied to the network so that it can drive its internal representation with speed $\mathbf{v}^*(t)$. Like a related model (Xie et al. 2002), this integration mechanism makes it possible for forward commands to allow the

network to predict its next state $\mathbf{p}(t)$ according to its supposed variation in time. As a consequence, the static marginal attractor states of the model become linear trajectories with constant velocity in the neural space defined on \mathcal{D}_r . Therefore, to contrast the effect of a stimulus input on the network, our network is said to possess an *intrinsic velocity* $\mathbf{v}^*(t) < \lambda/\tau$ determined by its background input, which is defined as the velocity of the trajectories of its attractors.

2.2 Stimulus encoding

In this section we describe how a stimulus input should be defined in order to drive the network dynamics toward that of the input. A nonlinear form is first defined, since its derivation from the background input is the most straightforward. We will then slightly modify this description by approximating it with a linear form more suitable for transferring information across multiple neural populations. Indeed, this linear input form will further help us derive appropriate synaptic projections. As a consequence, within a large network, instances of our model may easily transfer their positional and velocity-dependent information across each other. So, let us introduce the first form of stimulus input.

We consider a stimulus located at $\mathbf{r}_0(t)$ in the neural space \mathcal{D}_r moving in phase with the network intrinsic velocity, i.e., $\dot{\mathbf{r}}_0(t) = \mathbf{v}^*(t)$. Its input to the network $x(\mathbf{r}, \mathbf{s}, t)$, shown in Fig. 3b, is given by

$$x(\mathbf{r}, \mathbf{s}, t) = h_1 \left[G(\mathbf{r} - \mathbf{r}_0(t)) - \tau \dot{\mathbf{r}}_0(t) \cdot \nabla G(\mathbf{r} - \mathbf{r}_0(t)) \right] \left[1 + \frac{\tau}{\lambda \gamma} \dot{\mathbf{r}}_0(t) \cdot \mathbf{s} \right] \quad \|\dot{\mathbf{r}}_0(t)\| \ll \frac{\lambda}{\tau}, \tag{9}$$

² Except for rare special cases, γ must be found numerically.

where $h_1 > 0$ corresponds to the stimulus amplitude and G to the fundamental shape of the input. We assume this shape to be Gaussian-like and centered on the stimulus location $\mathbf{r}_0(t)$ (see a precise definition in Appendix C). The second term of the first factor compensates for the neuron’s integration time so that the network representation of the input shape is unaffected by stimulus movement. It stays symmetric, as if it was static in the neural medium. Finally, comparing the equation of the stimulus input with that of the background input [Eqs. (9) and (8)], it can be noticed that the asymmetric term in the second factor is responsible for driving the push–pull mechanism of the neural field. Through this modulatory excitation, the asymmetric recurrent connectivity will help the neural field tracking the stimulus. Since the asymmetry of the stimulus input is driven by its speed, we introduce here a directional input factor $\hat{\mathbf{h}}_1$ that has a role similar to that of the background input. It is given by

$$\hat{\mathbf{h}}_1(t) = \frac{\tau}{\lambda\gamma} \dot{\mathbf{r}}_0(t). \tag{10}$$

This relationship relates to the amount of strength of the input asymmetry that is necessary for the network to follow its speed. Rewriting Eq. (9) then gives

$$x(\mathbf{r}, \mathbf{s}, t) = h_1 \left[G(\mathbf{r} - \mathbf{r}_0(t)) - \lambda\gamma \hat{\mathbf{h}}_1(t) \cdot \nabla G(\mathbf{r} - \mathbf{r}_0(t)) \right] \left[1 + \hat{\mathbf{h}}_1(t) \cdot \mathbf{s} \right]. \tag{11}$$

Since the asymmetric factors of both the stimulus and the background input are proportional, respectively, to the stimulus and intrinsic network velocity, the use of either notation will be considered equivalent in the further analysis of our experiments. Again, the reader interested in a more detailed description of this mathematical development is encouraged to refer to Appendix B.1.

2.2.1 Linear encoding

As mentioned earlier, an alternative stimulus input form can be defined so as to avoid the nonlinear multiplicative factor found in its previous definition [Eq. (9)]. This multiplicative factor is needed to drive the internal network dynamics toward that of the input. In order to replace it, the idea is to apply the same principle as that described in the case of the background input. A constant term, homogeneous along the stimulus space \mathcal{D}_r and with a velocity-dependent asymmetry along \mathbf{s} , can be used. Detailed calculations given in Appendix B.2

result in the following input form:

$$x(\mathbf{r}, \mathbf{s}, t) = h_1 \left[G(\mathbf{r} - \mathbf{r}_0(t)) - \tau \dot{\mathbf{r}}_0(t) \cdot \nabla G(\mathbf{r} - \mathbf{r}_0(t)) + \eta \hat{\mathbf{h}}_1 \cdot \mathbf{s} \right], \tag{12}$$

where η is a constant that depends on both the network recurrent weights W and on the input shape G . This form of linear encoding is particularly interesting for transmitting information across populations. Indeed, as will be described next, our model may transfer both its positional and velocity-related information to another field by means of strictly linear synaptic projections chosen appropriately.

2.3 Information transmission across neural fields

In large-scale neural networks, it may be envisaged that several instances of our model may communicate and transfer information. The synaptic projections W_{AB} from one neural field A to another B are defined so that the input $x_B(\mathbf{r}, \mathbf{s}, t)$ of population B is given by

$$x_B(\mathbf{r}, \mathbf{s}, t) = \iint W_{AB}(\mathbf{r}, \mathbf{s}, \mathbf{r}', \mathbf{s}') f(u_A(\mathbf{r}', \mathbf{s}', t)) d\mathbf{r}' d\mathbf{s}'. \tag{13}$$

In this section, our aim is to allow these synaptic projections to be equivalent to the stimulus input form given by Eq. 12. Consequently, the information conveyed by the source population A would be completely transferred such that $\mathbf{p}_B(t) = \mathbf{p}_A(t)$ and $\dot{\mathbf{p}}_B(t) = \dot{\mathbf{p}}_A(t)$. Usually, in classical neural field implementation, information is transmitted using a weight kernel having a symmetric and center-surround shape. However, as shown previously, the ability for velocity integration is the result of an asymmetric coupling across the network sublayers. Thus, we here use a similar technique to that described in the context of whole network dynamics [Eq. (1)]. It can be shown (Appendix B.3) that the following weights

$$W_{AB}(\mathbf{r}, \mathbf{s}, \mathbf{r}', \mathbf{s}') = \left[W_T(\mathbf{r} - \mathbf{r}') - \lambda \nabla W_T(\mathbf{r} - \mathbf{r}') \cdot \mathbf{s}' + \mu_{AB} \mathbf{s} \cdot \mathbf{s}' \right] \tag{14}$$

produce the desired effect. W_T is a symmetric and center-surround convolution kernel strictly defined on the neural space \mathcal{D}_r . The last term, $\mu_{AB} \mathbf{s} \cdot \mathbf{s}'$, where μ_{AB} is a constant depending on the recurrent weights and on the input shape, is the analog to that of Eq. (12). The input source may hence drive the network according to its own dynamics.

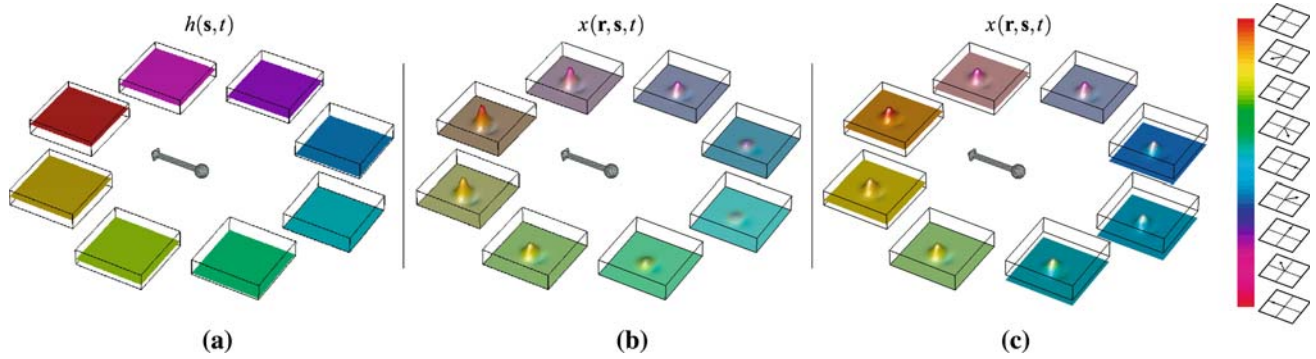


Fig. 3 Examples of external inputs where the same representation as in Fig. 2 is used. The *color code* on the *right* indicates the corresponding preferred movement direction of each sublayer. **a** Background input $h(s, t)$ [Eq. (4)]. **b** stimulus input $x(r, s, t)$

[Eq. (9)]. **c** Linear form of stimulus input [Eq. (12)]. The *arrow* in the *middle* of the figures denotes the intrinsic velocity of the network resulting when the corresponding input is applied

3 Experiments and results

In this section, we present several experiments that were performed using our model. Sections 3.1 and 3.2 show numerical simulations aimed at illustrating the relevance of our mathematical developments. Our model’s ability to integrate velocity commands and its response to both types of stimulus inputs and synaptic projections across multiple instances of our networks are described. We further investigate, in Sect. 3.3, how our model may reproduce neurophysiological data concerning the preferential tuning to stimulus velocity of several groups of neurons in the visual cortex (Orban et al. 1986). Finally, we report two additional simulation results addressing mechanisms that we suggest to be respectively related to motor imagery and sensory discrimination. The former experiment describes how the excitation level of the network may liberate it from its sensory influences. The latter considers a neural population that receives two contradictory and ambiguous sensory feedback inputs and that should select the input corresponding best to its own internal dynamics.

3.1 Dynamic velocity integration

In the hippocampal formation of the rat, it has been shown that the so-called head-direction cells code for the current heading direction of the animal. It has also been observed that this group of neurons displays a fast updating of its representation according to motor efference copies or vestibular inputs (Sharp et al. 2001). Moreover, in other brain regions, neural representations encoding for different sensory states have also been shown to exhibit a similar updating (Roitman et al. 2005; Schwartz and Moran 1999).

Here, we show that our model may also update its internal representation according to the commands $\mathbf{v}^*(t)$ given through the background input $h(s, t)$. The simulation parameters that were used in each of the following experiments are summarized in Appendix C. Figure 4 shows the measured velocity response $\dot{\mathbf{p}}(t)$ of the neural field as a function of $\|\hat{\mathbf{h}}_0\|$. Each point on the graph was obtained in a different trial with a different value for $\|\hat{\mathbf{h}}_0\|$. The approximation, given by Eq. (7), is displayed in the same graph. Note that, in the ring attractor case, for a sufficiently strong asymmetric input, the network reaches the maximum velocity λ/τ . However, in the torus case, it can be seen that the system only tends toward that maximum asymptotically. Indeed, in the former case, the stronger sublayer is capable of completely inhibiting its opposite and hence driving the whole network alone. The continuous nature of the direction preference in the second case forbids a single sublayer to win against all the others. Nevertheless, for a relatively small asymmetric component $\hat{\mathbf{h}}_0$, the theoretical approximation fits the simulation results well. In Figs. 5 and 6, some trajectories of the command \mathbf{v}^* are compared to the velocity response $\dot{\mathbf{p}}(t)$ of the network.

3.2 Information transfer

The purpose of this section is to show how our network can integrate the representation given by a stimulus input and by projections from another network. It was mentioned previously that a population of neurons may update its encoded variable through the integration of velocity commands. Similarly, a stimulus input should also be able to do so. In what follows, our model will be shown to capture such an effect. In addition, we will also compare its performance with that of a more commonly used single-layered attractor network. Recall that the

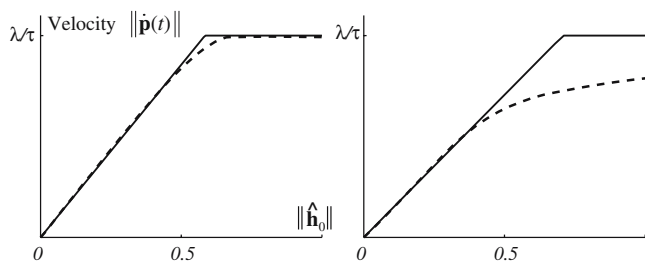


Fig. 4 Network velocity response (dotted line) to different asymmetric background input drives $\hat{\mathbf{h}}_0$, as displayed by a ring (left) and torus attractor (right). The straight line corresponds to the approximation of the velocity response [Eq. (4)]

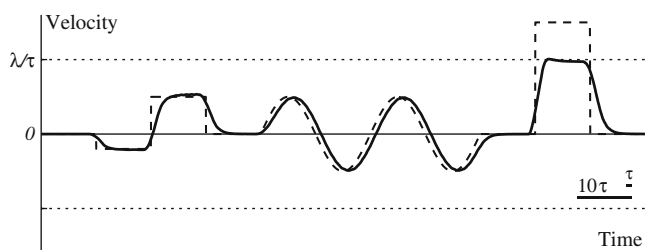


Fig. 5 Velocity response $\hat{\mathbf{p}}(t)$ of a ring network to external commands $\mathbf{v}^*(t)$ provided through background input [Eq. (8)]. Dotted and straight lines correspond, respectively, to the desired velocity command $\mathbf{v}^*(t)$ and the network velocity response $\hat{\mathbf{p}}(t)$. As can be seen when a high-velocity command is given, the network saturates at its maximum velocity integration boundary given by λ/τ

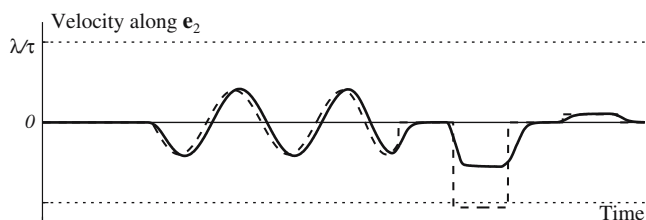
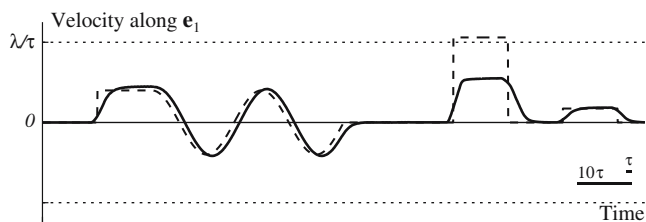


Fig. 6 Velocity response $\hat{\mathbf{p}}(t)$ of a torus network to external commands $\mathbf{v}^*(t)$. Data are shown relative to the two principal axes of the network representation space \mathcal{R}_4 given by the canonical base $\{\mathbf{e}_1 = (1, 0), \mathbf{e}_2 = (0, 1)\}$. The same notation as in Fig. 5 is used

latter neural dynamics are given by

$$\tau \dot{u}(\mathbf{r}, t) = -u(\mathbf{r}, t) + x(\mathbf{r}, t) + h(t) + \oint W(\mathbf{r} - \mathbf{r}') f(u(\mathbf{r}', t)) d\mathbf{r}' \quad (15)$$

When compared to Eq. (1), this corresponds to a network with a reduced dimensionality. In this case, the stimulus and the background inputs are given, respectively, by $x(\mathbf{r}, t) = h_1 G(\mathbf{r} - \mathbf{r}_0(t))$ and $h(t) = h_0$.

3.2.1 Transfer from stimulus to network representation

To illustrate the validity of our mathematical development concerning stimulus integration, we applied separately to the network the nonlinear and the linear forms of the stimulus input [Eqs. (9) and (12), respectively]. The background input amplitude h_0 was set to zero. For different stimulus speeds, we measured the spatial lag between the effective stimulus spatial location $\mathbf{r}_0(t_n)$ and the neural population vector $\mathbf{p}(t_n)$ at a given time t_n . The results are plotted in Fig. 7a and b. These figures show that, under both conditions, our model outperforms the simpler model. Indeed, below the maximum integration velocity given by λ/τ , the lag stays close to zero, while an almost linear velocity-dependent lag is observed for the other model.

3.2.2 Transfer across network representation

Next, we performed similar simulations while considering the transfer of information across the representations of two interconnected neural fields using Eqs. (13) and (14). The projection weights kernel W_τ is defined in Appendix C. Figure 7c shows the resulting lag measured across the neural populations. As expected, the lag effectively stays close to zero below the velocity integration boundary λ/τ , which corresponds to a property similar to that mentioned in the case of a stimulus input.

3.3 Dynamic velocity tuning

In the context of the visual system, neurons have been found to be preferentially tuned to both a stimulus position and velocity in retinal space. This property has further been suggested to be the basis of the human ability for velocity discrimination (Cheng et al. 1994; Chey et al. 1998; Goodwin and Henry 1975; Mineiro and Zipser 1998). Our interest is to see if our model may first reproduce this experimental result and further suggest an alternative approach to this neural process. Indeed, it has already been shown that a single-layered attractor network having a fixed asymmetric recurrent connectivity can exhibit a certain tuning to input stimulus speed (Mineiro and Zipser 1998). However, such a model is restricted to be tuned to a single preferred velocity. It thus needs to be replicated a sufficiently large number of times with different synaptic strengths so that

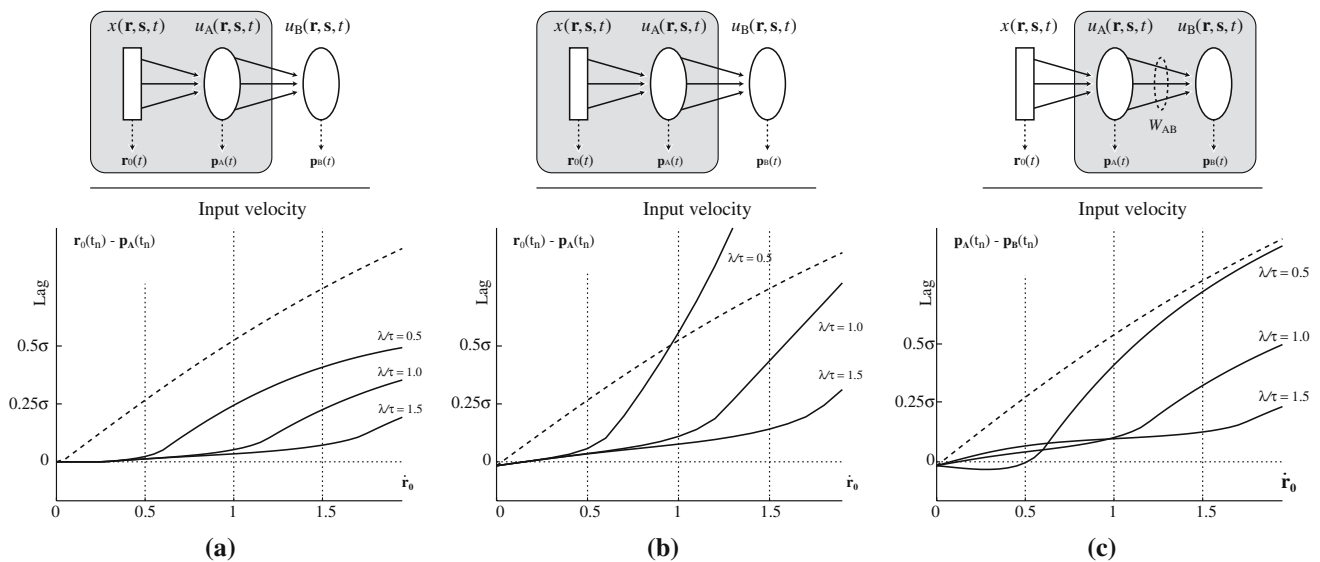


Fig. 7 *Bottom* Positional lag at a given time t_n between the network representation $\mathbf{p}(t_n)$ and **a** a moving stimulus encoded using the nonlinear method (Eq. 9), **b** a moving stimulus encoded with the linear method (Eq. 12), and **c** another network driven by a moving stimulus (Eq. 13). To quantify the amount of latency between these representations, the lag is given relative to the breadth σ of the input shape G (Appendix C). *Top* System schematic shows which part of the network is considered in each

case. Data are shown for different values of the ratio λ/τ , which corresponds to the maximum allowable integration velocity. Note that the lag is very low below this maximal value but then grows almost linearly with the input speed. Moreover, this figure also compares our neural field model with classical continuous attractor network implementation (*dotted line*). Note that such a model suffers from a lag that is almost linearly related to the input speed

the whole system will then exhibit a preferential tuning to a broad range of velocities (Chey et al. 1998).

By considering the changes in the network response resulting from the modulation of the stimulus speed, we show that our model can display this large range of preferential tuning by simply varying the asymmetry of the background input $h(\mathbf{s}, t)$. Indeed, by dynamically setting the intrinsic network velocity \mathbf{v}^* , a stimulus with a close speed will resonate or cooperate more with the network than divergent ones. Since cooperative interactions in such an attractor network result in higher activation patterns than competitive ones, we assume that the mean global firing rate of all the neurons is a good measure of this resonance effect. The network response energy $E(t)$ is defined by

$$E(t) = \iint f(u(\mathbf{r}, \mathbf{s}, t)) d\mathbf{r} d\mathbf{s}. \tag{16}$$

Several simulation trials were performed while the model was given various intrinsic velocities \mathbf{v}^* . Then, a single stimulus moving according to a large range of speeds and directions was applied. The measured network energy for each trial is shown in Fig. 8. It can be seen that the network effectively responds preferentially to stimuli with a close velocity and that these tuning curves show a high similarity to those reported in the visual cortex (Cheng et al. 1994; Orban et al. 1986).

3.4 Stimulus to background input strength ratio

The experiments described above were mostly performed while the mean background homogeneous input h_0 was kept sufficiently small so that the influence of the network recurrent connectivity was relatively weak compared to the strength of the stimulus input h_1 . Indeed, the ratio h_0/h_1 is of critical importance when considering the network dynamic behavior. Indeed, it determines which of these inputs is driving the network. A small value corresponds to a predominance of the stimulus input, while a larger one corresponds to a dominance of the network intrinsic dynamics.

This network property may allow the same neural substrate to be used for different purposes. For instance, several motor areas of the brain have been shown to be activated similarly by movement execution, imagery, and observation (Fogassi and Gallese 2002; Jeannerod and Decety 1995; Porro et al. 1996; Rizzolatti et al. 2001), processes that are not completely equivalent. During motor execution, it is necessary for the brain to keep track of the real sensory states, whereas during the other processes, it should avoid perceiving them. Considering now an analogy for our model, in the former case ($h_1 \gg h_0$), the network stays locked to the input stimulus. In the latter case ($h_0 \gg h_1$), the intrinsic dynamics become sufficiently strong to free itself from the stimulus

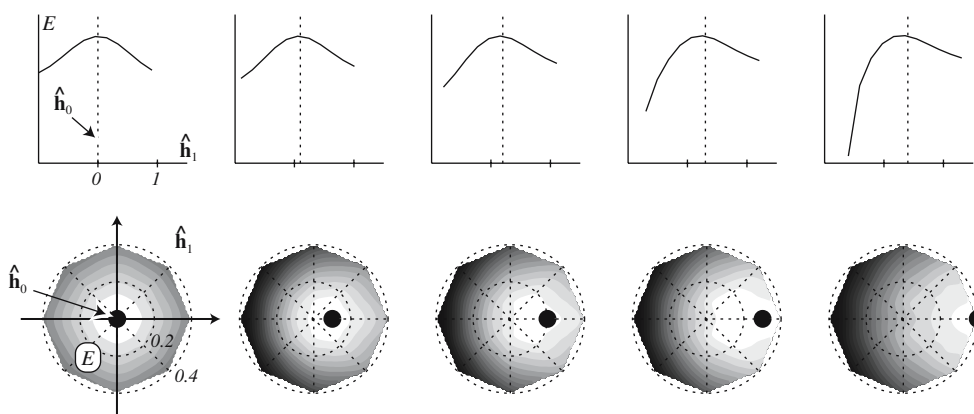


Fig. 8 *Top* Ring network mean energy relative to the speed of the input stimulus. Each *plot* corresponds to a different intrinsic network speed. Notice the shift in the preferential velocity tuning

of the network with its intrinsic speed. *Bottom* The same results for a torus attractor. The *black dot* indicates the network intrinsic speed. The brighter the plot is, the higher the energy

input. Hence, the network evolves strictly according to its background input. Figure 9 shows, for different values of the ratio h_0/h_1 , the velocity space in which the network is mostly driven by either the stimulus or the background input. By mostly driven by the stimulus input we mean that the network response $\mathbf{p}(t)$ is strongly dependent on the stimulus location $\mathbf{r}_0(t)$. Indeed, if the lag between these values is almost constant through time, it means that the network actually follows that input. In contrast, an uncorrelated difference between $\mathbf{p}(t)$ and $\mathbf{r}_0(t)$ indicates that the network is only driven by its background input. The data shown in Fig. 9 was obtained by first letting the network and the stimulus evolve for a given and sufficiently long period of time. Then, we measured the lag between the stimulus location and that of the network representation. If, for a given trial, the lag was less than twice the breadth of the stimulus input σ (Appendix C), the network was considered to be stimulus driven.

3.5 Sensory discrimination

The last simulation concerns sensory discrimination. Our inspiration was taken from a behavioral experiment addressing the recognition of a stimulus moving according to self-generated movements in the presence of one ambiguous distractor. Indeed, the neural mechanisms related to self-recognition are supposed to be grounded in the brain by means of a forward model taking as inputs motor efference copies whose predictions are compared with actual sensory consequences (Decety and Sommerville 2003). A close match between the prediction and the sensory feedback is supposed to signify that the observed stimulus is controlled by the self, whereas a discrepancy would mean that it is

under an external influence. Neurophysiological data suggest that this discrimination process may be partly grounded within brain regions containing shared representations (Decety and Sommerville 2003). This would signify that a neural population receiving two ambiguous inputs should be capable of performing such a selection.

Therefore, we applied this principle to our model. Two external stimulus inputs located, respectively, at $\mathbf{r}_0^c(t)$ and $\mathbf{r}_0^i(t)$ in neural space are fed to our network. Since they are supposed to be ambiguous, their corresponding amplitude is equivalent, i.e., $h_1^c = h_1^i$, and their initial location in neural space is assumed to be identical. Respectively designated by the indices *c* and *i*, the *compatible* stimulus has a speed that is equally fed to the network through its background input, i.e., $\hat{h}_1^c = \hat{h}_0$ such that $\mathbf{r}_0^c(t) \approx \mathbf{v}^*(t)$, while the *incompatible* stimulus is different. Simulation results are shown in Figs. 10 and 11. Figure 10 illustrates the network neural activity at different time steps. It can be seen that when the stimulus inputs are separated in neural space, the network selects the input with a speed corresponding best to its own intrinsic dynamics. Figure 11 shows the temporal dynamics of stimulus selection for different cases. Below the maximum intrinsic network speed (Fig. 11a, b), the selection is successful, but above, the network may either lag behind the compatible stimulus or even select the incompatible stimulus (Fig. 11c, d). Finally, Fig. 12 summarizes the range of speeds where the network successfully discriminates the right stimulus. In the case of a right decision, the spatial lag between the population vector $\mathbf{p}(t)$ and the compatible stimulus location \mathbf{r}_0^c is shown. Note that the lag is almost zero when the compatible stimulus velocity lies within the allowable range of velocity integration. Indeed, as described in Sect. 3.2.1, outside that range, the lag increases with

Fig. 9 Input space in which the network is mostly driven by the stimulus or the background input is shown for different value of the ratio h_0/h_1 . The *straight lines* indicate the separations of the input velocity space between areas where the network is mostly driven by the stimulus input (*straight arrows*) and where the background input dominates (*dashed arrows*)

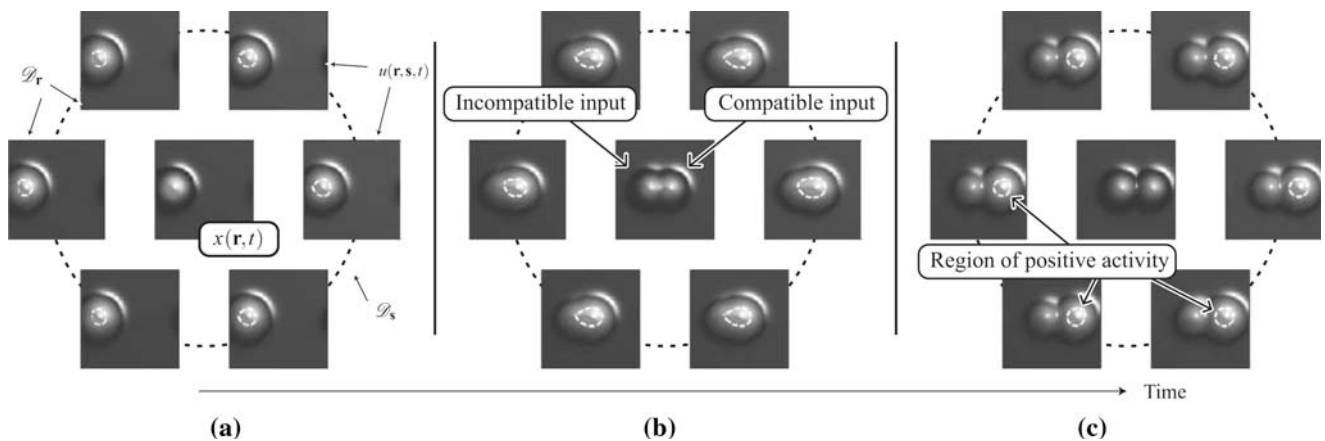
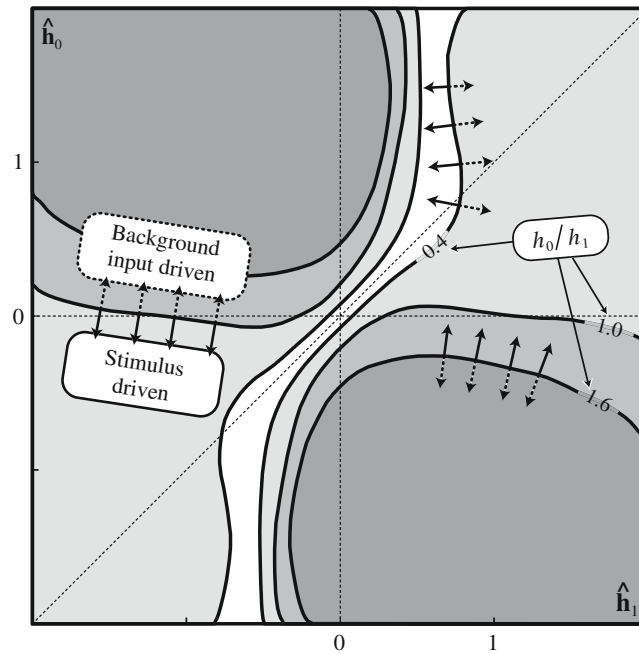


Fig. 10 Network response during a velocity discrimination task. Each *subplot* corresponds to a snapshot of the membrane potential $u(\mathbf{r}, \mathbf{s}, t)$ of selected sublayers sorted according to their preferred movement direction. The surfaces enclosed with the *white dotted line* indicate regions of neural space in which activity is above zero, i.e., the neurons within these areas are actually firing. In the *middle* of each *subfigure*, the external input $x(\mathbf{r}, \mathbf{s}, t)$ averaged over the preferred directions of movement \mathbf{s} are shown. **a** Beginning

of discrimination task: both the compatible and the incompatible stimuli are at the same location in neural space; they are indistinguishable. **b** The stimuli, moving at different speeds, start to separate but are not distinguishable on the network representation yet. **c** The stimuli are clearly disjoint, and, by means of its recurrent interactions, the network naturally selects the stimulus that is the most compatible with its own intrinsic velocity

the stimulus speed. Moreover, the regions of false discrimination are strictly located over the system integration boundary. Since the network cannot accurately

update its internal representation, it naturally selects the input that is the closest to its actual intrinsic velocity. When the input with incompatible speed is closer to the

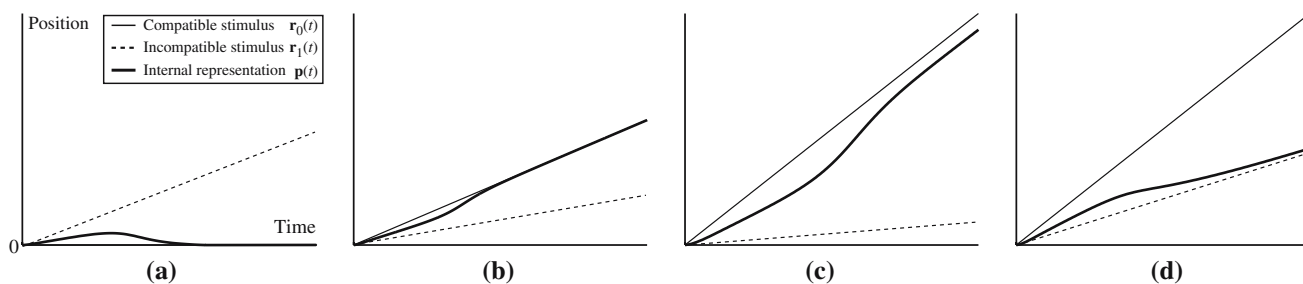


Fig. 11 Illustration of network response $p(t)$ when a stimulus compatible with the intrinsic network velocity and an incompatible one are applied to the inputs of the network. The positions of the compatible stimulus (*light filled line*), the incompatible stimulus (*dotted line*), and the network response (*dark filled line*) are shown over time in several situations. **a–c** As soon as the stimuli are sufficiently separated in neural space, the network successfully selects the compatible stimulus. Note that the trajectory of the internal representation $p(t)$ exhibits, initially, a deviation from

the compatible stimulus trajectory. As described in Fig. 10, this is the result of the temporary overlap between the two stimulus representations on neural space; the network thus cannot discriminate between the two. However, as soon as the stimuli are sufficiently separated, the selection is performed. **c** Since the speed of the compatible stimulus is above the maximal network velocity, a constant lag can be observed. **d** In a similar out-of-bounds situation, the network selects the wrong input. Indeed, the input speed is here closer to that of the network

integration boundary than that of the compatible one, it is selected.

These results confirm that our model can account for the importance of the precise timing between the prediction of the movement outcome and the sensory feedback. Indeed, this allows our model to select which of the stimuli is under self-control (Haggard and Clarke 2003). Consequently, it also proposes a neural mechanism contributing to this cognitive function and suggests how it may be realized within a shared representation.

4 Discussion

In this paper, we have presented a continuous attractor network model capable of integrating velocity commands for updating its internal representation. Moreover, we have also considered and defined adequate external input forms such that they can positively influence the network integration dynamics toward their own. Further, the analysis of the network dynamics reveals that our model should not be restricted to representing the mechanism of a single brain region. Rather, its various properties suggest that its architecture might be more dispersed across brain areas, such as the cerebellum and the motor, visual, and associative cortices.

Other models addressing the integration of velocity commands have already been described in the literature. The vast majority of them consider the rat head direction system, its hippocampal places fields, and its abilities for path integration (Redish et al. 1996; Stringer et al. 2004; Xie et al. 2002; Zhang 1996). Despite the fact that the biological plausibility of their implementation is still under debate, a series of models assumes the existence of

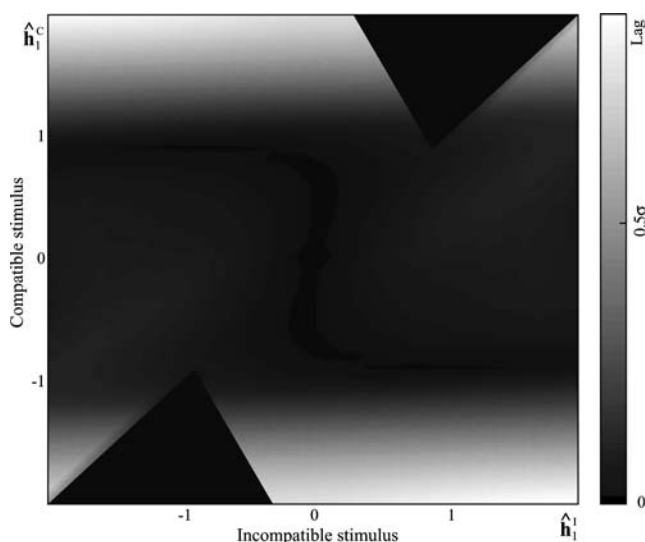


Fig. 12 Range of compatible-incompatible stimulus velocities leading the network to a false decision (*triangular black regions*). Moreover, in the case of a correct decision, the figure also shows the lag measured between the location r_0^c of the compatible stimulus and the network response p . The lag is given relative to the breadth σ of the stimulus input. As can be seen, when the compatible stimulus speed is over the network integration limit, an increasing lag between the input and the network response is observed

sigma-pi units, i.e., neurons performing both a sum and a product of their inputs (Redish et al. 1996; Stringer et al. 2004; Zhang 1996). Although our architecture allows for avoiding the use of such computational units, our model is nevertheless largely inspired from the computational principles that were described by Zhang (1996) and may even, under certain assumptions, be reduced to his model. Indeed, by averaging over the sublayers sharing the same preferential movement direction, only

a neural field spanning the neural space of the main variable remains. Then, by replacing the background input by a multiplicative scaling of the asymmetric recurrent weights, our network model matches the above-mentioned velocity integration model (Zhang 1996). Despite this similarity, our model, whose structure is also very close to that described by (Xie et al. 2002), we first consider explicitly external inputs to the system and their interaction properties. Second, our model analysis has given several new predictions and hypotheses that will be described throughout the following discussion.

4.1 Velocity tuning

Velocity and direction tuning are properties of neurons that have been found primarily in visual cortices, such as V1 and V2. They have been shown to fire preferentially for stimuli moving at a specific speed and direction within their receptive field (Orban et al. 1986). This neurophysiological finding has been suggested to be the basis of the human ability to discriminate velocity (De Bruyn and Orban 1988). Earlier work by Mineiro et al. (1998) has already demonstrated how a neural field endowed with a fixed asymmetric recurrent connectivity can exhibit such a sensitivity. However, this network is constrained to exhibit a single velocity tuning (Mineiro and Zipser 1998). Consequently, in order to be sensitive to a broader range of velocities, this network should be replicated a large number of times with different weight strengths. This approach has been followed by other models addressing the problem of stimulus velocity discrimination (Chey et al. 1998). In contrast, the present model proposes a mechanism relying on sublayers having opposite directional tuning, where the precise velocity tuning is then driven dynamically by the background input. Indeed, velocity discrimination is a dynamic process where a stimulus is first presented to subjects so that they can internalize its velocity for further discrimination against other stimuli. In addition, although it has been shown that visual areas, such as V1 and V2, possess a wide range of velocity-tuned neurons (Goodwin and Henry 1975; Orban et al. 1986), these areas are likely to be located in earlier stages of the whole visual process, as compared to area MT. This area sends direct projections to the parietal cortex, which is suggested to be the locus of some decisional processing (Wise et al. 1997). Therefore, the information carried by MT is more likely to be responsible for the discrimination process. Moreover, in this area, it has been shown that the distribution of velocity-tuned cells appears to be not uniform but rather centered near some extreme values (Cheng et al. 1994). These findings seem to confirm our modeling approach. It suggests that the brain may be able to

adapt the dynamics of its internal representation toward that of the stimulus, and hence to build a model to be used for further discrimination. Finally, for sensitivities in 2D retinal space, the complexity of our model is of order $O(n^3)$, where n is the number of neurons along a single dimension, whereas the previously mentioned approach is of order $O(n^4)$ (Chey et al. 1998; Mineiro and Zipser 1998).

4.2 Sensorimotor transformations and motor control

The transmission time of information across neural structures is critical for time-dependent tasks such as movement control. However, the brain is known to suffer from delays arising from the substrate where information is encoded, i.e., the neurons. In addition to the single neuron's integration time constant, the high density of the recurrent connectivity among cortical columns and across brain regions, while providing the brain with high computational power, adds an even stronger inertia to the information flow and hence increases the overall system time constant. Both neurophysiological and modeling studies have already described the consequences in response latency of intralayer recurrent and center-surround connectivity during simple stimulus-response tasks (Panzeri et al. 2001; Raiguel et al. 1999). Moreover, the recurrent connectivity has another side effect, not captured by these studies, which is the difficulty of moving from one attractor state to another.

This problem has several implications regarding recently proposed neural mechanisms that may be the basis of sensorimotor transformations (Burnod et al. 1999; Deneve et al. 1999; Salinas and Thier 2000; Sauser and Billard 2005; Scherberger and Andresen 2003). Indeed, these approaches rely on populations of neurons encoding basis functions and grouped within gain fields, which are neural substrates connected reciprocally and which receive changing inputs from several external sources. For example, to compute the location of a visual target in body-centered coordinates, the brain is supposed to merge information related to the target location in retinal space, to the displacement of the eyes relative to the head, and to the direction of the head in body-centered coordinates. Then, despite the inertia resulting from the recurrent connectivity, the desired information must be read out from this mixed representation with a latency as short as possible. Nevertheless, the brain successfully performs that operation. For instance, in the visual cortices of the monkey, neurons have been found to fire even before a saccade brings a stimulus into their receptive field (Nakamura and Colby 2002; Unema and Goldberg 1997). This suggests that this neural activity may correspond to a predictive

updating of the visual representation. This mechanism might explain, at least partially, why eye tracking of predictable targets such as one's own hand can be performed with almost no latency (Miall and Reckess 2002; Vercher et al. 1991). Neurophysiological findings also indicate that large groups of neurons in the cerebellum display a preferential tuning to arm position and movement direction. These neurons also fire with almost no latency when compared to real arm dynamics (Fu et al. 1997; Roitman et al. 2005). Together, these results highly suggest that the brain may use internal forward models in order to predict the consequences of upcoming movements.

Based on these biological results and hypotheses, our model proposes *a neural mechanism that links the velocity tuning properties of groups of neurons to their abilities to update their own sensory representation through velocity integration*. For instance, by using such a mechanism, the overall integration time of gain fields would be reduced drastically and thus would result in more efficient sensorimotor transformations. Moreover, neurons having a preferential tuning to position and velocity like those of our model have been found in the cerebellum (Roitman et al. 2005), a brain area suggested to be involved in both the inverse and forward control of movements (Miall and Reckess 2002; Vercher and Gauthier 1988; Wolpert and Kawato 1998). Therefore, in addition to the suggested role of neurons of the cerebellum and of motor cortices in the direct control of movements (Georgopoulos 1996; Schweigenhofer et al. 1998; Todorov 2000), their sensitivity to nonlinear mixtures of information, such as arm position and velocity, may also suggest an intrinsic dynamic process within a cortical column. As proposed by our model architecture, their firing might be used directly within their internal representation for a rapid and predictive-like updating, which therefore would not need to wait for slow sensory feedbacks.

4.3 Shared representations: motor imagery and imitation

Our model also has some implications concerning motor imagery and imitation-related mechanisms. First, motor imagery is the ability to mentally imagine oneself or someone else performing a movement. Recent findings suggest that this mental operation is performed in motor terms, i.e., by activating parts of the motor cortices that would be effectively involved in overt movement execution (Fogassi and Gallese 2002; Jeannerod and Decety 1995; Porro et al. 1996). Computationally, this hypothesis implies that some motor areas, receiving projections from proprioceptive feedback during nor-

mal motor execution, should also be able to process motor commands without being influenced by this aforementioned information. Under this problematic issue, our model analysis provides some insight as to a potential neural mechanism resolving this problem. Indeed, it suggests that the same neural substrate, by increasing its global excitation level, can detach itself from the efferent sensory inputs, and hence perform this imagery task freely. By lowering this excitation, the external inputs can lock the network back to the actual sensory state, resetting its internal representation. In addition, our model may also, to some extent, contribute to the explanation of the neural mechanisms underlying illusory perception of one's own body resulting from epileptic seizures, such as the autoscopic phenomenon (Blanke and Mohr 2005; Blanke et al. 2002). Indeed, since epileptic seizures are defined as an abnormal synchronization of electrical neuronal activity, this would correspond in our model to an abnormally high excitation of the network. Let us consider, for example, the proprioceptive and the vestibular systems, which are responsible for the representation of self-body schema and self-localization in space. An epileptic seizure affecting these systems would produce in an "unlocking" or loss of control of the network representation from the real sensory state. As mentioned by the authors of these studies, such an effect may result in a disintegration of self processing in brain areas related to self and others (Blanke and Mohr 2005). Backpropagating this conflicting information to visual areas may produce the reported hallucinations.

Furthermore, shared representations have also been found when considering movement observation and execution. Indeed, recent neurophysiological studies indicate that both the observation and the execution of actions activate a shared complex of brain areas, usually called the mirror neuron system (Iacoboni et al. 1999; Rizzolatti et al. 2001). The discovery of this network of brain regions has led to several very attractive hypotheses related to the underlying neural mechanisms of imitation and of theory of mind (Arbib 2002; Decety and Sommerville 2003; Gallese and Goldman 1988; Keysers and Perrett 2004; Wolpert et al. 2003). It is, for instance, suggested that when we observe another human being performing an action, we unconsciously simulate the same action with our own representation of our body and world. This may allow us, to some extent, to predict what the outcome of that observed movement will be. However, as with the problem of the shared representation during motor imagery, this calls for explanations as to what happens when we simultaneously execute an action and observe others, and as to how the brain minimizes conflicts arising from these potentially contradictory sources of information. Again, a partial answer

may be derived from our modeling framework. In contrast to motor imagery, sensory feedback is here needed by motor areas to realize normal performance. Nevertheless, our model also provides here a computational hypothesis. As shown in Sect. 3.5, when the network receives sensory feedback that is compatible with the background input controlling its intrinsic dynamics, it can track and keep locking to that reafferent input, even if a distractor such as an observed movement of others is present. Moreover, our model also predicts that errors may increase with the decreasing ability of the network to update its internal representation accurately with speed. At very high movement velocities, the pattern of errors should reflect a bias toward a boundary corresponding to the internal limit for movement integration.

Finally, our modeling study also gives some insights on a plausible neural medium for learning by imitation through motor resonance (Gallese and Goldman 1988; Oztop et al. 2006; Wolpert et al. 2003). This principle has been proposed to explain the mirror response of the neurons in premotor areas by suggesting that the perception of others' movements and actions activates in parallel our internal representations of motor plans. Through competition, the plan corresponding best would be selected and may further be used for prediction of movement outcome, action understanding, and imitation. Since our model can potentially represent any state space, it may implement a motor plan by setting its intrinsic dynamics to correspond to that of the motor plan. Then, the state of a demonstrator's movement may drive the network. And finally, by monitoring the mean energy of the neural ensemble that is maximal when the internal dynamics match that of the external input, it may thus be possible for the brain to select the best among activated motor plans. Moreover, as already suggested by earlier modeling works (Demiris and Hayes 2002; Oztop et al. 2006; Wolpert et al. 2003), imitation processes may use such a comparison value given here by our energy function to perform a gradient ascent on that function. Within this framework, movement imitation would consist in maximizing the energy of the shared representation of self and others' movements.

5 Conclusion

This paper describes an extension of classical dynamic neural field models, which are known to exhibit marginal attractor states corresponding to compact activity packets on the neural surface. Our extension provides additional properties to these models. One such property is to allow the network to modify its intrinsic

dynamics using external commands provided, for instance, by motor efference copies or forward models. Consequently, the typical marginal and static attractor states of such models become marginal linear trajectories in neural space. This property allows the updating of the internal representation in a predictive fashion, without the need to wait for slow sensory feedbacks. Moreover, the model dynamic properties were analyzed with a special focus on the interactions between external stimuli and the network intrinsic dynamics. Finally, the biological plausibility of the model and its implications concerning several brain computational mechanisms were discussed in light of neurophysiological and behavioral data. The discussed topics include velocity tuning to visual stimuli, sensory discrimination, sensorimotor transformations, motor control, motor imagery, and imitation.

Acknowledgements E. L. Sauser would like to specially thank G. Schöner for a helpful discussion concerning the problem of response latency occurring in recurrent models of sensorimotor transformations. This work was supported by the Swiss National Science Foundation, through grant no. 620-066127 of the SNF Professorships program.

A A coupled attractor model

In this paper, the following continuous attractor network dynamics are considered;

$$\begin{aligned} \tau \dot{u}(\mathbf{r}, \mathbf{s}, t) = & -u(\mathbf{r}, \mathbf{s}, t) + h(\mathbf{s}, t) + x(\mathbf{r}, \mathbf{s}, t) \\ & + \iint [W(\mathbf{r} - \mathbf{r}') - \lambda \nabla W(\mathbf{r} - \mathbf{r}') \cdot \mathbf{s}'] \\ & \times f(u(\mathbf{r}', \mathbf{s}', t)) d\mathbf{r}' d\mathbf{s}', \end{aligned} \quad (17)$$

where $u(\mathbf{r}, \mathbf{s}, t)$ is the membrane potential of a neuron with time constant τ , preferentially tuned to \mathbf{r} , a variable in stimulus space \mathcal{D}_r , and to $\mathbf{s} \in \mathcal{D}_s$. We consider both a ring and a torus attractor (Table 1). $f(u)$ is the activation function chosen to be the linear threshold function: $\max(0, u)$. W is a center-surround, symmetric, and Gaussian-like recurrent weights kernel, $\lambda > 0$ a scaling factor, and ∇ the gradient operation along \mathbf{r} . As will be shown later, a resulting effect of the second term of the recurrent connectivity is that the neuron's sensitivity to the variable \mathbf{s} implicitly corresponds to a preferred movement direction along the other variable \mathbf{r} . The external inputs are decomposed into two parts: the background input $h(\mathbf{s}, t)$ and the stimulus inputs $x(\mathbf{r}, \mathbf{s}, t)$.

In what follows, our interest is primarily in the dynamics of the interactions between the proposed network and its external inputs. These interactions will be shown to modify the attractor states of the network. We will

begin by introducing the case where the background input alone is sufficient to drive the network marginal attractor states to become marginal linear trajectories in the space \mathcal{D}_r . Then, we will consider how a stimulus input should be defined in order to drive the network toward its own motion dynamics. We will first define a nonlinear form of input, since its derivation from the background input form is the most straightforward. Then, we will slightly modify this input description by defining a linear form of it, which is more suitable for transferring information across neural populations. Indeed, this linear form of input will further help us to derive adequate synaptic projections so that, within a larger network of neural populations, instances of our model may transfer their encoded information across each other. Further, since such network dynamics can only be described fully analytically for certain rare special cases (Amari 1977; Sauser and Billard 2005; Xie et al. 2002; Zhang 1996), we will assume some linear approximations around equilibrium points. We start by considering general stable solutions as existing and then develop our mathematical argument based on them. Indeed, the nature of the marginally stable solutions of such systems, known as activity bumps, has already been long described in the literature (Amari 1977; Salinas and Thier 2000; Sauser and Billard 2005; Wilson and Cowan 1973; Xie et al. 2002; Zhang 1996).

A.1 Static case

Before addressing inhomogeneous background inputs along \mathbf{s} , let us begin by addressing the static case, where the background input is balanced and where no spatial input is applied to the network, i.e., respectively, $h(\mathbf{s}, t) = h_0$ and $x(\mathbf{r}, \mathbf{s}, t) = 0$. Omitting the system variables, we consider a marginally stable and static solution given by u^* and $f^* = f(u^*)$. In this case, since the system is constant in \mathbf{s} , the second term of the recurrent connectivity vanishes through the closed integral, which, by rewriting Eq. (17), gives

$$u^* - \oint W * f^* ds' = h_0, \tag{18}$$

where $*$ denotes the convolution operation along \mathbf{r} . Then, in order to extract the homogeneous term h_0 from the solutions, we define a normalization of these solutions by the following substitution:

$$u^* = h_0 U_0^* \text{ and } f^* = h_0 F_0^*, \tag{19}$$

where $U_0^*(\mathbf{r})$ and $F_0^*(\mathbf{r})$ are solutions only defined on \mathbf{r} . Further, rewriting Eq. (18) gives

$$U_0^* - 1 = \oint W * F_0^* ds'. \tag{20}$$

From this, the multiplicative effect of the background homogeneous input h_0 on the solutions of the system can be observed. This property is the basis of the non-linear behavior of the neural fields in general (Salinas and Thier 2000; Sauser and Billard 2005). This implies that we can restrict our analysis, without loss of generality, by considering normalized solutions of the system, since h_0 only acts as a scaling factor.

A.2 Velocity integration

We are now interested in how the background input may drive a stable activity packet toward a similar but traveling bump with velocity \mathbf{v}^* along the space defined on \mathbf{r} . A usual transformation performed in such a situation is to focus on a frame of reference moving with the bump so that, in this new frame, it appears static (Xie et al. 2002; Zhang 1996). We consider the following variable substitution:

$$\tilde{\mathbf{r}} = \mathbf{r} - \int_0^t \mathbf{v}^*(t') dt'. \tag{21}$$

In this new frame, $\dot{u}(\mathbf{r}, \mathbf{s}, t) = -\mathbf{v}^*(t) \cdot \nabla u(\tilde{\mathbf{r}}, \mathbf{s}, t) + \dot{u}(\tilde{\mathbf{r}}, \mathbf{s}, t)$, where ∇ is the gradient operator along $\tilde{\mathbf{r}}$. This leads Eq. (17) to become

$$\begin{aligned} -\tau \mathbf{v}^*(t) \cdot \nabla u(\tilde{\mathbf{r}}, \mathbf{s}, t) \\ + \tau \dot{u}(\tilde{\mathbf{r}}, \mathbf{s}, t) = -u(\tilde{\mathbf{r}}, \mathbf{s}, t) + h(\mathbf{s}, t) + x(\tilde{\mathbf{r}}, \mathbf{s}, t) \\ + \oint [W(\tilde{\mathbf{r}} - \tilde{\mathbf{r}}') - \lambda \nabla W(\tilde{\mathbf{r}} - \tilde{\mathbf{r}}') \cdot \mathbf{s}'] \\ \times f(u(\tilde{\mathbf{r}}', \mathbf{s}', t)) d\tilde{\mathbf{r}}' ds'. \end{aligned} \tag{22}$$

As mentioned earlier in the static case, a constant input $h(\mathbf{s}, t) = h_0$ leads the system to be constant along \mathbf{s} , which suppresses the second term of the recurrent connectivity. Moreover, as can be guessed, this term will be responsible for compensating the new term on the left-hand side of Eq. (22). Therefore, we need to break the symmetry by adding $\hat{\mathbf{h}}_0(t) \cdot \mathbf{s}$, an asymmetric term, to the driving background input such that

$$h(\mathbf{s}, t) = h_0 [1 + \hat{\mathbf{h}}_0(t) \cdot \mathbf{s}], \tag{23}$$

where $\hat{\mathbf{h}}_0$ corresponds to the normalized strength of the asymmetry breaking relative to the constant term h_0 . If, for a while, we omit terms containing a gradient expression in Eq. (22), an approximate solution \tilde{u}^* along \mathbf{s} can be found, which is equivalent to Eq. (19) up to a constant given by the asymmetric term in Eq. (23). In a compact form, the solution of the system can be written as

$$\tilde{u}^*(\tilde{\mathbf{r}}, \mathbf{s}, t) = h_0 [\hat{\mathbf{h}}_0(t) \cdot \mathbf{s} + U_0^*(\tilde{\mathbf{r}})], \tag{24a}$$

where $U^*(\tilde{\mathbf{r}})$ is given by Eq. (20). Then, from Eq. (19), where the output function f^* is proportional to the background homogenous input, we assume in this case that an approximate \tilde{f}^* can still be obtained by such a nonlinear transformation. We hence consider a linear approximation around the equilibrium point, i.e., the static case, which is given by

$$\tilde{f}^*(\tilde{\mathbf{r}}, \mathbf{s}, t) = h_0 [1 + \gamma_0 \hat{\mathbf{h}}_0(t) \cdot \mathbf{s}] F_0^*(\tilde{\mathbf{r}}), \tag{24b}$$

where γ_0 corresponds to the linear approximation factor, which, except for rare cases, may not be found analytically. It corresponds to the slope of the relation between the asymmetry of the background input and the velocity of the network response. It strictly depends on the recurrent weight profile, whose convolution is hard to solve analytically (Zhang 1996). Now, we return to the full form of Eq. (22) and show how the background input should be set in order for the network representation to exhibit the desired traveling activity profile. Let us begin by considering the recurrent synaptic drive. First, developing its symmetric part gives

$$\begin{aligned} \oint W * \tilde{f}^* ds' &\stackrel{(24b)}{=} h_0 \oint W * [1 + \gamma_0 \hat{\mathbf{h}}_0 \cdot \mathbf{s}'] F_0^* ds' \\ &= h_0 \left[\oint W * F_0^* ds' + \gamma_0 \oint W * F_0^* [\hat{\mathbf{h}}_0 \cdot \mathbf{s}'] ds' \right] \\ &\stackrel{(20)}{=} h_0 [(U_0^* - 1) + 0] = h_0 (U_0^* - 1). \end{aligned}$$

Similarly, considering its asymmetric component, we obtain

$$\begin{aligned} - \oint \lambda \nabla W \cdot \mathbf{s}' * \tilde{f}^* ds' &\stackrel{(24b)}{=} -h_0 \lambda \oint \nabla W \cdot \mathbf{s}' * [1 + \gamma_0 \hat{\mathbf{h}}_0 \cdot \mathbf{s}'] F_0^* ds' \\ &= -\lambda h_0 \left[\oint \nabla W \cdot \mathbf{s}' * F^* ds' + \gamma_0 \oint \nabla W \cdot \mathbf{s}' * F_0^* [\hat{\mathbf{h}}_0 \cdot \mathbf{s}'] ds' \right] \\ &= -\lambda h_0 [0 + \gamma_0 \hat{\mathbf{h}}_0 \cdot \nabla [W * F_0^*]] \\ &\stackrel{(20)}{=} -\lambda \gamma_0 h_0 \hat{\mathbf{h}}_0 \cdot \nabla (U_0^* - 1) \\ &= -\lambda \gamma_0 h_0 \hat{\mathbf{h}}_0 \cdot \nabla U_0^*. \end{aligned}$$

And, finally, the complete recurrent synaptic drive is given by

$$\oint [W - \lambda \nabla W \cdot \mathbf{s}'] * \tilde{f}^* ds' = h_0 [U_0^* - 1 - \lambda \gamma_0 \hat{\mathbf{h}}_0 \cdot \nabla U_0^*]. \tag{25}$$

Then, by substituting Eqs. (23), (24a), and (25) into the system Eq. (22), we find that the velocity of the activity blob is determined by the asymmetric component of the background input, following

$$\mathbf{v}^*(t) \approx \frac{\lambda \gamma_0}{\tau} \hat{\mathbf{h}}_0(t) \Leftrightarrow \hat{\mathbf{h}}_0(t) \approx \frac{\tau}{\lambda \gamma_0} \mathbf{v}^*(t). \tag{26}$$

Further, because the approximate linear relationship between the external commands and the network response is unbounded, let us determine the integration limit of the network. Indeed, since the network integration property relies on a competition across the sublayers, the maximum intrinsic speed is reached strictly when a single sublayer has a positive activation. In the ring attractor case, this specific network state occurs when the asymmetry in the background input $h(\mathbf{s}, t)$ is sufficiently strong to completely inhibit one of the sublayers. However, in the torus case, because of the continuous nature of the space \mathcal{D}_s spanned by the sublayers, one sublayer may eventually drive the network alone only for $\|\hat{\mathbf{h}}_0\| \rightarrow \infty$. As a consequence, regarding the network dynamics, the close integral along \mathbf{s} of the recurrent connectivity may be removed. The full recurrent synaptic drive thus becomes

$$[W(\tilde{\mathbf{r}} - \tilde{\mathbf{r}}') - \lambda \nabla W(\tilde{\mathbf{r}} - \tilde{\mathbf{r}}') \cdot \mathbf{s}'] * \tilde{f}^*(\tilde{\mathbf{r}}', \mathbf{s}', t). \tag{27}$$

where \mathbf{s}' denotes the only active sublayer given by $\mathbf{s}' = \hat{\mathbf{h}}_0 / \|\hat{\mathbf{h}}_0\|$. This recurrent drive has already been shown to lead a traveling activity peak to move with a constant velocity equal to λ/τ and with direction \mathbf{s}' (Zhang 1996). This is thus the integration limit that our model can reach.

B Stimulus input

B.1 General input form

In this section, we are interested in how a stimulus input that conveys information related to the spatial localization of a stimulus may drive the network representation to reflect the motion dynamics of that input. We consider a stimulus located at $\mathbf{r}_0(t)$ in the stimulus space and moving in phase with the network intrinsic velocity $\mathbf{v}^*(t)$, i.e., $\dot{\mathbf{r}}_0(t) = \mathbf{v}^*(t)$. The shape G of the stimulus input is assumed to be a Gaussian-like shape centered on its current location $\mathbf{r}_0(t)$. Moreover, in order to compensate for the neural dynamics and for the effect of the network recurrent connections, an additional differential term and a scaling factor are respectively needed to define a “well-behaving” input, i.e., one that does not turn into an asymmetric activity peak in the network representation when moving within the neural field. It is defined by

$$x(\mathbf{r}, \mathbf{s}, t) = h_1 [G(\mathbf{r} - \mathbf{r}_0(t)) - \tau \dot{\mathbf{r}}_0(t) \cdot \nabla G(\mathbf{r} - \mathbf{r}_0(t))] [1 + \epsilon \dot{\mathbf{r}}_0(t) \cdot \mathbf{s}], \tag{28}$$

where ϵ is a temporary constant. As can be noticed while looking at the network dynamics, the second differential term here compensates for the similar term in Eq. (22).

Then, considering the asymmetry in the strength of the input relative to the neuron preferred movement direction, it modifies the network intrinsic dynamics in a similar way to that of the background input (Eq. (23)). Moreover, we also assume that the velocity term $\dot{\mathbf{r}}_0(t)$ of the stimulus input can be expressed in a form comparable to that of the background homogenous input given by Eqs. (23) and (26) such that

$$\dot{\mathbf{r}}_0(t) = \frac{\lambda \gamma_1}{\tau} \hat{\mathbf{h}}_1(t), \tag{29}$$

where $\hat{\mathbf{h}}_1(t)$ is a directional vector similar to $\hat{\mathbf{h}}_0(t)$ defined previously. It determines the strength of the stimulus asymmetry and, thus, the stimulus velocity. γ_1 is like γ_0 and corresponds to a linear approximation factor. Then, rewriting Eq. (28) using (29) and writing $\epsilon = \tau/\lambda\gamma_0$ gives

$$x(\mathbf{r}, \mathbf{s}, t) = h_1 [G(\mathbf{r} - \mathbf{r}_0(t)) - \lambda \gamma_1 \hat{\mathbf{h}}_1(t) \cdot \nabla G(\mathbf{r} - \mathbf{r}_0(t))] [1 + \hat{\mathbf{h}}_1(t) \cdot \mathbf{s}]. \tag{30}$$

In the moving frame of reference, using the substitution given in Eq. (21), the stimulus input equation becomes

$$x(\tilde{\mathbf{r}}, \mathbf{s}, t) = h_1 [G(\tilde{\mathbf{r}} - \tilde{\mathbf{r}}_0) - \lambda \gamma_1 \hat{\mathbf{h}}_1(t) \cdot \nabla G(\tilde{\mathbf{r}} - \tilde{\mathbf{r}}_0)] [1 + \hat{\mathbf{h}}_1(t) \cdot \mathbf{s}], \tag{31}$$

where $\tilde{\mathbf{r}}_0$ is constant since we would like the stimulus to move according to $\dot{\mathbf{r}}_0(t) = \mathbf{v}^*(t)$. Again, similarly to the background input case described before [Eqs. (24a) and (24b)], we also assume here a linear approximation of the network activity around the equilibrium point. In addition, we also assume that a superposition of solutions may result from both forms of inputs, i.e., $h(\mathbf{s}, t)$ and $x(\tilde{\mathbf{r}}, \mathbf{s}, t)$. This leads to a family of solutions given by

$$\tilde{u}^* = h_0 [U_0^* + \hat{\mathbf{h}}_0 \cdot \mathbf{s}] + h_1 [U_1^* + \hat{\mathbf{h}}_1 \cdot \mathbf{s}G] \tag{32a}$$

$$\tilde{f}^* = h_0 [1 + \gamma_0 \hat{\mathbf{h}}_0 \cdot \mathbf{s}] F_0^* + h_1 [1 + \gamma_1 \hat{\mathbf{h}}_1 \cdot \mathbf{s}] F_1^*, \tag{32b}$$

where

$$-1 + U_0^* = \oint W * F_0^* ds' \tag{33a}$$

$$-G + U_1^* = \oint W * F_1^* ds'. \tag{33b}$$

From these equations we can determine the value that the parameters of the stimulus input should take in order to respect the network dynamics. Again developing the system Eq. (22) results here in

$$-\tau \mathbf{v} \cdot \nabla \tilde{u}^* + \tilde{u}^* = h + x + \oint [W - \lambda \nabla W \cdot \mathbf{s}'] * \tilde{f}^* ds' \tag{23),(26),(30),(32),(33)}$$

$$\begin{aligned} & -\lambda \gamma_0 \hat{\mathbf{h}}_0 [h_0 \nabla U_0^* + h_1 \nabla [\hat{\mathbf{h}}_1 \cdot \mathbf{s}G + U_1^*]] \\ & + h_0 [\hat{\mathbf{h}}_0 \cdot \mathbf{s} + U_0^*] + h_1 [\hat{\mathbf{h}}_1 \cdot \mathbf{s}G + U_1^*] \\ & = h_0 [1 + \hat{\mathbf{h}}_0 \cdot \mathbf{s}] + h_1 [G + \lambda \gamma_1 \hat{\mathbf{h}}_1 \cdot \nabla G] [1 + \lambda \gamma_1 \hat{\mathbf{h}}_1 \cdot \mathbf{s}] \\ & + h_0 [U_0^* - 1 - \lambda \gamma_0 \hat{\mathbf{h}}_0 \cdot \nabla U_0^*] \\ & + h_1 [U_1^* - G - \lambda \gamma_1 \hat{\mathbf{h}}_1 \cdot \nabla [U_1^* - G]] \end{aligned}$$

\Leftrightarrow

$$\lambda \gamma_0 \hat{\mathbf{h}}_0 [h_1 \nabla [\hat{\mathbf{h}}_1 \cdot \mathbf{s}G + U_1^*]] = \lambda \gamma_1 \hat{\mathbf{h}}_1 [h_1 \nabla [\hat{\mathbf{h}}_1 \cdot \mathbf{s}G + U_1^*]].$$

And thus we find that

$$\gamma_1 \stackrel{!}{=} \gamma_0 = \gamma \quad \text{and} \quad \hat{\mathbf{h}}_1 \stackrel{!}{=} \hat{\mathbf{h}}_0 \tag{34}$$

Therefore, in order for the network intrinsic dynamics to follow its input, the asymmetric breaking of the network along \mathbf{s} must match that of the stimulus input.

B.2 Linear input form

From now on, the form of the stimulus input is defined using a velocity-dependent scaling factor, which needs a nonlinear operation to be constructed. In what follows, we will thus try to define a different form of input containing only linear terms. Indeed, as will be shown further, this form of input will help us to derive an expression for the synaptic projections across instances of our model, which will allow them to completely transfer information they are encoding. Let us rewrite the stimulus input in the moving frame of reference, as defined in Eq. (31), by replacing the velocity-dependent nonlinear factor with a velocity-dependent linear term such that

$$x(\tilde{\mathbf{r}}, \mathbf{s}, t) = h_1 [G(\tilde{\mathbf{r}} - \tilde{\mathbf{r}}_0) - \lambda \gamma \hat{\mathbf{h}}_0(t) \cdot \nabla G(\tilde{\mathbf{r}} - \tilde{\mathbf{r}}_0) + \eta \hat{\mathbf{h}}_0(t) \cdot \mathbf{s}], \tag{35}$$

where $\eta > 0$ is a scaling factor. Then, in a similar fashion to the previous sections, we also consider a family of solutions given by

$$\tilde{u}^* = h_0 [U_0^* + \hat{\mathbf{h}}_0 \cdot \mathbf{s}] + h_1 [\eta \hat{\mathbf{h}}_0 \cdot \mathbf{s}] + h_1 U_1^* \tag{36a}$$

$$\tilde{f}^* = h_0 [1 + \gamma \hat{\mathbf{h}}_0 \cdot \mathbf{s}] F_0^* + h_1 [\gamma \eta \hat{\mathbf{h}}_0 \cdot \mathbf{s}] F_0^* + h_1 F_1^*, \tag{36b}$$

where the system of Eqs. (33) still apply. Again, inserting these equations into the system dynamics (22) gives

$$-\tau \mathbf{v} \cdot \nabla \tilde{u}^* + \tilde{u}^* = h + x + \oint [W - \lambda \nabla W \cdot \mathbf{s}'] * \tilde{f}^* ds' \tag{(23),(26),(33),(35),(36)}$$

$$\begin{aligned}
 & -\lambda\gamma \hat{\mathbf{h}}_0 [h_0 \nabla U_0^* + h_1 \nabla U_1^*] \\
 & + h_0 [U_0^* + \hat{\mathbf{h}}_0 \cdot \mathbf{s}] + h_1 [\eta \hat{\mathbf{h}}_0 \cdot \mathbf{s}] + h_1 U_1^* \\
 & = h_0 [1 + \hat{\mathbf{h}}_0 \cdot \mathbf{s}] + h_1 [G - \lambda\gamma \hat{\mathbf{h}}_0 \nabla G + \eta \hat{\mathbf{h}}_0 \cdot \mathbf{s}] \\
 & + h_0 [U_0^* - 1 - \lambda\gamma \hat{\mathbf{h}}_0 \cdot \nabla U_0^*] \\
 & - \lambda\gamma \eta h_1 \hat{\mathbf{h}}_0 \cdot \nabla U_0^* + h_1 [U_1^* - G] \\
 & \Leftrightarrow \\
 & \nabla [U_1^* - G] \stackrel{!}{=} \eta \nabla U_0^*. \tag{37}
 \end{aligned}$$

Therefore, in order for the network intrinsic dynamics to follow that of the input, the gradient of the stimulus-related solution $[U_1^* - G]$ must be proportional to that of the homogeneous background input U_0^* . However, depending on the shape of G , this may not necessarily be the case. Nevertheless, since the solutions U_i^* are the result of an equilibrium between the input and the drive of the recurrent weights, we can assume that the latter dominate. As a result, both sides of Eq. (37) can be considered to be very similar and hence proportional. We can thus find an approximate value for η such that

$$\nabla [U_1^* - G] \approx \eta \nabla U_0^*, \tag{38}$$

which, subsequently, using the system of Eqs. (33) is equivalent to

$$\nabla \oint W * F_1^* ds' \approx \eta \nabla \oint W * F_0^* ds'. \tag{39}$$

B.3 Synaptic projections

In this section, we are interested in synaptic projections across populations. Indeed, since the model we propose might be integrated in a larger network of neural populations, we must determine how these populations may transfer information they are encoding. We consider here two neural populations, A and B. In what follows, these names will be used as indices in order to clearly identify the corresponding variables and parameters of each population. The synaptic projections W_{AB} refer to a full directional connectivity between populations A and B. It is given by

$$x_B(\mathbf{r}, \mathbf{s}, t) = \iint W_{AB}(\mathbf{r}, \mathbf{s}, \mathbf{r}', \mathbf{s}') f(u_A(\mathbf{r}', \mathbf{s}', t)) d\mathbf{r}' d\mathbf{s}' \tag{40}$$

such that input x_B is fed to population B. Moreover, W_{AB} will be defined so that information related to both position and velocity will be transferred from population A to B, such that $\mathbf{p}_A = \mathbf{p}_B$ and $\dot{\mathbf{p}}_A = \dot{\mathbf{p}}_B$, where \mathbf{p}_i is the population vector of population i defined in the main text by Eq. (2). Throughout this section, we consider the

following convolution weights:

$$W_{AB}(\mathbf{r}, \mathbf{s}, \mathbf{r}', \mathbf{s}') = [W_T(\mathbf{r} - \mathbf{r}') - \lambda_A \nabla W_T(\mathbf{r} - \mathbf{r}') \cdot \mathbf{s}' + \mu_{AB} \mathbf{s} \cdot \mathbf{s}'], \tag{41}$$

where W_T is a center-surround, Gaussian-like convolution kernel (see Appendix C for a precise definition). Similarly to the fundamental input shape G of the linear input form [Eq. (35)], the convolution through the projection weights will produce a fundamental input shape, denoted by G_{AB} . Moreover, since the recurrent weights preferentially link neurons sharing similar tuning properties, this guarantees that the position-related information \mathbf{p}_A of population A is correctly transferred to population B. Then, since the activity profile \tilde{f}_A^* of population A [Eq. (36)b] is composed of a linear combination of two terms $F_{A_0}^*$ and $F_{A_1}^*$, we assume that the input shape G_{AB} can be written as

$$G_{AB} = h_0 \underbrace{\oint W_T * F_{A_0}^* ds'}_{G_{AB_0}} + h_1 \underbrace{\oint W_T * F_{A_1}^* ds'}_{G_{AB_1}}, \tag{42}$$

where G_{AB_0} and G_{AB_1} correspond, respectively, to the contribution of the background and the stimulus input of population A to the input x_B feeding population B.

Further, from Eq. (39) we know that there exists an η_A such that

$$\nabla G_{AB_1} \approx \eta_A \nabla G_{AB_0}. \tag{43}$$

Then, by constraining the weight convolution kernel such that

$$\oint W_T(\mathbf{r}) d\mathbf{r} = 0 \quad \Rightarrow \quad \oint G_{AB_0} d\mathbf{r} = \oint G_{AB_1} d\mathbf{r} = 0, \tag{44}$$

both sides of Eq. (43) can be integrated and we then find

$$G_{AB_1} \approx \eta_A G_{AB_0}, \tag{45}$$

which implies

$$G_{AB} = (h_0 + \eta_A h_1) G_{AB_0}. \tag{46}$$

Further, we can define an exact input form equivalent to Eq. (35) with which we can match the projection through the weights W_{AB} . It is given by

$$\begin{aligned}
 x_B(\mathbf{r}, \mathbf{s}, t) = (h_0 + \eta_A h_1) & \left[G_{AB_0}(\mathbf{r} - \mathbf{p}_A(t)) - \lambda_B \gamma_B \hat{\mathbf{h}}_{B_0}(t) \right. \\
 & \left. \cdot \nabla G_{AB_0}(\mathbf{r} - \mathbf{p}_A(t)) + \eta_B \hat{\mathbf{h}}_{B_0}(t) \cdot \mathbf{s} \right], \tag{47}
 \end{aligned}$$

where η_B may be different from η_A since these constants depend on the recurrent weights of their respective populations and on the shape of their respective inputs, i.e.,

G and G_{AB} . Similarly, the population’s intrinsic parameters, such as γ and λ , may also be different.³ Thus let us define the relationship between them using the constraint on the velocity information transfer:

$$\dot{\mathbf{p}}_A = \dot{\mathbf{p}}_B \quad \Leftrightarrow \quad \frac{\lambda_A \gamma_A \hat{\mathbf{h}}_{A0}}{\tau} = \frac{\lambda_B \gamma_B \hat{\mathbf{h}}_{B0}}{\tau}. \quad (48)$$

We can then look for a μ_{AB} satisfying the following equation:

$$\oint W_{AB} * \tilde{f}_A^* ds' = (47), \quad (49)$$

where \tilde{f}_A^* is given by rewriting Eq. (36b) with the corresponding population indices:

$$\tilde{f}_A^* = h_0 [1 + \gamma_A \hat{\mathbf{h}}_{A0} \cdot \mathbf{s}] F_{A0}^* + h_1 [\gamma_A \eta_A \hat{\mathbf{h}}_{A0} \cdot \mathbf{s}] F_{A0}^* + h_1 F_{A1}^* \quad (50)$$

Developing the left-hand side of Eq. (49) gives

$$\begin{aligned} \oint [W_T - \lambda_A \nabla W_T \cdot \mathbf{s}' + \mu_{AB} \mathbf{s} \cdot \mathbf{s}'] * \tilde{f}_A^* ds & \quad (51) \\ \stackrel{(50)}{=} h_0 [G_{AB0} - \lambda_A \gamma_A \hat{\mathbf{h}}_{A0} \cdot \nabla G_{AB0} + \mu_{AB} \gamma_A \hat{\mathbf{h}}_{A0} \cdot \mathbf{s} \iint F_{A0}^*] \\ & + h_1 [G_{AB1} - \lambda_A \gamma_A \hat{\mathbf{h}}_{A0} \cdot \nabla G_{AB1} + \eta_A \mu_{AB} \gamma_A \hat{\mathbf{h}}_{A0} \cdot \mathbf{s} \iint F_{A0}^*] \\ \stackrel{(46)}{\approx} (h_0 + h_1 \eta_A) [G_{AB0} - \lambda_A \gamma_A \hat{\mathbf{h}}_{A0} \cdot \nabla G_{AB0} + \mu_{AB} \gamma_A \hat{\mathbf{h}}_{A0} \cdot \mathbf{s} \iint F_{A0}^*] \\ \stackrel{(48)}{\approx} (h_0 + h_1 \eta_A) [G_{AB0} - \lambda_B \gamma_B \hat{\mathbf{h}}_{B0} \cdot \nabla G_{AB0} + \mu_{AB} \frac{\lambda_B \gamma_B}{\lambda_A} \hat{\mathbf{h}}_{B0} \cdot \mathbf{s} \iint F_{A0}^*] \end{aligned}$$

And finally, by comparing Eqs. (47) and (51), it remains

$$\mu_{AB} \approx \frac{\lambda_A}{\lambda_B} \frac{\eta_B}{\gamma_B \iint F_{A0}^*}. \quad (52)$$

Therefore, using the synaptic projections W_{AB} described by Eq. (41), a population A can transfer its encoded information related to both position and velocity to another population B.

C Simulation parameters

In our experiments, in order to implement our model, precise forms of the recurrent weights and the stimulus input were defined. We used a common periodical Gaussian kernel function \mathcal{G} known as the von Mises distribution, normalized so that $\mathcal{G}(\theta) \in [0, 1]$. It is given by

$$\mathcal{G}(\theta) = \frac{e^{\frac{\cos(\theta)-1}{2\sigma^2}} - \chi}{1 - \chi} \quad \text{where} \quad \chi = e^{-1/\sigma^2}. \quad (53)$$

³ For simplicity, we assume τ to be constant across the populations.

Table 2 Simulation parameters

	Parameters values
Ring attractor	$\alpha_G = 1, \alpha_W = 2, \alpha_T = 2, \sigma = 0.3, \lambda = 0.1, h_0 = 1, h_1 = 1, \tau = 0.1$ Discretization along \mathcal{D}_r : 128 neurons \mathcal{D}_s : 2 sublayers
Torus attractor	$\alpha_G = 1, \alpha_W = 1, \alpha_T = 2, \sigma = 0.3, \lambda = 0.1, h_0 = 1, h_1 = 1, \tau = 0.1$ Discretization along \mathcal{D}_r : 32×32 neurons \mathcal{D}_s : 8 sublayers

The synaptic weights and the external input shapes were then defined by

$$W(\mathbf{r}, \mathbf{r}') = \alpha_W \left[\prod_{i=1}^N \mathcal{G}((\mathbf{r} - \mathbf{r}') \cdot \mathbf{e}_i) - 1 \right] \quad (54)$$

$$W_T(\mathbf{r}, \mathbf{r}') = \alpha_T \left[\prod_{i=1}^N \mathcal{G}((\mathbf{r} - \mathbf{r}') \cdot \mathbf{e}_i) - \delta \right] \quad (55)$$

and

$$G(\mathbf{r}, t) = \alpha_G \left[\prod_{i=1}^N \mathcal{G}((\mathbf{r} - \mathbf{r}_0(t)) \cdot \mathbf{e}_i) - \delta \right], \quad (56)$$

respectively, where $N \in \{1, 2\}$ corresponds to the dimension of the network domain \mathcal{D}_r ; \mathbf{e}_i denotes the i th component of the canonical base such that $\mathbf{e}_1 = (1, 0)$ and $\mathbf{e}_2 = (0, 1)$ in two dimensions; α_W, α_T , and $\alpha_G \in \mathbb{R}_+^*$ are scaling factors; and $\delta = \int \mathcal{G}$ removes the constant component of \mathcal{G} . The simulations were performed using the fourth-order Runge–Kutta numerical algorithm for the resolution of dynamic systems. Except where explicitly defined in the description of each experiment, the simulation parameters that were used in this paper are given in Table 2.

References

Amari S (1977) Dynamics of pattern formation in lateral-inhibition type neural fields. *Biol Cybern* 27:77–87

Arbib MA (2002) The mirror system, imitation and the evolution of language. In: *Imitation in animals and artefacts*, a bradford book. MIT Press, Cambridge pp 229–280

Ben Hamed S, Duffy C, Pouget A (2003) MSTd neuronal basis functions for the population encoding of heading direction. *J Neurophysiol* 90:549–558

Ben-Yishai R, Bar-Or RL, Sompolinsky H (1995) Theory of orientation tuning in visual cortex. *Proc Nat Acad Sci USA* 92:3844–3848

Blanke O, Mohr C (2005) Out-of-body experience, heautoscopy, and autoscopic hallucination of neurological origin: implications for neurocognitive mechanisms of corporeal awareness and self-consciousness. *Brain Res Rev* 50(1):184–199

Blanke O, Ortigue S, Landis T, Seeck M (2002) Stimulating illusory own body perceptions. *Nature* 419:269–270

- Burnod Y, Baraduc P, Battaglia-Mayer A, Guigon E, Koechlin E, Ferraina S, Laquaniti F, Caminiti R (1999) Parieto-frontal coding of reaching: an integrated framework. *Exp Brain Res* 129:325–346
- Chaminade T, Franklin D, Oztop E, Cheng G (2005) Motor interference between humans and humanoid robots: Effect of biological and artificial motion. In: *Proceedings of the 4th IEEE International Conference on Development and Learning*, Osaka, Japan pp 96–101
- Cheng K, Hasegawa T, Kadharbatcha SS, Tanaka K (1994) Comparison of neuronal selectivity for stimulus speed, length, and contrast in the prestriate visual cortical areas V4 and MT of the macaque monkey. *J Neurophysiol* 71(6):2269–2280
- Chey J, Grossberg S, Mingolla E (1998) Neural dynamics of motion processing and speed discrimination. *Vision Res* 38:2769–2786
- De Bruyn B, Orban GA (1988) Human velocity and direction discrimination measured with random dot patterns. *Vision Res* 28(12):1323–1335
- Decety J, Sommerville JA (2003) Shared representations between self and others: a social cognitive neuroscience view. *Trends Cogn Sci* 7:527–533
- Demiris Y, Hayes G (2002) Imitation as a dual process featuring predictive and learning components: A biologically plausible computational model. In: *imitation in animals and artefacts*, a bradford book, MIT Press, Cambridge pp 327–362
- Deneve S, Latham PE, Pouget A (1999) Reading population codes: a neural implementation of ideal observers. *Nat Neurosci* 2(8):740–745
- Erlhagen W, Schöner G (2002) Dynamics field theory of movement preparation. *Psychol Rev* 109(3):545–572
- Fogassi L, Gallese V (2002) The neural correlates of action understanding in non-human primates. In *Mirror Neurons and the Evolution of Brain and Language*, *Advances in Consciousness Research*, John Benjamins Publishing pp 13–35
- Fu QG, Flament D, Coltz JD, Ebner TJ (1997) Relationship of cerebellar purkinje cell simple spike discharge to movement kinematics in the monkey. *J Neurophysiol* 78(1):478–491
- Gallese V, Goldman A (1998) Mirror neurons and the simulation theory of mind-reading. *Trends Cogn Sci* 2:493–501
- Georgopoulos AP (1996) On the translation of directional motor cortical commands to activation of muscles via spinal interneuronal systems. *Cogn Brain Res* 3:151–155
- Giese M (2000) Neural model for the recognition of biological motion. In: Baratoff G, Neumann H (eds) *dynamische perception 2*, Infix Verlag, Berlin pp 105–110
- Goodwin AW, Henry GH (1975) Direction selectivity of simple striate cells: properties and mechanisms. *J Neurophysiol* 38:1524–1540
- Haggard P, Clarke S (2003) Intentional action: conscious experience and neural prediction. *Conscious Cogn* 12(4):695–707
- Hubel D, Wiesel T (1977) Ferrier lecture: functional architecture of the macaque monkey visual cortex. *Proc R Soc Lon B* 198:1–59
- Iacoboni M, Woods RP, Brass M, Bekkering H, Mazziotta JC, Rizzolatti G (1999) Cortical mechanisms of human imitation. *Science* 286:2526–2528
- Jeannerod M (2003) The mechanism of self-recognition in human. *Behav Brain Res* 142:1–15
- Jeannerod M, Decety J (1995) Mental motor imagery: a window into the representational stages of action. *Curr Opin Neurobiol* 5(6):727–732
- Jellema T, Maassen G, Perrett DI (2004) Single cell integration and aminate form, motion and location in the superior temporal cortex of the macaque monkey. *Cereb Cortex* 14:781–790
- Kettner RE, Schwartz AB, Georgopoulos AP (1988) Primate motor cortex and free arm movements to visual targets in three-dimensional space. III. positional gradients and population coding of movement direction from various movement origins. *J Neurosci* 8(8):2938–2947
- Keysers C, Perrett DI (2004) Demystifying social cognition: a hebbian perspective. *Trends Cogn Sci* 8(11):501–507
- Kilner JM, Paulignan Y, Blakemore SJ (2003) An interference effect of observed biological movement on action. *Curr Biol* 13:522–525
- Miall RC, Reckess GZ (2002) The cerebellum and the timing of coordinated eye and hand tracking. *Brain Cogn* 48:212–226
- Miall RC, Wolpert DM (1996) Forward models for physiological motor control. *Neural Netw* 9:1265–1279
- Mineiro P, Zipser D (1998) Analysis of direction selectivity arising from recurrent local interactions. *Neural Comput* 10:353–371
- Nakamura K, Colby CL (2002) Updating of the visual representation in monkey striate and extrastriate cortex during saccades. *Proc Nat Acad Sci USA* 99(6):4026–4031
- Orban GA, Kennedy H, Bullier J (1986) Velocity sensitivity and direction selectivity of neurons in areas V1 and V2 of the monkey. *J Neurophysiol* 56(2):462–480
- Oztop E, Kawato M, Arbib M (2006) Mirror neurons and imitation. *Neural Netw* 19(3):254–271
- Panzeri S, Rolls ET, Battaglia F, Lavis R (2001) Speed of feed-forward and recurrent processing in multilayer networks of integrate-and-fire neurons. *Netw Comput Neural Syst* 12:423–440
- Porro CA, Francescato MP, Cettolo V, Diamond ME, Baraldi P, Zuiani C, Bazzocchi M, di Prampero PE (1996) Primary motor and sensory cortex activation during motor performance and motor imagery: a functional magnetic resonance imaging study. *J Neurosci* 16(23):7688–7698
- Raiguel SE, Xiao DK, Marcar VL, Orban GA (1999) Response latency of macaque area MT/V5 neurons and its relationship to stimulus parameters. *J Neurophysiol* 82:1944–1956
- Redish DA, Elga AN, Touretzky DS (1996) A coupled attractor model of the rodent head direction system. *Netw Comput Neural Syst* 7(4):671–685
- Rizzolatti G, Fogassi L, Gallese V (2001) Neurophysiological mechanisms underlying the understanding of actions. *Nat Rev Neurosci* 2:661–670
- Roitman AV, Pasalar S, Johnson MTV, Ebner TJ (2005) Position, direction of movement, and speed tuning of cerebellar Purkinje cells during circular manual tracking in monkey. *J Neurosci* 25(40):9244–9257
- Rougier N (2006) Dynamic beural field with local inhibition. *Biol Cybern* 94:169–179
- Salinas E, Thier P (2000) Gain modulation: a major computational principle of the central nervous system. *Neuron* 27:15–21
- Sausser EL, Billard AG (2005) Three dimensional frames of references transformations using recurrent populations of neurons. *Neurocomputing* 64:5–24
- Sausser EL, Billard AG (2006) Parallel and distributed neural models of the ideomotor principle: an investigation of imitative cortical pathways. *Neural Netw* 19(3):285–298
- Scherberger H, Andresen RA (2003) Sensorimotor transformations. In: Chalupa LM, Werner JS (eds): *The visual neurosciences*. MIT Press, Cambridge pp 1324–1336
- Schöner G (2002) Dynamical systems approaches to neural systems and behavior. *Int Encyclopedia Soc Behav Sci* pp 10571–10575
- Schwartz A, Moran D (1999) Motor cortical activity during drawing movements: population representation during lemniscate tracing. *J Neurophysiol* 82:2705–2718

- Schweigenhofer N, Arbib MA, Kawato M (1998) Role of the cerebellum in reaching movements in humans. I. distributed inverse dynamics control. *Eur J Neurosci* 10:86–94
- Sharp PE, Blair HT, Cho J (2001) The anatomical and computational basis of the rat head-direction cell signal. *Trends Neurosci* 24(5):289–294
- Stringer SM, Rolls ET, Trappenberg TP (2004) Self-organizing continuous attractor networks with multiple activity packets and the representation of space. *Neural Netw* 17:5–27
- Todorov E (2000) Direct cortical control of muscle activation in voluntary arm movements: a model. *Nat Neurosci* 3(4):391–398
- Unema MM, Goldberg ME (1997) Spatial processing in the monkey frontal eye field. I. predictive visual responses. *J Neurophysiol* 78:1373–1383
- Vercher JL, Gauthier GM (1988) Cerebellar involvement in the coordination control of the oculo-manual tracking system: effects of cerebellar dentate nucleus lesion. *Exp Brain Res* 73:155–166
- Vercher JL, Gauthier GM, Guedon O, Blouin J, Cole J, Lamarre Y (1996) Self-moved target eye tracking in control and deafferented subjects: roles of arm motor command and proprioception in arm-eye coordination. *J Neurophysiol* 76(2):1133–1144
- Wersing H, Steil JJ, Ritter H (2001) A competitive-layer model for feature binding and sensory segmentation. *Neural Comput* 13:357–387
- Wilson H, Cowan J (1973) A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik* 13:55–80
- Wise SP, Broussaoud D, Johnson PB, Caminiti R (1997) Premotor and parietal cortex: corticocortical connectivity and combinatorial computations. *Annu Rev Neurosci* 20:25–42
- Wolpert DM, Doya K, Kawato M (2003) A unifying computational framework for motor control and social interaction. *Philos Trans R Soc* 358:593–602
- Wolpert DM, Kawato M (1998) Multiple paired forward and inverse models for motor control. *Neural Netw* 11(7–8):1317–1329
- Xie X, Hahnloser R, Seung H (2002) Double-ring network model of the head-direction system. *Phys Rev E* 66:041902
- Zhang K (1996) Representation of the spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. *J Neurosci* 16(6):2112–2126