Perceptual learning in visual hyperacuity: A reweighting model

Vision Research 51 (2011) 585–599

Contents lists available at ScienceDirect

Vision Research

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

Perceptual learning in visual hyperacuity: A reweighting model

Grigorios Sotiropoulos [a], Aaron R. Seitz [b], Peggy Seriès [a,*]

[a] School of Informatics, University of Edinburgh, Edinburgh, UK
[b] Department of Psychology, University of California, Riverside, Riverside, CA, USA

ARTICLE INFO

Article history:
Received 8 June 2010
Received in revised form 6 January 2011
Available online 18 February 2011

Keywords:
Perceptual learning
Reweighting
Hyperacuity
Specificity
Disruption

ABSTRACT

Improvements of visual hyperacuity are a key focus in research of perceptual learning. Of particular interest has been the specificity of visual hyperacuity learning to the particular features of the trained stimuli as well as disruption of learning that occurs in some cases when different stimulus features are trained together. The implications of these phenomena on the underlying learning mechanisms are still open to debate; however, there is a marked absence of computational models that explore these phenomena in a unified way. Here we implement a computational learning model based on reweighting and extend it to enable direct comparison, by means of simulations, with a variety of existing psychophysical data. We find that this very simple model can account for a diversity of findings, such as disruption of learning of one task by practice on a similar task, as well as transfer of learning across both tasks and stimulus configurations under certain conditions. These simulations help explain existing results in the literature as well as provide important insights and predictions regarding the reliability of different hyperacuity tasks and stimuli. Our simulations also shed light on the model's limitations, for example in accounting for temporal aspects of training procedures or dependency of learning with contextual stimuli, which will need to be addressed by future research.

1. Introduction

Perceptual learning has been defined by Gibson (1963) as "*any relatively permanent and consistent change in the perception of a stimulus array following practice or experience with this array*". Examples of perceptual learning in vision are improvements following training on discrimination of contrast (Adini, Sagi, & Tsodyks, 2002; De Valois, 1977; Mayer, 1983; Sowden, Rose, & Davies, 2002), orientation (Fahle, 1997; Schoups, Vogels, & Orban, 1995; Shiu & Pashler, 1992), motion (Koyama, Harner, & Watanabe, 2004; Kuai, Zhang, Klein, Levi, & Yu, 2005; Liu & Vaina, 1998) and Vernier offsets (Fahle & Edelman, 1993; Herzog & Fahle, 1997; Poggio, Fahle, & Edelman, 1992). In some of these perceptual tasks, such as orientation and Vernier offset discrimination, humans exhibit thresholds far smaller (an order of magnitude) than the diameter of foveal photoreceptors. Westheimer (1976) coined the term *hyperacuity* to describe this remarkable ability.

One of the hallmarks of perceptual learning, particularly in the hyperacuity realm, is specificity: learning improvements following training are highly specific to task, stimulus type, orientation, retinal location and eye trained (Ahissar & Hochstein, 1993; Fahle, 1997; Fahle & Morgan, 1996; Poggio et al., 1992; Shiu & Pashler, 1992). Along with a considerable body of research on specificity of perceptual learning, there are notable cases of generalization of learning, such as the experimental findings of Webb, Roach, and McGraw (2007) on transfer across tasks and orientations for three-dot stimuli (when there is a common axis of stimulus variability) and the theoretical work of Zhaoping (2009), which predicts generalization across orientations of the background but not the foreground ("target") stimulus elements as well as generalization across target locations. Understanding where learning is specific and where it generalizes provides key constraints to models of perceptual learning.

Another related, but less studied, phenomenon is disruption, whereby learning improvements in one task diminish after a similar task is subsequently practiced (e.g. Seitz et al., 2005). Disruption of learning on stimulus by practice with other stimuli has been the focus of increasing research in the last few years (Aberg, Tartaglia, & Herzog, 2009; Tartaglia, Aberg, & Herzog, 2009; Yotsumoto, Chang, Watanabe, & Sasaki, 2009; Yu, Klein, & Levi, 2004; Zhang et al., 2008) and provides further constraints for models of perceptual learning.

In the last 20 years, various computational models of perceptual learning have appeared in the literature. These models can be grouped into different categories, depending on which stage learning occurs. The first category corresponds to the theory that learning occurs through changes in the properties of neurons at early levels of the visual cortex (Karni & Sagi, 1991). These sensory

* Corresponding author.
E-mail addresses: s0563640@sms.ed.ac.uk (G. Sotiropoulos), aseitz@ucr.edu (A.R. Seitz), pseries@inf.ed.ac.uk (P. Seriès).

0042-6989/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2011.02.004

neurons are typically regarded as filters (e.g. orientation-selective neurons in V1 as Gabor filters) that can change shape (e.g. sharpen their tuning curves – Schoups, Vogels, Qian, & Orban, 2001) following practice on a perceptual task. Such practice-induced changes are known in the literature as *representation modification*. Models of the second category are based on the view that what changes during perceptual learning is the contribution of sensory neurons in the decision, i.e. the relative importance (weight) of each neuron's response in the perceptual judgment. This view implies changes in the strengths of connections between cortical sensory representations and higher-level decision-making areas. Such changes are commonly referred to as *readout modification* or *'reweighting'*. Models of this type have been put forth by several researchers (Dosher & Lu, 1999; Mollon & Danilova, 1996; Petrov, Dosher, & Lu, 2005; Vaina, Sundareswaran, & Harris, 1995; Weiss, Fahle, & Edelman, 1993) and provide an alternative to the traditional notion that the remarkable specificity of perceptual learning implies changes in low-level cortical areas (where neurons are still quite selective to various stimulus properties). Finally, other models of perceptual learning are based on the hypothesis that the mechanism of learning is the modification of recurrent connections within a single visual area (e.g. Zhaoping, Herzog, & Dayan, 2003) or top-down connections from higher level areas (e.g. Roelfsema & Ooyen, 2005; Schäfer, Vasilaki, & Senn, 2007). Mechanisms of learning are a subject of considerable debate, both among theorists and experimentalists. Not surprisingly, there is a diversity of models using one or more learning mechanisms.

Poggio et al. (1992) were the first to show that hyperacuity performance could be easily achieved by fast learning in 'task-specific modules' in early visual processing. Their model used a feedforward network whose representation units correspond to retinal photoreceptors and are equivalent to basis functions (HyperBF network). Learning consisted of adapting the locations of the basis function centres (a form of representation modification) as well as the coefficients of the basis functions (a form of reweighting). The model could learn from a small number of examples, exhibited specificity with respect to orientation and replicated the experimental finding that learning of a Vernier discrimination task does not transfer to other (distant) orientations of the line elements.

Shortly afterwards, the authors proposed a more biologically realistic feedforward network (Weiss et al., 1993) based on the HyperBF technique. Here, the representation units model orientation-selective cells, such as those found in V1. Unlike in their previous work (Poggio et al., 1992), learning occurs solely through reweighting: the basis functions of the representation units remain unchanged, with only their weights to the decision unit changing. This simple network was able to qualitatively account for the dependence of threshold on the length and the separation between the line elements of a Vernier stimulus, as determined experimentally by Westheimer and McKee (1977). The authors further showed that learning could also be achieved using an unsupervised learning rule, independent of feedback.

Following similar ideas, Vaina et al. (1995) presented a reweighting HyperBF model of perceptual learning of direction discrimination for random dots, which described motion-sensitive cells in the middle temporal area (MT). They showed that the model learned the task by 'ignoring' noise in the image and spatially integrating the motion signal. The model also showed lack of transfer to orthogonal directions – a well-supported experimental finding.

The idea that learning is associated with noise filtering was further supported by the model of Dosher and Lu (1998). These authors investigated mechanisms of perceptual learning in an orientation identification task in the periphery, in the presence of systematically varying amounts of external noise. They demonstrated experimentally that perceptual learning was a result of improvements in external noise exclusion and stimulus enhancement. They further showed that this could be accounted for in a reweighting model. They postulate that "*perceptual learning primarily serves to select or strengthen the appropriate channel and prune or reduce inputs from irrelevant channels*" and suggest that "*reducing the weights on irrelevant channels reduces the contributions of external noise and additive internal noise*".

Observing that a simple reweighting model such as a single layer perceptron can only learn very locally, e.g. around only one orientation angle, Mato and Sompolinsky (1996) sought to understand which multilayer network architectures could efficiently learn angle discrimination over a range of angles. They found that a two-layer perceptron and a 'gating' model (where the gating units select, for each stimulus, the appropriate hidden units that will dominate the decision), where plasticity occurs in the form of representation modification and reweighting, could both perform well at all angles, but differed in their predictions for transfer at angles away from the learned angle. The gating model predicts partial transfer, while the two-layer perceptron predicts negative transfer. The authors remark that there is some evidence for negative transfer (Ball & Sekuler, 1987; Fahle & Edelman, 1993; Weiss et al., 1993) but not in all subjects.

Contrasting from previous models where changes of synaptic weights were determined by the presented stimuli only, Herzog and Fahle (1998) proposed that perceptual learning might be primarily guided by top-down operations. They suggest that internal criteria actively control the learning process, for e.g. selecting which neurons must be updated and controlling the learning rate.

Zhaoping et al. (2003) defended yet another idea which is that recurrent connections might also play a crucial role in perceptual learning. These authors were interested in the fact that high hyperacuity performance can be achieved despite the fact that eye movements lead to large variations in the positions of the stimulus on the retina during learning. They showed that a nonlinear recurrent network performs much better than a pure feedforward network on the bisection task in the face of positional variance. Here, recurrent plasticity serves to counter the variabilities of the eye position in a task and stimulus dependent way, while reweighting is used to guide the decision units and learn the stabilized task.

Another perceptual learning model based on plasticity of the recurrent interactions between orientation-selective neurons in V1 is that of Teich and Qian (2003). Here, however, perceptual learning leads to the modification of the same local connections as the ones that are responsible for orientation selectivity and to changes in tuning curves, in ways consistent with the data of Schoups et al. (2001).

Petrov et al. (2005) explicitly question whether modifications of early cortical representation are necessary for perceptual learning. They considered an orientation discrimination experiment, involving a manipulation of the visual context in which the orientations were presented. They showed that the drop in performance (or 'switch costs') observed experimentally when the context was changed during learning could be accounted for in a reweighting model. They argue that representation modification are not implied by the available psychophysical and physiological evidence, nor functionally necessary.

The work of Roelfsema and Ooyen (2005) re-introduced the idea that feedback signals should be involved in perceptual learning. These authors were interested in designing a learning algorithm for multilayer networks that could simultaneously be as efficient as the backpropagation algorithm, and as biologically plausible as reinforcement learning schemes. They showed that this can be achieved by including an attentional feedback signal that limits plasticity to those units at earlier processing levels that are crucial for the stimulus response mapping. The model is shown to be able

to account for changes in the tuning of neurons in sensory areas that are induced during face categorization or orientation discrimination (representation modification) but also involves changes in the read-out connections (reweighting).

Lastly, Schäfer et al. (2007) further support the role of top-down signals in learning. Here, however, top-down connections are the only ones that are being modified during learning. Their role is to remove contextual biases due to lateral interactions in V1, thus linearizing the representation, as well as to increase the gain of excitatory neurons. The model is applied to a brightness discrimination task, where the bar stimulus to be judged is surrounded by other bars, and can be cued or not.

Evidently, in terms of complexity and of the locus of perceptual learning, there is a variety of models that is indicative of the controversy around the issues of where learning takes place and how complex a model of perceptual learning should be.

It is also noteworthy that most of the aforementioned models have been evaluated against just one (at most two, as in Roelfsema & Ooyen (2005)) set of experimental findings. One of the possible reasons for this might be that, with the exception of the models of Weiss et al. (1993) and Petrov et al. (2005), the inputs to these models were features (typically scalar, such as angle values), assuming to come from cells tuned for each value of the feature under study. For example, the input in the model of Mato and Sompolinsky (1996) is an array of units, each tuned to a particular orientation. In other words, these networks were not presented with raw stimulus images but with an abstraction of those. While this can work well in theory, it can make comparison with experimental findings using diverse stimuli and tasks less practical. Indeed, there are subtleties in the similarities or differences between different stimuli (e.g. dots and curves) that are not straightforward to capture using symbolic input units. With the notable exception of the recent work of Zhaoping (2009) for pop-out detection tasks, there has been little modeling investigation regarding which dimensions (task, stimulus, stimulus configuration) learning can generalize across. This is despite the fact that the necessity of such investigation has been pointed out by researchers in the field (Fahle, 2005). Furthermore, only the model of Petrov et al. (2005) has been used to explore (partial) disruption of learning. Some models have shown negative transfer (Mato & Sompolinsky, 1996; Zhaoping et al., 2003), which is related to disruption: if training on one task/condition results in deterioration of performance in another task, it is natural to assume that training on two tasks should disrupt learning of one or both tasks. However, it is not clear which of these tasks would be disrupted – for example, it may be that one task is dominant, and thus simultaneous training only disrupts learning of the other task. Another possibility is a "last one wins" scenario, where the task that is disrupted is the one trained first within a session (see Seitz et al. (2005) for relevant data and discussion). Modeling disruption explicitly might thus be very useful for clarifying the constraints imposed on cortical plasticity.

The central question we try to answer in this study is how far a simple reweighting model based on plasticity of the readout can go towards explaining various findings on perceptual learning. To this end, we implement a model based on the HyperBF network model of Weiss et al. (1993). In this model, learning is implemented as changes in readout weights from a network of orientation-selective simple cells (such as those found in the primary visual cortex). By means of simulation, we compare the model's predictions with published psychophysical results and subject it to a variety of tasks and stimuli, focusing on its behavior with respect to the aforementioned phenomena of specificity and disruption. Furthermore, we perform simulations using combinations of tasks and stimuli that have not been experimentally studied and derive testable predictions.

2. General methods

In all simulations presented hereafter, we use a simple model of a network of orientation-selective neurons in the primary visual cortex. The model is based upon the *HyperBF* (*hyper basis function*) *network* described in Weiss et al. (1993). We first implement the published model and compare our results to both the authors' own results and to psychophysical data presented in the same paper. We then implement an expanded model and test it against recently published studies of perceptual learning of hyperacuity as well as simulate novel stimulus and task combinations as to make predictions based upon other simulated datasets.

2.1. Model description

HyperBF networks belong to the family of radial basis function (RBF) networks (Broomhead & Lowe, 1988; Moody & Darken, 1989). Units in the hidden (middle) layer of the network represent orientation-selective neurons in early cortical areas (such as V1), i.e. simple cells that possess elongated receptive fields (RFs) with a preferred orientation. Simple cells are essentially linear filters: the response r of the neuron can be estimated by integrating the RF function over the surface of the stimulus:

$$h = \int\int_{\mathbb{R}^2} G(x,y)I(x,y)\,dx\,dy + \xi \tag{1}$$

where $I(x,y)$ the stimulus intensity at point (x,y) and

$$G(x,y) = Ae^{-\frac{y^2}{\sigma_y^2}}\left(e^{-\frac{x^2}{\sigma_1^2}} - Be^{-\frac{x^2}{\sigma_2^2}} + Ce^{-\frac{x^2}{\sigma_3^2}}\right) \tag{2}$$

the RF value at that point. ξ is a zero-mean Gaussian random variable modeling noise in the output of the oriented units and σ_y, σ_1, σ_2, σ_3, A, B, C are constants (the σ parameters control the widths of the Gaussian components and A can be thought of as a "gain" parameter).

In the published model (Weiss et al., 1993), the hidden layer consists of three oriented units with preferred orientations of $-15°$, $0°$ and $15°$. The RF function values for orientations other than $0°$ are obtained by standard rotation of coordinates:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} \tag{3}$$

Apart from the orientation-selective units, the hidden layer contains several (of the order of 100) "noise" units, modeling nearby neurons whose activity is uncorrelated with the input. This "decisional noise" is distinct from noise present in the oriented filter responses ("early noise", Weiss et al., 1993). The network output is a linear combination of the outputs of the hidden units $\mathbf{h}\cdot\mathbf{w}$, passed through a non-linear activation function (in this case a step function, *sign*( )):

$$O = sign(\mathbf{h}^{\mathrm{T}}\mathbf{w}) \tag{4}$$

$$sign(x) = \begin{cases} 1 & x \geqslant 0 \\ -1 & otherwise \end{cases} \tag{5}$$

where $\mathbf{h}$ is the hidden unit output vector (input to the decision unit) and $\mathbf{w}$ is the vector of the weights on the connections between the hidden layer and the decision unit (see Fig. 1).

2.2. Application to perceptual task

Weiss et al. (1993) tested their model on a Vernier acuity task, where the objective is to discriminate between left and right offsets of vertical Vernier lines. In the network described above, the

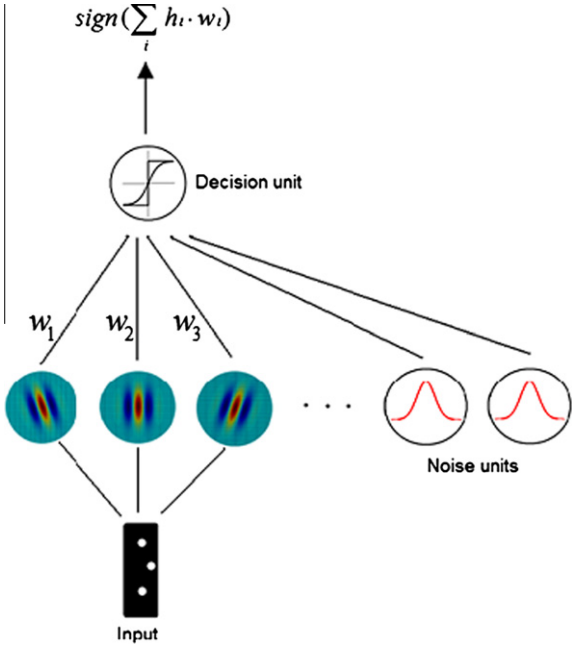

Fig. 1. HyperBF network architecture. The stimulus representation is passed through the oriented filters in the hidden layer. The responses of these filters, as well as the activities of several noise units, are weighted by the decision unit, which outputs 1/−1 for right/left offsets respectively.

unit with a preferred orientation of −15° is maximally activated by a Vernier stimulus whose top line is to the left of the bottom one (left-offset). The authors examined the response of each of the three oriented units to the presentation of Vernier stimuli with offsets ranging from −30″ to 30″ (arcsec) in steps of 3″. The oriented unit response as a function of the stimulus image is given by the noiseless version of Eq. (1) ($\xi = 0$).

For verification purposes, we performed the same test. The parameters used in this experiment are given in Table 1. It should be noted that these are the unit responses prior to learning (which, in any case, does not involve a change in the response properties of individual units in the middle layer, as we explain shortly).

Fig. 2 shows the results of our simulation, presented alongside the simulation results of Weiss et al. (1993). In both instances, it can be seen that the vertically oriented unit yields the same response for offsets of opposite sign but same magnitude, whereas the responses of the other two units are monotonically increasing (or decreasing) functions of the (signed) offset. Therefore the vertically oriented unit provides the least amount of information for solving the task.

Table 1
Model parameters for the responses of the three oriented units to a Vernier stimulus (see Section 2.2 and Fig. 2). "′" denotes minutes of arc.

Parameter	Symbol	Value	Comment
Parameter for y-width of Gaussian envelope	σ_y	19′	See Eq. (2)
First parameter for x component of DoG	σ_1	2′	See Eq. (2)
Second parameter for x component of DoG	σ_2	7′	See Eq. (2)
Third parameter for x component of DoG	σ_3	3′	See Eq. (2)
Fourth parameter for x component of DoG	A	2.95	See Eq. (2)
Fifth parameter for x component of DoG	B	0.9	See Eq. (2)
Sixth parameter for x component of DoG	C	0.2	See Eq. (2)
Vernier bar length	–	5′	
Vernier bar thickness	–	0.033′	
Vernier bar separation	–	1′	

2.3. Learning

Network training can be performed with various degrees of feedback. In general, as Weiss et al. (1993) discuss, there exists a trade-off between the amount and nature of feedback supplied and the prior information that must be encoded in the network: a network whose connection weights are initialized according to the specifics of the task being learnt must necessarily rely on feedback in order to improve. Conversely, a network that encodes in its weights prior knowledge about the task does not need feedback: as long as the weights are in the right basin of attraction, the network can improve without supervision.

In our simulations, learning occurs solely by modifying the weight vector **w**, which corresponds to the reweighting view, i.e. that learning occurs through strengthening (or weakening) of connections between representational (low-level) and decisional (high-level) areas.

Training is performed in an "online" fashion: weights are updated after the presentation of each data point (stimulus). In the supervised learning mode used throughout the simulations, weights are updated according to the Widrow-Hoff rule:

$$\mathbf{w}^{(bft+1)} = \mathbf{w}^{(t)} + \eta \mathbf{h}(Y(\mathbf{x}) - O(\mathbf{x})) \tag{6}$$

where $Y(\mathbf{x})$ and $O(\mathbf{x})$ are the desired and the actual output for stimulus **x**, respectively. η is the learning rate, typically a small constant.

Since, in the psychophysical experiments we simulated, subjects received feedback, the Widrow-Hoff rule was a natural choice of learning rule. It has been shown, however, that subjects are able

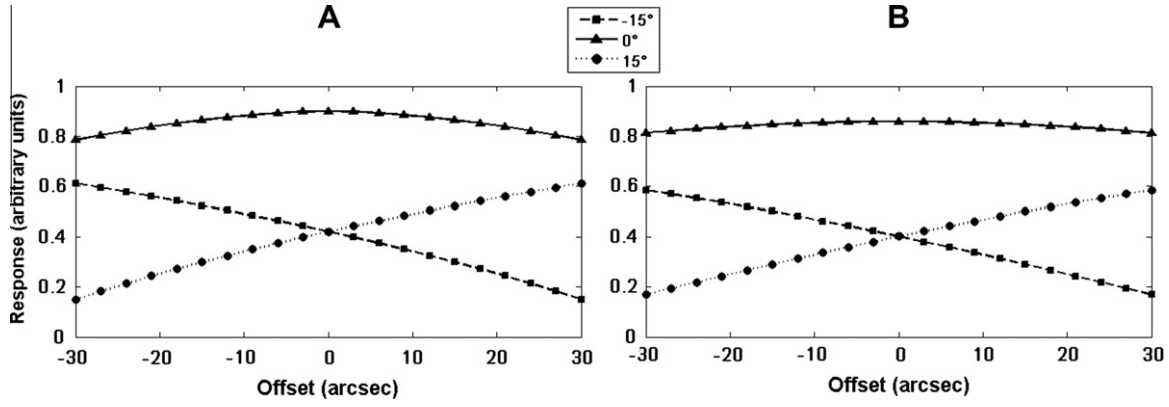

Fig. 2. Responses of the three oriented units in the model as a function of offset in a Vernier stimulus. The top curve represents the activity of the vertical (0°) unit; the crossing bottom curves the activity of the 15°- and −15°-oriented units: (A) adapted from Weiss et al. (1993): (B) our implementation. Model responses are in qualitative agreement with electrophysiological data from cat striate cortex (Swindale & Cynader, 1986).

to learn when feedback is given in a sparse manner, as a percentage of correct responses in a block of presentations, with no indication as to which responses in the block were the correct ones (block feedback) (Herzog & Fahle, 1997). Learning is even possible with total absence of feedback (Fendick & Westheimer, 1983; McKee & Westheimer, 1978), although at a slower rate (Fahle, Edelman, & Poggio, 1995). These results would indicate that feedback is not important; however, one of the same studies that shows the effect of partial feedback also shows that when subjects receive *uncorrelated* (that is, random) feedback, their performance drops dramatically: they show virtually no improvement from baseline (Herzog & Fahle, 1997). Thus it seems that subjects can solve perceptual problems even in the absence of feedback but when feedback is present, it is used.

Furthermore, as Weiss et al. (1993) show via modelling, a self-supervised rule, where the teacher signal $Y(\mathbf{x})$ is provided only for the larger (and thus easier) offsets, can be just as effective as a supervised one. According to the authors, this self-supervised rule "*was designed to simulate the conditions in psychophysical experiments in which subjects receive no feedback at all, but nevertheless possess a clear indication of the correctness of their response for the large values of Vernier offset when the stimulus looks trivially easy. Under these conditions, the subjects' thresholds improve with practice, albeit at a slower rate than when explicit feedback is available* (Fahle & Edelman, 1993)". This view sits comfortably with the Reverse Hierarchy theory (Ahissar & Hochstein, 2004): easier instances of the task (for example, trials with large offsets) can be solved at higher-level processing stages that require less precision. This can lead to feedback that helps train low-level processing stages. Mato and Sompolinsky (1996) used the same self-supervised approach to initialize the weights in their network. Petrov et al. (2005) also use a supervised rule to update the weights in their model, mentioning that it was a natural choice since their data was collected with feedback but pointing out that "*self-supervised Hebbian learning drives the weights to the same optimal solution in the long run*". Our simulations with self-supervision (not shown here) also confirmed this: the network converged to the correct solution, although somewhat more slowly.

2.4. Changes to the model

The original model of Weiss et al. (1993) was designed with simplicity in mind. The authors' intention was to adopt this simple model "*as a minimalist platform for studying the improvement of hyperacuity with practice*" (Weiss et al., 1993). While we tried to maintain simplicity, certain aspects of the model had to be modified, both in order to render the model more plausible biologically and to enable direct comparison with existing psychophysical data. The modifications were as follows:

- We converted the oriented filter response to a firing rate by passing it through a static nonlinearity which ensures saturation and nonnegativity of the output. The mean firing rate of an oriented filter with response h is given by

$$\hat{r} = r_{max}[\tanh(g(h - h_0))]_+ \tag{7}$$

where $r_{\max}$ is the maximum possible firing rate, h_0 is the threshold value that h must have for the neuron to fire, g is a constant that determines how quickly the firing rate increases as a function of h and $[\,\cdot\,]_+$ is the half-wave rectification operation ($[x]_+ = max(0,x)$).

- Additive noise in the oriented unit response – known as "early noise" (Weiss et al., 1993) or "representation noise" (Petrov et al., 2005) – is replaced by multiplicative noise, where the trial-to-trial firing rate is given by

$$r = \hat{r} + \sqrt{\hat{r}}\xi \tag{8}$$

where ξ is a random number drawn from a Gaussian distribution of zero mean and unit variance ($\mathcal{N}(0,1)$). The use of $\sqrt{\hat{r}}$ for the standard deviation of the Gaussian noise reflects the experimentally determined fact that the Fano factor (the variance/mean ratio) for neuronal responses is close to 1, i.e. mean and variance are often approximately equal (Dayan & Abbott, 2001, chap. 1.4, pp. 30–31).

- Following Jones and Palmer (1987) and Ringach (2002), we replaced the difference-of-Gaussians RF function used by Weiss et al. (1993) with a Gabor:

$$G(x,y) = e^{-\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)}\cos(2\pi f x + \varphi) \tag{9}$$

Gabors fit physiological data well and contain fewer parameters which determine the RF structure in a more natural way. σ_x and σ_y control the width of the 2D Gaussian envelope along the x and y axes, respectively; f and φ are the spatial frequency and phase of the sinusoidal component, respectively.

- *Baseline performance modeling.* In the Weiss et al. (1993) implementation, no prior knowledge is encoded in the network and the weights between the hidden and output layer are initialized randomly. Thus the model's performance during the first few trials of any perceptual task is near chance, regardless of the difficulty of the task. However, naive human subjects exhibit baseline performance, which is above chance to a degree proportional to the easiness of the task. This above-chance performance may be explained by the Reverse Hierarchy theory (Ahissar & Hochstein, 2004), mentioned earlier: when the perceptual problem is easy, higher-level cortical areas are able to solve it, without the need for long-term training. To allow for quantitative comparisons with experimental data, we incorporated baseline performance levels probabilistically: with probability p, the actual output $O(\mathbf{x})$ of the decision unit is set to the desired output $Y(\mathbf{x})$. From a neurobiological perspective, $Y(\mathbf{x})$ corresponds to the answer given by higher-level areas. This is not always available or correct; it is so only with probability p, which is inversely proportional to the difficulty of the task performed. This is quantified in different ways, depending on the experiment. When simulating psychophysical experiments, the baseline performance observed in subjects is used to estimate p (see Appendix A):

$$p = 2F(e) - 1 \tag{10}$$

where e is a measure of the "easiness" of the problem being solved (in a Vernier experiment, this could be the stimulus offset) and F the function that, for each e, yields the observed baseline success rate. F can be approximated from psychophysical data.

- Nonzero spatial phases of varying magnitude were incorporated to the model (setting φ in Eq. (9) to a nonzero value). This feature resembles the ability for position-invariant judgments, and apart from its biological fidelity, it proved a necessary addition, as the network was unable to solve certain tasks (such as dot alignment) using only zero-phase units.
- A wider range of preferred orientations for the oriented filters (from $-90°$ to $90°$), modeling overcomplete representation – a common property of population codes (Dayan & Abbott, 2001).

3. Simulations

To evaluate the expanded model, we simulated a number of psychophysical experiments from the perceptual learning literature: one studying the dependence of performance on stimulus attributes (Westheimer & McKee, 1977), one studying disruption of perceptual learning under various conditions (Seitz et al.,

2005), one studying transfer of learning across task and stimulus configurations (Webb et al., 2007) and one studying transfer of learning when the target stimulus remains the same while the context changes (Petrov et al., 2005). To our knowledge, no single model has simulated all of these experiments; furthermore, the results of Seitz et al. (2005) and Webb et al. (2007) have never been modeled.

3.1. Stimulus attributes

Here we examine the dependence of the offset threshold (defined as the offset that is detectable 75% of the time) to both the length and the separation of the lines in a Vernier offset discrimination task.

3.1.1. Methods

Following Westheimer and McKee (1977), the network was trained on a Vernier offset discrimination task using the method of constant stimuli. There were eight sessions of 80 trials each. In each trial, a Vernier stimulus with either positive or negative offset was presented, the network reported the offset sign and received feedback on the correctness of the choice. On the last session, network performance was recorded. Offsets ranged from $-30''$ to $30''$ (arcsec) in steps of $3''$.

The first set of simulations examined the dependence of the post-training offset detection threshold on the length of the line elements of a Vernier stimulus. Each simulation of this set was performed with a different value for the length (ranging from $1'$ to $20'$ (arcmin) in steps of $1'$), while the separation of the line elements was kept constant (at $0'$ in both the data and the model). Data is from two subjects. The second set of simulations was completely analogous, except the independent variable was the separation (ranging from $0'$ to $10'$ in steps of $1'$), with the length kept constant (at $4'$ and $8'$ in both the data and the model). In both cases, data is from two subjects. In Fig. 3B, data for each length is from a different subject. Model parameters for these simulations are listed in Table 1.

3.1.2. Results and discussion

Fig. 3 shows the results of both Weiss et al.'s (1993) and our simulation as well as experimental data from Westheimer and McKee (1977). In both the data and the models, threshold is an asymptotically decreasing function of the length of the lines that make up the stimulus: increasing the length of very short lines makes the task easier but only up to a length of about $4'$ – further increase has no effect. Threshold is also an increasing function of line separation: it is harder to judge alignment of line segments that are far apart. Weiss et al. (1993) note that while their length–threshold curve agrees reasonably well with the psychophysical data, the separation–threshold curve agrees only qualitatively. In our implementation, the curves in both graphs agree reasonably well with the data.

The fact that, following the aforementioned changes, the model behaves similarly to the published implementation is indicative of a desirable property of the model: a relative insensitivity to details. Our changes to the published model simply refined it to enable it to model perceptual learning in a variety of conditions and in a more realistic and plausible way while retaining, and even improving, the model's ability to simulate the Westheimer and McKee (1977) data set.

3.2. Disruption of learning

Here we simulate the experiment of Seitz et al. (2005), who investigated how learning of a 3-dot hyperacuity task was disrupted by subsequent training on the same task with a slightly different stimulus.

3.2.1. Methods

In Seitz et al. (2005), subjects conducted a 2-interval-forced-choice task (2IFC). Two pairs of images were displayed in succession; one image of the pair depicted three dots aligned vertically whereas in the other image the central dot was slightly offset to one side and subjects reported whether the middle dot was offset in the first or in the second image. The experiment lasted 5 days. Each day subjects conducted either one or two sessions, depending on the task condition. Each session consisted of 400 trials (80 trials per offset size). Five different offset sizes were tested: $0.9'$, $1.8'$, $2.7'$, $3.6'$, $4.5'$ (arcmin). In each session, the stimuli were presented in random order with respect to offset size. The task conditions that we examine here were as follows:

- In condition **A**, there was a single session per day. All offset stimuli in this session had right offset.
- In condition **AB**, there were two sessions per day. The first session contained solely offsets to one side (right or left) whereas the second session contained solely offset to the other side (left or right).
- In condition **20RI**, trials were interleaved between between vertically oriented stimuli and horizontally oriented stimuli.

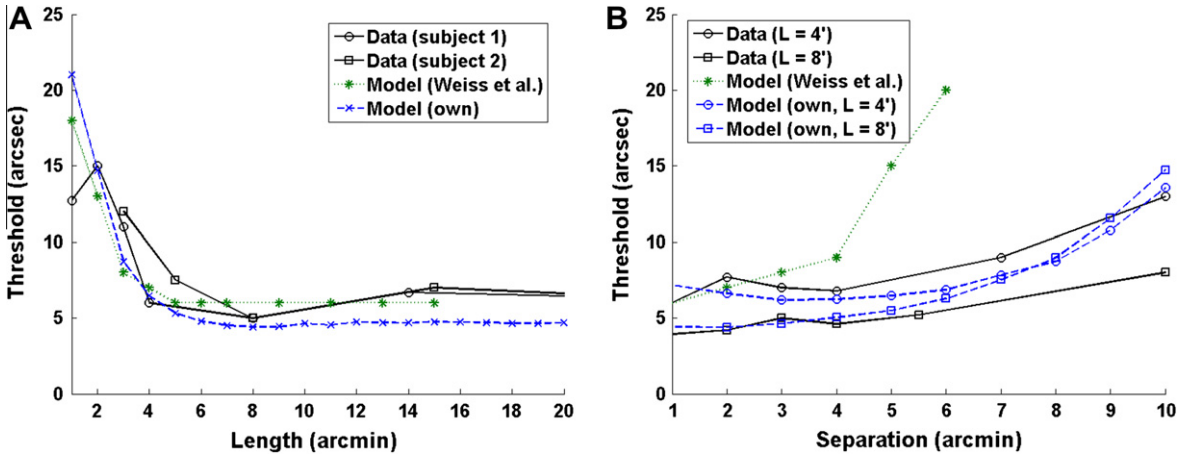

Fig. 3. Dependence of offset discrimination threshold on line length and line separation in a Vernier task, after training: (A) data is from two different subjects. Separation of line elements is zero for both subjects; (B) data is from two different subjects, each tested with a different length of the line elements (first subject: $4'$; second subject: $8'$). All data replotted from Westheimer and McKee (1977).

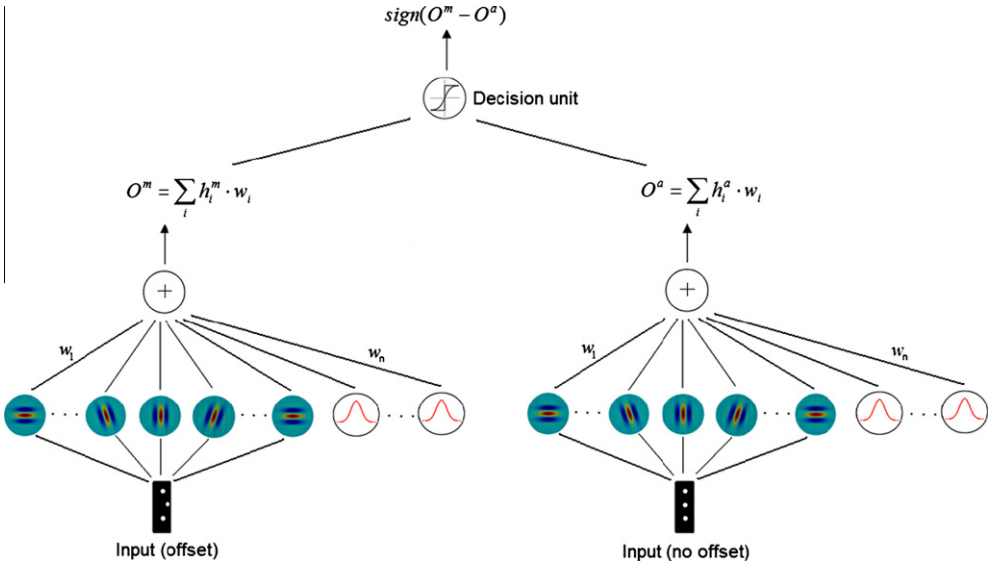

Fig. 4. Network architecture for the simulation of the experiment of Seitz et al. (2005). Note that the same network is used for both the aligned and the misaligned stimulus; the two responses are then compared by the decision unit so as to simulate the 2IFC design of Seitz et al. (2005).

In all conditions, after each individual presentation of the stimulus pair, subjects where given feedback on the correctness of their response. The objective of this experiment was to see if, and under what circumstances, training on a second stimulus disrupts learning on a first stimulus.

To simulate this 2IFC experiment, the output layer (decision unit) of the model had to be modified. The subjects were asked to judge the position (index) of the misaligned stimulus in the pair; a natural way to encode this in the model is for the decision unit to output 1 if the first stimulus is the misaligned one and −1 otherwise. Thus, the decision unit must compare the responses from the two stimuli. This is accomplished by setting the output O of the decision unit to the sign of the difference between the two responses:

$$O = sign((\mathbf{h}^m - \mathbf{h}^a)^{\mathrm{T}}\mathbf{w}) \tag{11}$$

where $\mathbf{h}^a$ and $\mathbf{h}^m$ are the outputs of the hidden layer for the aligned and the misaligned stimulus, respectively. Fig. 4 shows the corresponding network architecture.

The parameters used in the model are given in Table 2.

3.2.2. Results and discussion

Fig. 5 shows the results of the simulation, which largely agree with the experimental data: in Condition AB, the curves for Day 1 and Day 5 are closer together than the respective curves in Conditions A and 2ORI; moreover, the Day 1 curve in Condition AB is lower than the Day 1 curves in the other two conditions, especially for the toughest trials (small offset). This means that learning of task A was disrupted by learning of the same-orientation, opposite offset side, stimulus B but not by learning of the different-orientation stimulus B. In the model this happens as follows: without loss of generality, suppose that in each trial (regardless of session), it is always the first stimulus in the pair that is offset (with the zero-offset stimulus appearing second). In this case, the network output must be 1. Thus in Condition AB, the network must respond with 1 in both sessions (A and B). However, the individual responses of most of the oriented units are different for left and right offsets: due to the nonzero-phase RF in most units, their response within each session is a monotonic (increasing or decreasing) function of (signed) offset (cf. Fig. 2, bottom crossing curves). Since the response vector of the hidden layer has a different direction in the two sessions, it follows that the weight vector must also be different if the network is to always output 1. Therefore training in session B shifts the weight vector to a direction different from that of session A, disrupting learning of the latter.[1]

There is another subtle but important point to be discussed. Seitz et al. (2005) point out that disruption occurred only in task A in Condition AB, whereas subjects were able to learn task B. This is not the case in our simulation because our model treats tasks A and B equivalently. It is not clear how the asymmetry in the (Seitz et al., 2005) experiment comes about, but it is very likely that time is a factor. By performing task B last each day, learning for this task consolidates during the following hours. Thus the next day, training on task A does not disrupt the learning that occurred the previous day on task B. A possible mechanism for this is that consolidation may cause the two tasks to be handled by separate readout or representation modules (see Petrov et al. (2005) for a more in-depth discussion on the possible multiplicities of readouts and representations). An interesting expansion of the model would be to implement a mechanism for multiple readouts so as to address data consolidation (Karni, Tanne, Rubenstein, Askenasy, & Sagi, 1994; Mednick, Nakayama, & Stickgold, 2003; Seitz et al., 2005) as well as related phenomena such as stimulus tagging (Zhang et al., 2008).

3.3. Specificity of learning: tasks and configurations

The next simulation examined an interesting finding by Webb et al. (2007), where learning was found to generalize across tasks and stimulus configurations but not to other stimulus configurations within the same task.

3.3.1. Methods

Fig. 6A is a diagram of the scenarios Webb et al. (2007) explored. There were two different tasks: an alignment task (as in the previous section) and a bisection task (where subjects are asked to determine whether the central dot lies in the midpoint between the outer dots or whether it is closer to one or the other outer dot). Subjects were trained on one task using a particular stimulus orientation

[1] While the zero-phase units have a response that is symmetrical about zero offset (cf. top curve in Fig. 2), the response gradient of those units is too shallow around the peak of the RF and thus those units are not reliable offset discriminators in the presence of noise.

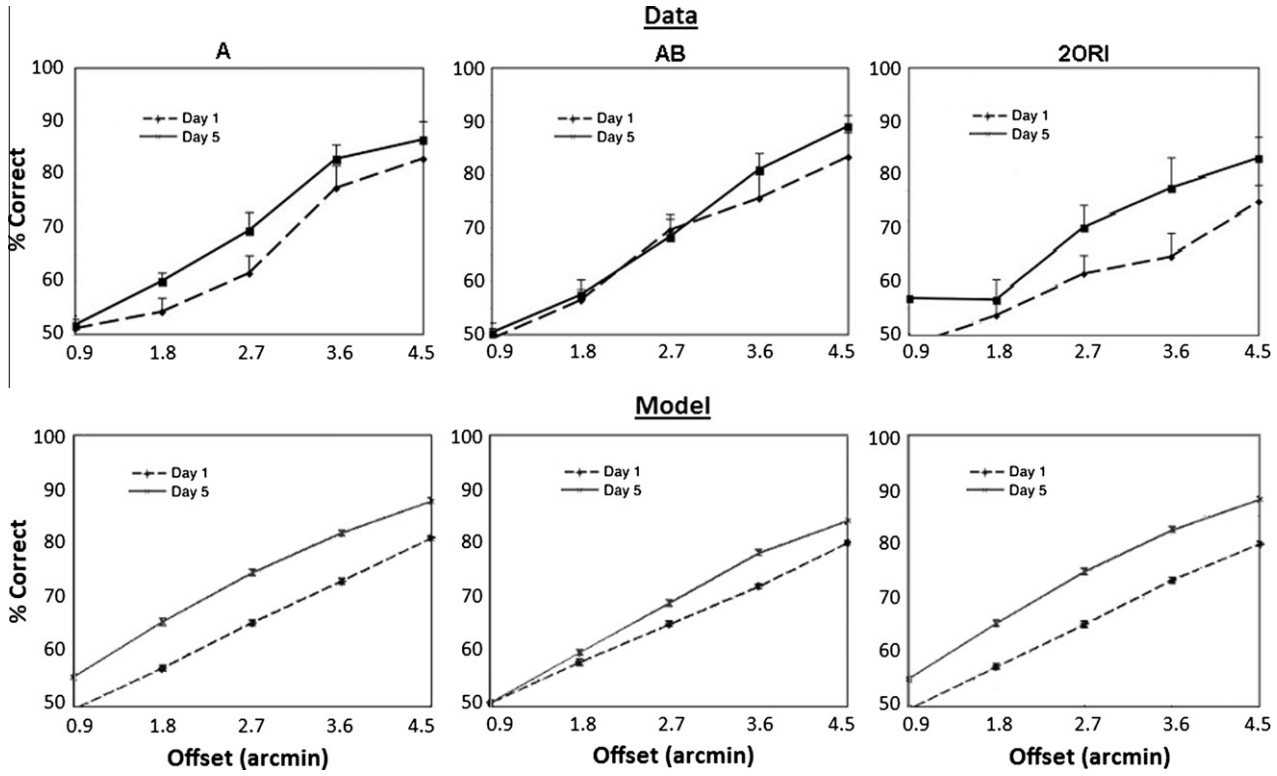

Fig. 5. Results of the experiment of Seitz et al. (2005). Percentage of correct responses in the dot alignment task plotted as a function of stimulus offset during the first and last day. Top row: psychophysical data (Seitz et al., 2005); bottom row: simulation. The columns correspond to the three conditions examined by Seitz et al. (2005) (see Section 3.2.1). Error bars are SEM.

(e.g. alignment-vertical); they were subsequently tested on the other three task/stimulus configurations (e.g. bisection-vertical, bisection-horizontal, alignment-horizontal). The purpose of this experiment was to see how learning generalizes across task and stimulus configurations. Here, the performance metric is the offset threshold, unlike the previous experiment, where the metric was percentage of correct responses for each offset.

The simulation involved eight training sessions, each consisting of 80 presentations on a particular task and stimulus configuration. At the end of the 8th session, the network was tested on the remaining tasks and stimulus configurations. The tasks and stimulus configurations were the same as in Webb et al. (2007).

On a sidenote, one of the variables in the experiment of Webb et al. (2007) is the distance between the outer stimulus elements, as the authors studied how the separation of the reference elements affects the magnitude of learning. This potentially interesting question cannot be addressed by the model. Webb et al. (2007) propose that as separation increases, larger RFs are recruited, which leaves more scope for moving to smaller spatial scales of analysis during training. However, our model contains units of fixed RF size and thus cannot be expected to accommodate stimuli at spatial scales as diverse as those used by Webb et al. (2007), who use separations from 5° up to 30° of visual angle.

3.3.2. Results and discussion

The results of Webb et al. (2007), as well as our simulation, are summarized in Fig. 6. In both the experiment and the simulation, learning failed to transfer across tasks (dash-dotted line in Fig. 6) or orientations alone (dotted line in Fig. 6), but it did transfer across tasks *and* orientations (diagonal dashed line in Fig. 6). This result challenges the traditional notion that learning is specific to stimulus and task configuration. Learning of a vertical dot alignment task does not transfer to the horizontal configuration of the same task or to the bisection task of the same orientation; it does, however, transfer to the horizontal bisection task. This can be explained by the common spatial axis of the stimulus variability between the conditions that show transfer of learning.

In the model, such generalization comes about naturally. In a horizontal alignment (HA) and a vertical bisection (VB) stimulus, the only variable element (middle dot) occupies the exact same positions in both stimuli, for each offset. The reference elements (outer dots) occupy cardinal positions, but these only serve as positional cues and their contribution to the activation of the oriented units is constant, i.e. does not depend on offset magnitude. Thus the relation between offset and activation level of the oriented units is identical in HA and VB and therefore a network that has optimal weights for HA is also optimized for VB (and vice versa). In other words, to the model, the desired input/output mapping corresponding to the vertical alignment and the horizontal bisection tasks (and associated stimuli) are identical.

3.4. Contextual effects on learning

We now turn to the experiment of Petrov et al. (2005), who examined whether learning of a discrimination task is specific to contextual elements of the visual scene (as opposed to the target stimulus) and, more critically, whether training in one particular context interferes with learning of the same task in a different context. The authors also proposed a model based on "selective reweighting" that can account for their experimental findings. Here we simulate the same experiments.

3.4.1. Methods

The experiment of Petrov et al. (2005) involved a binary (left/right) orientation discrimination task. Each trial consisted of a

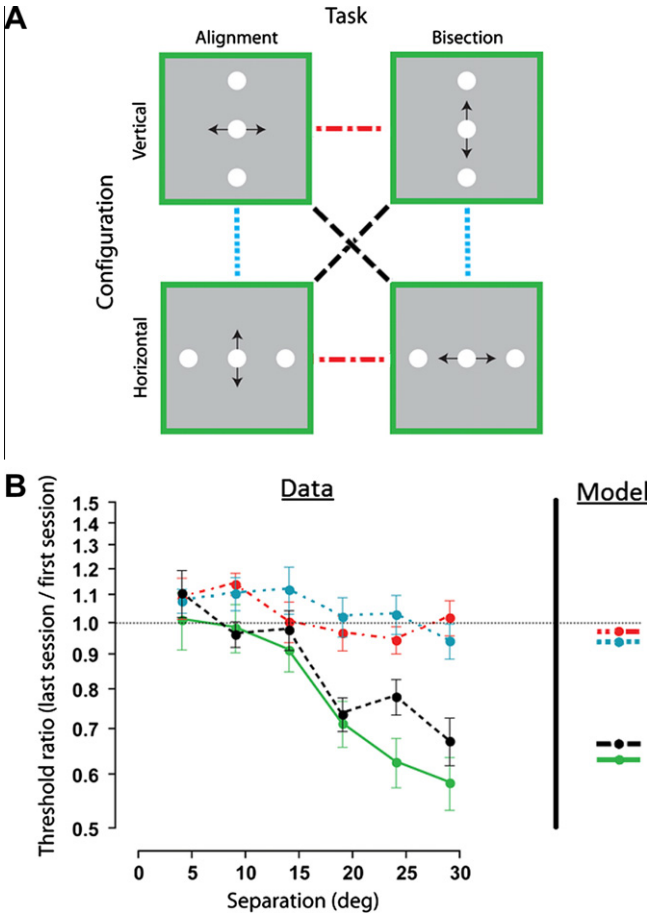

Fig. 6. (A) Decoupling the task and stimulus specificity of positional learning. In each case, the black arrow indicates the spatial axis of the required positional judgment. Adapted from Webb et al. (2007). (B) Transfer of learned improvements between task and stimulus configurations. Ratio of mean positional thresholds obtained on the first and last days, plotted as a function of stimulus separation. Dotted line: training on one task (e.g. alignment) and orientation, testing on the same task and the perpendicular orientation; dot-dashed line: training on one task (e.g. alignment) and orientation, testing on the other task (e.g. bisection) and the same orientation; dashed line: training on one task an orientation, testing on the other task and the perpendicular orientation. For comparison, the solid line shows averaged learned improvements obtained within each task-configuration combination (training and testing on the same task and orientation). Left part of B (labeled "Data") adapted from Webb et al. (2007). Error bars are SEM. Unlike in Webb et al. (2007), threshold ratios in the model were obtained for a single value of separation.

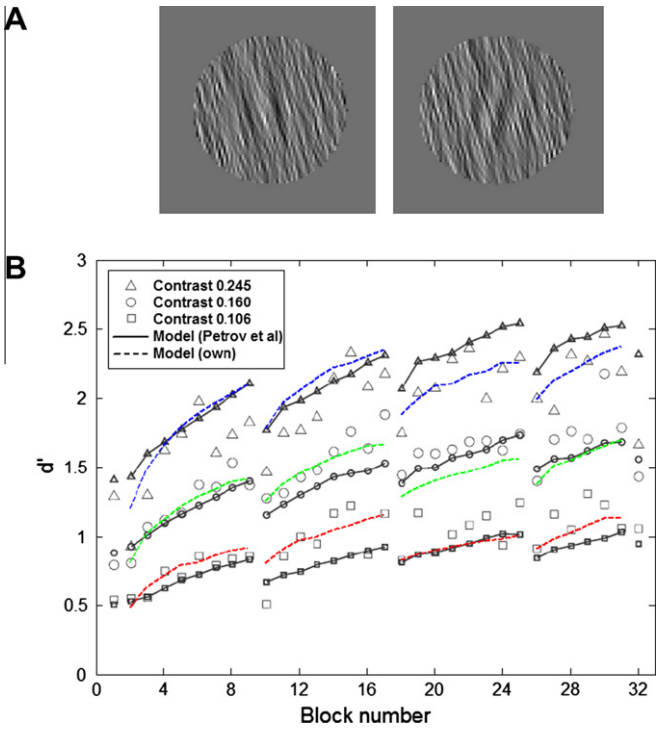

Fig. 7. (A) Stimulus examples used in the orientation discrimination experiment of Petrov et al. (2005). Congruent (left patch) and incongruent (right patch) stimuli are embedded in a noisy context oriented to the left. (B) Performance as a function of time. Sensitivity index (d') is plotted against block number, for three contrast levels. Context switches occur on blocks 2, 10, 18, 26 and 32.

presentation of a single stimulus, followed by a left/right response by the subject. The target stimulus was a windowed sinusoidal grating (Gabor patch) embedded in a circular field of Gaussian noise (referred to as *context*) that was filtered to form oriented textures (Fig. 7A). The Gabor patch had an orientation of either $-10°$ or $10°$ whereas the filtered noise context had an orientation of either $-15°$ or $15°$. In trials where the Gabor patch had an orientation of the same sign as the context, the stimulus is said to be *congruent* (Fig. 7A, left). There were 32 blocks of 300 trials, performed over a period of 8 days (four blocks per day). Within each block, the context orientation was kept fixed in each trial whereas the target stimulus varied randomly and equiprobably between the two orientations. The order of presentation of the 32 blocks during the course of the experiment was L–8R–8L–8R–6L–R, where L (R) denotes a block with exclusively left (right)-oriented noise, i.e. a block of trials of left-oriented noise was followed by eight blocks of right-oriented noise and so on. The transition between a block of a given noise orientation and a block of the opposite orientation is termed a *context switch*.

If learning is specific to context, then any improvements gained by training with blocks of a particular context orientation would not transfer to blocks of the opposite orientation. In that case, task performance, which should be monotonically increasing within a block as subjects become better at solving the task, would show a sudden drop after a context switch. Furthermore, if learning in both contexts occurs through changes in the same underlying neural circuitry, performance immediately after a context switch should be lower than performance immediately before the *previous* switch. In other words, learning in a certain context (e.g. L) should be disrupted by the intervening blocks of a different context (e.g. R), so that the network has to (at least partially) re-learn the task when it switches back to the original (L) context. Petrov et al. (2005) refer to this scenario as the *switch cost* hypothesis.

All simulation parameters were set to their experimental counterparts when applicable. Almost all model parameters were as in previous simulations and are given in Table 2.

3.4.2. Results and discussion

Fig. 7B shows the performance of human subjects as well as the two models – that of Petrov et al. (2005) and the present. Both the data and the models confirm the hypothesized specificity of learning with respect to context. Context switches, which occur on blocks 10, 18, 26 and 32, are accompanied by a significant performance drop. This result highlights the highly specific nature of perceptual learning: it is specific not only to the task or the target stimulus but also on contextual and unattended features of the visual scene.

Moreover, the switch cost hypothesis is confirmed by observing that performance immediately after a switch (e.g. block 18) is significantly lower than performance immediately before the previous switch (block 9 respectively), even though the context in these two blocks is the same. This shows that learning in a

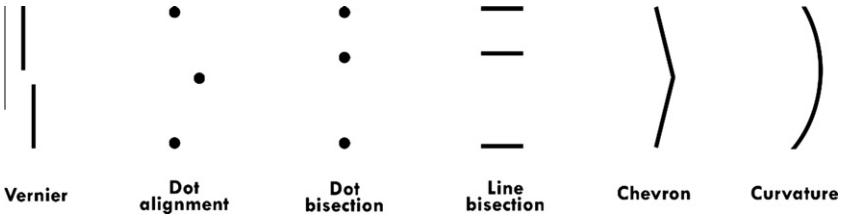

Fig. 8. Tasks and stimuli used in simulations. In the case of line bisection, the orientation of the stimulus is defined as in dot bisection, i.e. relative to the axis passing through the midpoints of the three elements. For the Chevron and curvature stimuli, the variable element is the angle between the line segments and the degree of curvature, respectively.

particular context can indeed be disrupted by intervening training on a different context.

Both models agree reasonably well both with the main experimental results and with each other, exhibiting the aforementioned specificity and interference (disruption) effects. A rather subtler aspect of the observed disruption, that is present in both the data (most easily seen in Fig. 4 of Petrov et al. (2005), which includes regression curves) and the two models, is that it is partial. When the network switches to a certain context, after an intervening sequence of different-context blocks, it does not learn from scratch: some improvement from previous exposure to the same context is retained. In Fig. 7B, this can be seen from the fact that performance in block 18 is higher than performance in block 2 (and the same holds for blocks 26 and 10, respectively), that is, performance immediately after a context switch is higher than performance immediately after the last switch to that context (which occurred 8 + 8 = 16 blocks ago).

Despite the agreement of our model with the average data, when performance is considered separately for congruent and incongruent trials, our model matches the incongruent data only qualitatively; for the congruent data, our model fails to replicate the small but statistically significant contrast reversal effect present in the congruent data – namely that performance drops slightly as contrast increases (Petrov et al., 2005, Fig. 9). The authors, who show that their model can account for this reversal, albeit within a narrow range of model parameters, state that "*the congruence effects reflect compressive nonlinearities in the representation subsystem*". Compressive nonlinearities are also present in our model but do not appear to be able to account for the congruence effect: increasing the gain (g in Eq. (7)), in order to saturate the responses of the oriented filters, did not change the results qualitatively. It will be interesting in future work to determine what minimal model could explain these effects.

As mentioned before, most parameters used in this simulation were the same as in previous simulations. In particular, we found that the model yielded very similar results under a wide range of parameter values. The only parameter that proved critical is the spatial frequency of the RF function used in our model (f in Eq. (9)), which had to be set to values close to the spatial frequency of the stimulus itself (which is generated by a Gabor function).

3.5. Predictions

Apart from simulating existing psychophysical experiments, the model was subjected to a variety of tasks and stimuli and generated a rich set of results. Our main question was whether there is a correlation between generalization and disruption; our hypothesis was that if a pair of tasks (each involving a particular stimulus, orientation and judgment type) shows high disruption under the conditions of the Seitz et al. (2005) experiments, it would also show high transfer under the conditions of the Webb et al. (2007) experiments.

3.5.1. Methods

The tasks and stimuli used are shown in Fig. 8. There were two sets of experiments, one studying disruption and another studying specificity. In each set, all combinations of tasks (six in total) and stimulus orientations (horizontal or vertical) were tested, totaling 42 cases.

In the disruption experiments, the conditions are the same as in Seitz et al. (2005): there are two 2IFC tasks (A and B), which can differ in one or more of the following:

- *Stimulus type:* Vernier (Poggio et al., 1992; Westheimer, 1976; Westheimer & McKee, 1977), dots (Fahle & Morgan, 1996; Seitz et al., 2005), parallel lines Crist, Kapadia, Westheimer, and Gilbert (1997), Li, Piëch, and Gilbert (2004), Tartaglia et al. (2009), Chevron (Aberg et al., 2009; Kramer & Fahle, 1996) or curve (Fahle, 1997; Kramer & Fahle, 1996);
- *Judgment type:* alignment, bisection or curvature;
- *Stimulus orientation:* horizontal or vertical.

In the specificity/transfer experiments, the conditions are the same as in Webb et al. (2007): there are two tasks (each trial consisting of a single stimulus presentation); the first (A) is the task that the network is trained on and the second (B) is the task that it is tested on.

At this point, it should be emphasized that even though both the disruption experiment of Seitz et al. (2005) and the transfer experiment of Webb et al. (2007) involve a binary decision task, the alternatives are different in the two experiments: the judgment in Webb et al. (2007) is the sign of the perceived offset of the middle element from the midpoint ("which side?") whereas the judgment in Seitz et al. (2005) is the order of presentation of a control (middle element at the midpoint) and an offset (middle element offset to one side) stimulus ("which stimulus?"). Furthermore, in Seitz et al. (2005) subjects in each session were exposed to stimuli offset to one side only (and to a control stimulus of zero offset), whereas in Webb et al. (2007) a single session included presentations of stimuli offset to either side. This observation will help explain the seemingly peculiar correlation between transfer and disruption, discussed below.

The degree of disruption (D) and the degree of transfer (T) between the tasks in the pair is determined for each of the 42 cases. Each case can thus be thought of as a point in the disruption-transfer space with coordinates (D_i, T_i) for $1 \leqslant i \leqslant 42$. D is percentage of performance drop from control (control is training only on the first condition of each pair). Negative values for D denote *facilitation*, that is, training on task B enhances previous improvements on task A. T_i is percentage of transfer of learning, when the network is trained on the first condition of the pair and evaluated on the second condition, and is defined as

$$T = 100 \times \frac{\theta_{pre}^{untrained} - \theta_{post}^{untrained}}{\theta_{pre}^{trained} - \theta_{post}^{trained}} \tag{12}$$

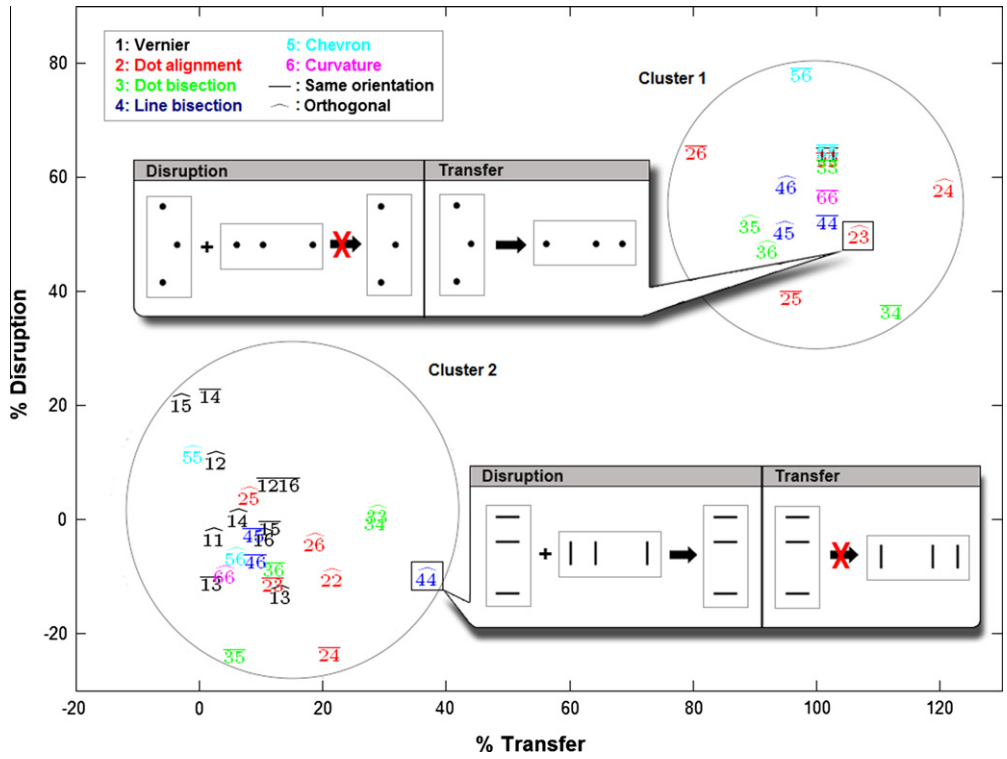

Fig. 9. Summary of all results on specificity and disruption generated by the model. Each data point represents one of the 42 cases and consists of two numbers and an overhead symbol, which is either a hat or an overline. Each number, in the range 1–6, corresponds to a task according to the order in Fig. 8. In the disruption simulations, the two numbers correspond to the tasks that were practiced in succession. As in Seitz et al. (2005), the stimulus offset sign in the second session is opposite of the sign in the first session (for the nonzero-offset stimulus in each trial). In the specificity simulations, the first number corresponds to the trained task, whereas the second number to the tested task (which is not trained), as in the Webb et al. (2007) experiment. In both kinds of simulations, a hat above the two numbers means the corresponding stimuli have orthogonal orientations, whereas an overline denotes same orientation. The floating insets depict, for two representative points, the associated stimuli used in the disruption and generalization simulations as well as qualitative (presence/absence of learning) results. In the transfer simulation, point $\widehat{23}$ corresponds to training on vertical dot alignment and testing on horizontal dot bisection. In both simulations, an arrow indicates learning: the trained tasks/stimuli are to the left of the arrow and the tested task/stimulus is to the right (also see Section 3.5.2).

where the numerator is the threshold difference for task B before and after training on task A, and the denominator is the before-after threshold difference for task A itself (control). When tasks A and B are the same, i.e. involve the same stimuli with the same orientation, T is trivially expected to be 100%.

3.5.2. Results and discussion

The results are summarized in Fig. 9. Most data points belong to two clusters, one centered around high (60%) disruption and high (100%) transfer (Cluster 1) and another centered around zero disruption and low (10–20%) transfer (Cluster 1). Cluster 1 contains exclusively same-task, same-orientation pairs, whereas Cluster 2 contains mainly different-task, different-orientation pairs.

3.5.3. Cluster 1

Points in Cluster 1 are of three types:

- $\overline{xx}$: there are six points of this type. In the disruption simulation, these points correspond to Condition AB in Seitz et al. (2005), i.e. tasks A and B are of the same type (dot alignment in Seitz et al., 2005) and involve stimuli with the same orientation but opposite-signed offsets. Therefore these points show high disruption. In the transfer simulation, these points correspond to the control condition in Webb et al. (2007), i.e. training and testing on the same task and orientation. Since the percentage of transfer was defined with respect to this control condition (see Eq. (12)), these points are on the 100% transfer mark.
- $\overline{xy}$ (with $x \neq y$): there are four points of this type, corresponding to conditions Chevron–Curvature, Dot alignment–Chevron, Dot alignment–Curvature and Dot bisection–Line bisection. Chevron and curvature stimuli are similar to each other, both in terms of appearance and in terms of how the stimulus changes as offset increases (both stimuli start from straight lines and progressively "bend" towards the same direction as offset increases). There is also similarity between dot alignment and Chevron/curvature, since the former essentially becomes the latter if the dots are connected with lines/curves.
- $\widehat{xy}$: there are six points of this type. $\widehat{23}$ is exactly the condition of the Webb et al. (2007) experiment, where significant transfer occurs. The rest of the points are similar to $\widehat{23}$ (especially $\widehat{24}$, whose only difference is that the bisection task includes line instead of dot stimuli) in the above sense; that is, as offset increases, the stimulus body moves towards the same direction and along the same spatial axes in the two tasks. In the less obvious case of $\widehat{35}$, the first task (dot bisection) is similar to the second (Chevron, orthogonally oriented relative to the first): in both tasks, the stimulus changes along the Y axis, its surface moving outwards from the center as offset increases.

The fact that pairs exhibiting high transfer also exhibit high (and not low) disruption may seem paradoxical but is explained by the different conditions in the two simulations, as pointed out in Section 3.5.1:

- In the disruption simulation, which is a generalization of the experiment of Seitz et al. (2005), all trials in the first task involve a right-offset stimulus and a zero-offset one, whereas in the other task of the pair, there are left-offset and zero-offset stimuli. Therefore, if in a particular trial of task A the control stimulus is presented first, the network must output 1 (otherwise −1), i.e.

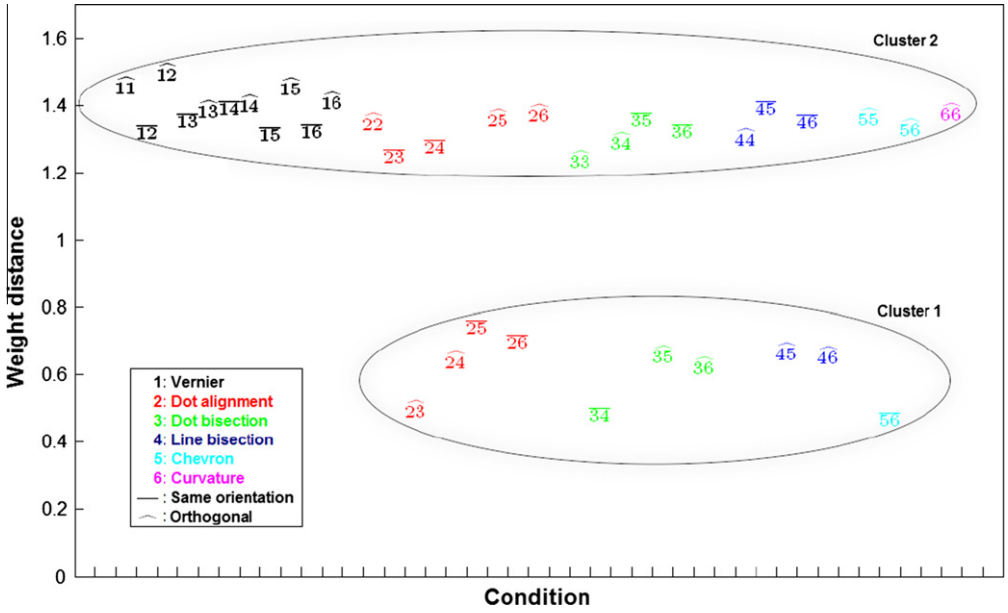

Fig. 10. Weight distances. Each point corresponds to a specificity simulation, as detailed in Fig. 9.Y axis: Euclidean distance of the (normalized) weight vectors of the two tasks in the pair at the end of the simulation.

1 is the desired response when the middle element (e.g. dot) in the second stimulus is to the *right* of the middle element in the first stimulus. In task B, however, the situation is reversed: the network must output 1 if the middle element in the second stimulus is to the *left* of the the middle element in the first stimulus. The result is that if the two tasks are similar (or the same, e.g. both dot alignment), the learned weight vector for task A has the opposite direction of the vector for task B. In other words, task similarity (i.e. points $\overline{xy}$ where x and y are similar or the same) entails disruption.

- In the transfer simulation, task A corresponds to training, whereas task B corresponds to testing. In both tasks, there is a single stimulus in each trial, ranging over both positive and negative offsets, as in Webb et al. (2007). Therefore if tasks A and B are similar, the learned weights for task A are also suitable for solving task B, i.e. task similarity entails transfer.

3.5.4. Cluster 2

Points in Cluster 2 represent pairs that either consist of different tasks or of different orientations (or both) and thus exhibit independence: there is neither disruption following sequential training nor transfer of learning from one to the other. Interestingly, however, relatively few points show near-zero transfer; Cluster 2 is centered at around 10%. That is, for most pairs, there is at least some small amount of transfer, suggesting that the network became slightly better at task B after training on task A by using any similarities, however minute, between the two tasks while ignoring noise (that is, the noisy units are assigned near-zero weights). A curious case is point $\widehat{44}$ (line bisection), which has the greatest transfer percentage (35%) in the cluster. This may have to do with the fact, unique to this case, that the spatial axis of stimulus variability (Y axis in the case of vertically oriented bisection stimuli) is orthogonal to the axis to which the bars are aligned (X axis). In other words, the orientation of the entire stimulus is perpendicular to the orientation of the bars that compose the stimulus. This does not apply to dot bisection because the dots do not have an orientation – they are radially symmetric.

At the end of the above simulations, we also measured, for each point, the distances between the weight vectors **w** (see Eq. (11)) of the two tasks in the pair, expecting a correlation between the distances and the amount of transfer between the tasks. The intuition is that a great difference in the final weights after training on two tasks suggests that the tasks are dissimilar and thus there is little transfer between them. Fig. 10 shows that this is indeed the case. Points are clustered in two columns, one centered around weight distance 0.6 and one around 1.4. Respective clusters in Figs. 9 and 10 consist of the same points (excluding the control conditions $\overline{xx}$ where the weight distance is trivially zero); in other words, the points with a small weight distance between trained and tested condition are exactly the ones exhibiting significant transfer whereas the points with a large weight distance are the ones characterized by little or no transfer.

4. General discussion

4.1. Key findings

Our simulations demonstrate that a simple reweighting model can account for a broad set of experimental results of hyperacuity research. In our first simulation we showed that the model can account for simple parametric manipulations of bar length and bar separation in a Vernier task. In the second simulation we found that the model can account for the finding of disruption (Seitz et al., 2005) between training of 3-dot hyperacuity stimuli of the same orientation but opposite offset sides and shows no disruption when training on orthogonal orientations. In the third simulation we found that the model can account for findings by Webb et al. (2007) that learning transfers across configurations and orientations because the axis of displacement of the central dot was conserved between these conditions. In the fourth simulation, we showed that our model can account for data showing interference between different contexts (Petrov et al., 2005). Finally, we simulated a large set of different stimuli and tasks to generate predictions of what degree of transfer and disruption are expected across different conditions. The results of these simulations show two clusters of data points; one cluster showing high transfer and high disruption and a second cluster showing low transfer and low disruption. Together, the simulations help provide an

understanding of the dominant mechanism involved in existing research and provide predictions to guide further research.

4.2. Model innovations

The aim of this work was not to introduce a new model, but rather to test previous ideas in a larger context. The model is similar to previous reweighting models. It was based on the model of Weiss et al. (1993), and like its precursor was aimed to provide "a minimalist platform for studying the improvement of hyperacuity with practice" (Weiss et al., 1993). Compared to the initial study, however, it includes several additions and modifications that render it more plausible biologically and enable comparison with a diversity of psychophysical results.

Our model is also similar to the other published reweighting models (Dosher & Lu, 1998; Petrov et al., 2005; Vaina et al., 1995). It differs in the details of the implementation, for example in that it is fully implemented neurally (unlike Dosher & Lu, 1998); it can process raw images (unlike Vaina et al., 1995); it uses a plausible model of noise (multiplicative instead of additive noise, Petrov et al., 2005; Vaina et al., 1995). Our model can be considered a simplification of the model of Petrov et al. (2005), which includes multiple stages of integration with respect to spatial phase and scale and features complex operations such as response normalization.

How does our model compare with the performance of the models outlined in the introduction, which employ reweighting, representation modification or both? Most of these models, including the ones that employ representation modification alone (e.g. Mato & Sompolinsky, 1996), have shown lack of transfer of learning to orthogonal orientations (in the case of oriented stimuli). Our model is in agreement with these results: regardless of the stimulus, there is very little (or no) transfer of learning to orthogonal orientations (points $\widehat{xy}$ in Fig. 9). Thus it seems that either learning mechanism is by itself able to account for at least some forms of specificity of learning.

Although we have not addressed the question of whether representation modification alone is sufficient to account for other experimental results, our simulations show that a number of results of visual hyperacuity learning can be accounted for without representational changes. In this sense, our modeling work is, to our knowledge, the first one that has been shown to unify a considerable diversity of experimental findings.

However, it should be noted that our model is very simple and it would be premature to conclude from our simulations anything regarding the neural locus at which this plasticity is taking place. Notably, numerous perceptual learning studies show plasticity in visual cortex (Furmanski, Schluppeck, & Engel, 2004; Raiguel, Vogels, Mysore, & Orban, 2006; Schoups et al., 2001; Schwartz, Maquet, & Frith, 2002; Vaina, Belliveau, Roziers, & Zeffiro, 1998; Yang & Maunsell, 2004; Zohary, Celebrini, Britten, & Newsome, 1994); changes have indeed been found (Schiltz et al., 1999; Schoups et al., 2001; Schwartz et al., 2002) but the extent to which they are important in the relative contributions of representational and decisional processes in perceptual learning is still under debate.

4.3. Model limitations

Despite the success of the model in explaining various findings on specificity and disruption, there are a number of phenomena that cannot be accounted for with the current form of the model.

First, one aspect of perceptual learning that the model cannot account for by design is timing. Seitz et al. (2005) found that the disruptive effects of training on a second task are mitigated if sufficient time is allowed to pass between the two sessions, a process known as *consolidation*. Sleep has also been found to have protective effects against disruption. The model does not incorporate any consolidation mechanisms. Doing so may be an interesting direction for future research.

Second, the model also fails to account for certain spatial phenomena. Otto, Herzog, Fahle, and Zhaoping (2006), and later Tartaglia et al. (2009), found that learning of a line bisection task fails to transfer to an (otherwise identical) task involving shorter line elements. Our model exhibits the opposite behavior, i.e. high transfer of learning. Likewise, the model cannot account for spatial manipulations of the Webb et al. (2007) experiment, where the observed improvements occurred in sessions where the reference element separation (outer dots) was relatively large (several degrees of visual angles). In the model, the distance of the reference elements does not significantly influence the results. The model consists of a single spatial scale with receptive fields large enough to contain the stimulus and thus no transition to different spatial scales occurs as a result of training at larger line separations (the putative mechanism suggested by Webb et al. (2007)).

In the simulation of the Petrov et al. (2005) experiment, our model fails to replicate the contrast reversal effect in congruent conditions (see Section 3.4.2). It is not yet clear what needs to be changed in our model to account for this effect, which, as Petrov et al. (2005) point out, is highly sensitive to model parameters.

Another set of findings that the model is not designed to address is the absence of retinal-location specificity under certain conditions. Retinal-location specificity, typically used in arguments in favor of the low-level nature of perceptual learning, has been challenged by Xiao et al. (2008), who showed that training two different tasks on two distinct retinal locations enables transfer of learning of one of the tasks across these locations. This suggests the involvement of higher brain areas, which are not modeled here. It would be interesting to investigate whether models that include top-down mechanisms, similar to those suggested by Herzog and Fahle (1998) and Schäfer et al. (2007), are able to account for the transfer observed by Xiao et al. (2008).

It should be noted that recurrent interactions, which are not included in our model, may be able to account for certain aspects of perceptual learning, such as the fact that learning occurs in the face of positional variability due to eye movements – a problem shown to be solved by the recurrent network of Zhaoping et al. (2003). Also, models based solely on plasticity of long-range recurrent connections in V1 (Zhaoping, 2009) make interesting predictions for more complex stimuli. In particular, they predict that learning of pop-out detection should be more specific when the background texture is denser – a prediction that has received some experimental support (Karni & Sagi, 1991; Sireteanu & Rettenbach, 1995). According to Zhaoping (2009), the reason is that intra-cortical connections are of a finite range in V1 (2–4 mm) and thus would not be activated by stimuli formed of distant texture elements. In this case, learning would have to occur at higher cortical areas and would not be as specific. Recurrent interactions could still be incorporated into a reweighting model of a larger scale than ours (namely one whose representation units collectively cover a broad part of the visual field). Recurrent plasticity would not permanently change the classical RF of the representational units and may not serve to better encode a stimulus in general: such changes can serve to modulate the responses of these units in a task-dependent manner when the stimulus consists of multiple visual elements and/or is embedded in context.

Lastly, there is accumulating evidence that disruption also depends on the way the training sessions are interleaved. It has been found that learning occurs when multiple stimuli are presented in a fixed sequence (temporal patterning) but not when they are presented randomly (roving) (Zhang et al., 2008). However, a later study by Tartaglia et al. (2009) showed that roving is not necessarily

Table 2
Model parameters for all simulations presented in Section 3 of the main text. "′" denotes minutes of arc. Parameter values are consistent with RF sizes and shapes of cortical cells receiving parafoveal input.

Parameter	Symbol	Value	Comment
x-Width of Gaussian envelope for the threshold–length and threshold–separation simulations (Westheimer & McKee, 1977)	σ_x	3.9′	See Eq. (9)
x-Width of Gaussian envelope for all other simulations	σ_x	12′	See Eq. (9)
y-Width of Gaussian envelope for the threshold–length and threshold–separation simulations	σ_y	5.2′	See Eq. (9)
y-Width of Gaussian envelope for all other simulations	σ_y	16′	See Eq. (9)
Spatial frequency for the threshold–length and threshold–separation simulations	f	$(5')^{-1}$	See Eq. (9)
Spatial frequency for the context simulation (Petrov et al., 2005)	f	$(8.6')^{-1}$	See Eq. (9)
Spatial frequency for all other simulations	f	$(24')^{-1}$	See Eq. (9)
Maximum firing rate	r_{max}	100 Hz	See Eq. (7)
Gain	g	0.005	Current value produced realistic firing rates across the range of stimuli presented to the oriented units. See Eq. (7)
Threshold	h_0	0 Hz	See Eq. (7)
Stimulus length (for the simulations in Section 3.1)	–	4′ and 8′	As in Westheimer and McKee (1977)
Stimulus length (for all other simulations)	–	20′	For all dot stimuli and for the line bisection stimulus, this is the distance between the outer elements; for Verniers, it is the combined length of the two bars plus their separation; for Chevrons, it is the combined length of the two bars
Bar thickness	–	0.1′	Applies to Verniers, Chevrons and Curvature stimuli
Dot radius	–	2′	Applies to dot alignment and bisection stimuli
Vernier bar separation (for the simulations in Section 3.1)	–	0′	As in Westheimer and McKee (1977)
Vernier bar separation (for all other simulations)	–	2′	Applies to Vernier stimuli
Line bisection bar length	–	6.75′	Applies to line bisection stimuli
Learning rate for the disruption simulations (Seitz et al., 2005)	η_d	0.00035	
Learning rate for the generalization (Webb et al., 2007) and context (Petrov et al., 2005) simulations	η_t	0.003	
Number of preferred orientations	–	13	Equally spaced in $[-\pi, \pi]$
Number of preferred spatial phases	–	7	Equally spaced in $[-\pi, \pi]$
Number of "noise" units	–	59	See Fig. 4

disruptive; if the stimuli are clearly distinct, learning is possible. The same study also showed timing dependencies in roving conditions. The model does not distinguish between different temporal patterns or roving conditions: the amount of learning is the same regardless of interleaving patterns. These findings are very recent and more investigation seems necessary in order to obtain conclusive results. It is, however, evident that an investigation of the mechanisms underlying the influences blocking and interleaving different stimuli is an important future direction for models of perceptual learning.

4.4. Conclusions

By means of extensive simulations, we have shown that a very simple reweighting model can account for several key findings of perceptual learning of hyperacuity. The model's ability to generalize with respect to task and stimulus attributes is comparable to that of human observers. Furthermore, the model accounts for disruption of learning of one task by subsequent practice on a similar task. The fact that important findings on perceptual learning (specificity/generalization and disruption), obtained under quite different experimental conditions, can be accounted for within a single framework is encouraging. While the current model has notable limitations, it explains a diversity of results and is a promising foundation to build upon to account for some of the subtler findings that are emerging in empirical research on perceptual learning.

Appendix A. Derivation of Eq. (10)

The probability of obtaining a correct response from the untrained network (i.e. the observed success rate) without the top-down influence is $F(e) = P(O = Y)$, with $P(O = Y) = 0.5$ (chance-level performance). With the top-down influence, this becomes $F(e) = p + (1 - p)P(O = Y) = p + (1 - p)0.5 \Rightarrow p = 2F(e) - 1$.

References

Aberg, K., Tartaglia, E., & Herzog, M. (2009). Perceptual learning with Chevrons requires a minimal number of trials, transfers to untrained directions, but does not require sleep. *Vision Research, 49*, 2087–2094.

Adini, Y., Sagi, D., & Tsodyks, M. (2002). Context-enabled learning in the human visual system. *Nature, 415*, 790–793.

Ahissar, M., & Hochstein, S. (1993). Attentional control of early perceptual learning. *Proceedings of the National Academy of Sciences, 90*, 5718–5722.

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences, 8*, 457–464.

Ball, K., & Sekuler, R. (1987). Direction-specific improvement in motion discrimination. *Vision Research, 27*, 953–965.

Broomhead, D., & Lowe, D. (1988). Multivariable functional interpolation and adaptive networks. *Complex Systems, 2*, 321–355.

Crist, R., Kapadia, M., Westheimer, G., & Gilbert, C. (1997). Perceptual learning of spatial localization: specificity for orientation, position, and context. *Journal of Neurophysiology, 78*, 2889–2894.

Dayan, P., & Abbott, L. (2001). *Theoretical neuroscience: Computational and mathematical modeling of neural systems*. MIT Press.

De Valois, K. (1977). Spatial frequency adaptation can enhance contrast sensitivity. *Vision Research, 17*, 1057–1065.

Dosher, B., & Lu, Z. (1998). Perceptual learning reflects external noise filtering and internal noise reduction through channel reweighting. *Proceedings of the National Academy of Sciences, 95*, 13988–13993.

Dosher, B., & Lu, Z. (1999). Mechanisms of perceptual learning. *Vision Research, 39*, 3197–3221.

Fahle, M. (1997). Specificity of learning curvature, orientation, and Vernier discriminations. *Vision Research, 37*, 1885–1895.

Fahle, M. (2005). Perceptual learning: Specificity versus generalization. *Current Opinion in Neurobiology, 15*, 154–160.

Fahle, M., & Edelman, S. (1993). Long-term learning in vernier acuity: Effects of stimulus orientation, range and of feedback. *Vision Research, 33*, 397–412.

Fahle, M., Edelman, S., & Poggio, T. (1995). Fast perceptual learning in hyperacuity. *Vision Research, 35*, 3003–3013.

Fahle, M., & Morgan, M. (1996). No transfer of perceptual learning between similar stimuli in the same retinal position. *Current Biology, 6*, 292–297.

Fendick, M., & Westheimer, G. (1983). Effects of practice and the separation of test targets on foveal and peripheral stereoacuity. *Vision Research, 23*, 145–150.
Furmanski, C., Schluppeck, D., & Engel, S. (2004). Learning strengthens the response of primary visual cortex to simple patterns. *Current Biology, 14*, 573–578.
Gibson, E. (1963). Perceptual learning. *Annual Revision of Psychology, 14*, 29–56.
Herzog, M., & Fahle, M. (1997). The role of feedback in learning a Vernier discrimination task. *Vision Research, 37*, 2133–2141.
Herzog, M., & Fahle, M. (1998). Modeling perceptual learning: Difficulties and how they can be overcome. *Biological Cybernetics, 78*, 107–117.
Jones, J., & Palmer, L. (1987). An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology, 58*, 1234–1236.
Karni, A., & Sagi, D. (1991). Where practice makes perfect in texture discrimination: Evidence for primary visual cortex plasticity. *Proceedings of the National Academy of Sciences, 88*, 4966–4970.
Karni, A., Tanne, D., Rubenstein, B., Askenasy, J., & Sagi, D. (1994). Dependence on REM sleep of overnight improvement of a perceptual skill. *Science, 265*, 679–682.
Koyama, S., Harner, A., & Watanabe, T. (2004). Task-dependent changes of the psychophysical motion-tuning functions in the course of perceptual learning. *Perception, 33*, 1139–1148.
Kramer, D., & Fahle, M. (1996). A simple mechanism for detecting low curvatures. *Vision Research, 36*, 1411–1419.
Kuai, S., Zhang, J., Klein, S., Levi, D., & Yu, C. (2005). The essential role of stimulus temporal patterning in enabling perceptual learning. *Nature Neuroscience, 8*, 1497–1499.
Li, W., Piëch, V., & Gilbert, C. (2004). Perceptual learning and top-down influences in primary visual cortex. *Nature Neuroscience, 7*, 651–657.
Liu, Z., & Vaina, L. (1998). Simultaneous learning of motion discrimination in two directions. *Cognitive Brain Research, 6*, 347–349.
Mato, G., & Sompolinsky, H. (1996). Neural network models of perceptual learning of angle discrimination. *Neural Computation, 8*, 270–299.
Mayer, M. (1983). Practice improves adults' sensitivity to diagonals. *Vision Research, 23*, 547–550.
McKee, S., & Westheimer, G. (1978). Improvement in vernier acuity with practice. *Perception & Psychophysics, 24*, 258–262.
Mednick, S., Nakayama, K., & Stickgold, R. (2003). Sleep-dependent learning: A nap is as good as a night. *Nature Neuroscience, 6*, 697–698.
Mollon, J., & Danilova, M. (1996). Three remarks on perceptual learning. *Spatial Vision, 10*, 51–58.
Moody, J., & Darken, C. (1989). Fast learning in networks of locally-tuned processing units. *Neural computation, 1*, 281–294.
Otto, T., Herzog, M., Fahle, M., & Zhaoping, L. (2006). Perceptual learning with spatial uncertainties. *Vision Research, 46*, 3223–3233.
Petrov, A., Dosher, B., & Lu, Z. (2005). The dynamics of perceptual learning: An incremental reweighting model. *Psychological Review, 112*, 715–743.
Poggio, T., Fahle, M., & Edelman, S. (1992). Fast perceptual learning in visual hyperacuity. *Science, 256*, 1018–1021.
Raiguel, S., Vogels, R., Mysore, S., & Orban, G. (2006). Learning to see the difference specifically alters the most informative V4 neurons. *Journal of Neuroscience, 26*, 6589–6602.
Ringach, D. (2002). Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. *Journal of Neurophysiology, 88*, 455–463.
Roelfsema, P., & Ooyen, A. (2005). Attention-gated reinforcement learning of internal representations for classification. *Neural Computation, 17*, 2176–2214.
Schäfer, R., Vasilaki, E., & Senn, W. (2007). Perceptual learning via modification of cortical top-down signals. *PLoS Computational Biology, 3*, e165.
Schiltz, C., Bodart, J., Dubois, S., Dejardin, S., Michel, C., Roucoux, A., et al. (1999). Neuronal mechanisms of perceptual learning: changes in human brain activity with training in orientation discrimination. *Neuroimage, 9*, 46–62.
Schoups, A., Vogels, R., & Orban, G. (1995). Human perceptual learning in identifying the oblique orientation: Retinotopy, orientation specificity and monocularity. *Journal of Physiology, 483*, 797–810.
Schoups, A., Vogels, R., Qian, N., & Orban, G. (2001). Practising orientation identification improves orientation coding in V1 neurons. *Nature, 412*, 549–552.
Schwartz, S., Maquet, P., & Frith, C. (2002). Neural correlates of perceptual learning: A functional MRI study of visual texture discrimination. *Proceedings of the National Academy of Sciences, 99*, 17137–17142.
Seitz, A., Yamagishi, N., Werner, B., Goda, N., Kawato, M., & Watanabe, T. (2005). Task-specific disruption of perceptual learning. *Proceedings of the National Academy of Sciences, 102*, 14895–14900.
Shiu, L., & Pashler, H. (1992). Improvement in line orientation discrimination is retinally local but dependent on cognitive set. *Perception and Psychophysics, 52*, 582–588.
Sireteanu, R., & Rettenbach, R. (1995). Perceptual learning in visual search: Fast, enduring, but non-specific. *Vision Research, 35*, 2037–2043.
Sowden, P., Rose, D., & Davies, I. (2002). Perceptual learning of luminance contrast detection: Specific for spatial frequency and retinal location but not orientation. *Vision Research, 42*, 1249–1258.
Swindale, N., & Cynader, M. (1986). Vernier acuity of neurones in cat visual cortex. *Nature, 319*, 591–593.
Tartaglia, E., Aberg, K., & Herzog, M. (2009). Perceptual learning and roving: Stimulus types and overlapping neural populations. *Vision Research, 49*, 1420–1427.
Teich, A., & Qian, N. (2003). Learning and adaptation in a recurrent model of V1 orientation selectivity. *Journal of Neurophysiology, 89*, 2086–2100.
Vaina, L., Belliveau, J., Roziers, E., & Zeffiro, T. (1998). Neural systems underlying learning and representation of global motion. *Proceedings of the National Academy of Sciences, 95*, 12657–12662.
Vaina, L., Sundareswaran, V., & Harris, J. (1995). Learning to ignore: psychophysics and computational modeling of fast learning of direction in noisy motion stimuli. *Cognitive Brain Research, 2*, 155–163.
Webb, B., Roach, N., & McGraw, P. (2007). Perceptual learning in the absence of task or stimulus specificity. *PLoS ONE, 2*, e1323.
Weiss, Y., Fahle, M., & Edelman, S. (1993). Models of perceptual learning in vernier hyperacuity. *Neural Computation, 5*, 695–718.
Westheimer, G. (1976). Diffraction theory and visual hyperacuity. *American Journal of Optometry and Physiological Optics, 53*, 362–364.
Westheimer, G., & McKee, S. (1977). Spatial configurations for visual hyperacuity. *Vision Research, 17*, 941–947.
Xiao, L., Zhang, J., Wang, R., Klein, S., Levi, D., & Yu, C. (2008). Complete transfer of perceptual learning across retinal locations enabled by double training. *Current Biology, 18*, 1922–1926.
Yang, T., & Maunsell, J. (2004). The effect of perceptual learning on neuronal responses in monkey visual area V4. *Journal of Neuroscience, 24*, 1617–1626.
Yotsumoto, Y., Chang, L., Watanabe, T., & Sasaki, Y. (2009). Interference and feature specificity in visual perceptual learning. *Vision Research, 49*, 2611–2623.
Yu, C., Klein, S., & Levi, D. (2004). Perceptual learning in contrast discrimination and the (minimal) role of context. *Journal of Vision, 4*, 169–182.
Zhang, J., Kuai, S., Xiao, L., Klein, S., Levi, D., & Yu, C. (2008). Stimulus coding rules for perceptual learning. *PLoS Biology, 6*, e197.
Zhaoping, L. (2009). Perceptual learning of pop-out and the primary visual cortex. *Learning & Perception, 1*, 135–146.
Zhaoping, L., Herzog, M., & Dayan, P. (2003). Nonlinear ideal observation and recurrent preprocessing in perceptual learning. *Network: Computation in Neural Systems, 14*, 233–247.
Zohary, E., Celebrini, S., Britten, K., & Newsome, W. (1994). Neuronal plasticity that underlies improvement in perceptual performance. *Science, 263*, 1289–1292.