Disparity tuning as simulated by a neural net

Jörg Lippert[1], **David J. Fleet**[2], **Hermann Wagner**[1]

[1] Institut für Biologie II, RWTH Aachen, Kopernikusstrasse 16, D-52074 Aachen, Germany
[2] Department of Computing and Information Science, Queen's University, Kingston, Ontario, Canada

Abstract. Previous research has suggested that the processing of binocular disparity in complex cells may be described with an energy formalism. The energy formalism allows for a representation of disparity by differences in the position or in the phase of monocular receptive subfields of binocular cells, or by combination of these two types. We studied the coding of disparities with an approach complementary to previous algorithmic investigations. Since realization of these representations is probably not genetically determined but learned during ontogeny, we used backpropagation networks to study which of these three possibilities were realized within neural nets. Three types of networks were trained with noise patterns in analogy to the three types of energy models. The networks learned the task and generalized to untrained correlated noise pattern input. Outputs were broadly tuned to spatial frequency and did not respond to anti-correlated noise patterns. Although the energy model was not explicitly implemented, we could analyze the outputs of the networks using predictions of the energy formalism. After learning was completed, the model neurons preferred position shifts over phase shifts in representing disparity. We discuss the general meaning of these findings and the correspondences and deviations between the energy model, V1 neurons, and our networks.

1 Introduction

Neurons with a specific sensitivity to retinal disparities are thought to play a fundamental role in the perception of stereoscopic depth. Such neurons have been found in the primary visual cortex of mammals, especially cats (Pettigrew et al. 1968; Ohzawa et al. 1997) and monkeys

(Poggio and Fischer 1977; Poggio 1995). They have also been found in the visual wulst of owls (Pettigrew and Konishi 1976; Wagner and Frost 1993, 1994).

Several studies have investigated the characteristics of neurons with excitatory disparity tuning by means of an algorithmic approach (experimental neurophysiology: Wagner and Frost 1993, 1994; Ohzawa et al. 1990, 1997; theory: Qian 1994; Fleet et al. 1996a, b; Zhu and Qian 1996; Qian and Zhu 1997). These studies led to the development of a model for disparity-sensitive neurons, the energy neuron model, which explains most of the known experimental facts. Yet, some questions concerning the mechanism of disparity selectivity remain open.

Three types of energy models are known, namely, phase-shift, position-shift, and hybrid models. In the phase-type model, disparity selectivity is a result of differences in the phases of the monocular receptive subfields of the binocular neurons. Differences in the retinal positions of monocular subfields determine the disparity selectivity in the position type model. Both phase and positional differences contribute to disparity selectivity in the hybrid-type model. These three types of energy models yield clear predictions of the responses of real disparity-sensitive neurons (Fleet et al. 1996b). Experimental data from the cat suggest that phase-shifts play a significant role, favoring phase-type or hybrid-type models (DeAngelis et al. 1991; Anzai et al. 1997, 1999; Ohzawa et al. 1997), while data for the owl favor the position-type model (Wagner and Frost 1993, 1994). However, the data published so far do not rule out any of the three models.

In this paper, we take a different approach, using a neural network to learn more about the effectiveness of the different models. We used bandpass monocular filters that differed in interocular position and interocular phase to provide input to neural nets that we trained with a conventional backpropagation algorithm to be disparity selective. Networks learned to exhibit disparity selectivity much like that of V1 cells. These artificial neurons are shown to have many properties in common with those often recorded from V1 cells. Moreover,

Correspondence to: J. Lippert
(e-mail: joerg@candide.bio2.rwth-aachen.de,
Tel.: +49-241-807531 (lab)/804835 (secretary),
Fax: +49-241-8888133)

although the networks were not explicitly designed in accordance with the energy model, we were also able to analyze the results in terms of predictions of the energy model.

2 Materials and methods

2.1 The backpropagation model

Three-layered feedforward networks were implemented (Fig. 1). Temporal characteristics of real cells were not incorporated into the model. The inputs to units in the hidden and output layers were weighted sums of the responses, x_i, from units at the previous layer. The total input to a single unit, y, can be written as:

$$y = \omega_0 + \sum x_i \omega_i \ ,$$

where ω_i are the weights. Every hidden unit and the output unit had an additional constant input equal to 1. The weights ω_0 of these supplementary inputs act as threshold values for each unit and are also learned. The output function R of the hidden units and the output unit was a Fermi function of the input:

$$R(y) = 1/[1 + \exp(-y)] \ .$$

The input stereograms consisted of two one-dimensional half-images $I_L(x)$ and $I_R(x) = I_L(x - d)$ that were translated versions of one another. As shown in Fig. 1, the
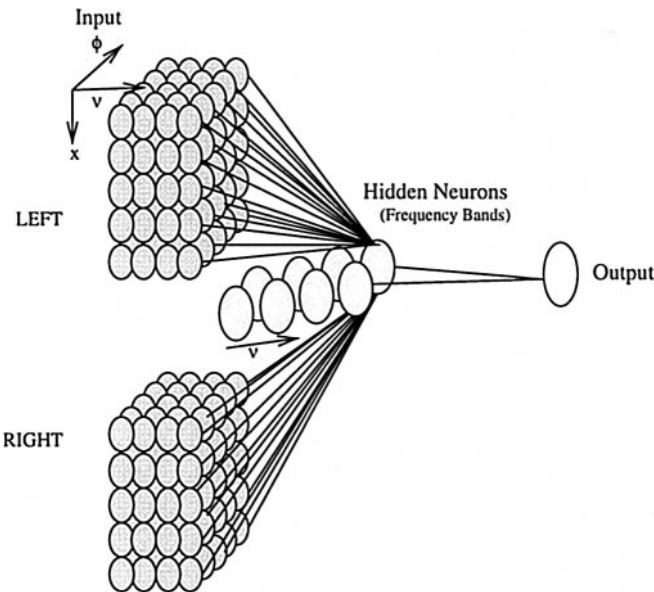

Fig. 1. Topology of the network model. The input layer consisted of two sets of inputs, representing left and right eye receptive field (RF) responses. RFs in each set differed in best frequency, v, phase, ϕ, and position, x (x was kept constant in the phase type networks and ϕ was kept constant in the position type networks). This resulted in a three-dimensional input space. The hidden units were divided into four groups of two neurons representing different frequency bands. Every hidden unit was connected to all input units from both eyes with a single preferred frequency. The output unit was connected to all hidden units. (For clarity, only a reduced set of connections is shown)
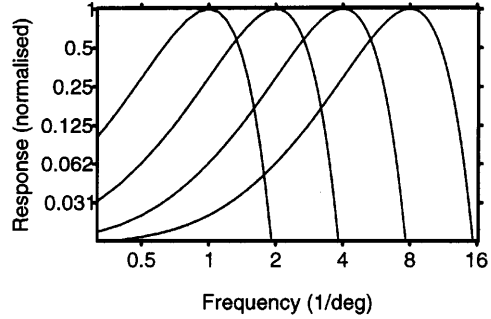

Fig. 2. Frequency filters. Four filters covering four octaves and having a constant ratio of best frequency and tuning width (defined in Sect. 2) were used

first layer units were divided into two groups, representing left and right receptive fields (RF), and were modeled as Gabor functions. Inputs to the second layer were the responses of the Gabor filters applied to the monocular half-images of the stereograms. Filter operations were performed on a discrete lattice of 1/320 deg unit length; i.e., receptive fields are spaced 1/320 deg apart.

The spatial frequency selectivity of the Gabor RFs was consistent with psychophysical data (Campbell and Robson 1968) and neurophysiological recordings (DeValois et al. 1982). The best frequencies, v, of the filters were set to 1, 2, 4 and 8 cyc/deg (Fig. 2). To provide a constant half-height bandwidth of about 1.1 octaves (1.4 octaves at e^{-1}-height) for all filters, the width of the Gaussian envelope σ was chosen as:

$$\sigma = 0.5/v \ .$$

Apart from differences in frequency selectivity, the RFs consisted of Gabor filters at five different phases ϕ ($\phi = -0.4, -0.2, 0, 0.2, 0.4$ periods) and at five different spatial positions x_0 ($x_0 = -0.4/v_i$ deg^{-1}; $-0.2/v_i$ deg^{-1}; $0/v_i$ deg^{-1}; $0.2/v_i$ deg^{-1}; $0.4/v_i$ deg^{-1}).

Three different types of RF combinations were used as input to the three types of networks:

Type 1: hybrid-type model, with five different phases and five different positions, resulting in 2 (half-image) $\times$ 4 (frequency) $\times$ 5 (phase) $\times$ 5 (position) = 200 filters.

Type 2: phase-type model (position constant), with $2 \times 4 \times 5 \times 1 = 40$ filters.

Type 3: position-type model (phase constant), with $2 \times 4 \times 1 \times 5 = 40$ filters.

The input-layer units distributed the Gabor filter outputs to the hidden layer. The 8 cells of the hidden layer were binocular and received input from both subsets of monocular RFs. They were subdivided into four groups representing different spatial frequency tunings. All eight hidden units projected onto a single output unit.

2.2 The training

The weights ω_i were determined during a training period using standard backpropagation (Rumelhart

et al. 1986). We determined suitable values for the learning rate α and the impulse γ during preliminary tests and kept them constant during training for all networks ($\alpha = 0.6$, $\gamma = 0.05$). The starting weights were determined using a random number generator (uniformly distributed between -3 and 3).

The input images used for learning were one-dimensional samples of random noise. Pixels were 1/32 deg wide. The gray values were 0, 1, or 2 (independent and uniformly distributed). The image width was 6.25 deg. Thirty-one stereo pairs, each having a different disparity, were constructed from 100 different random dot patterns, yielding a training set of 3100 stereograms. Disparity values were evenly distributed between -0.235 deg to $+0.235$ deg, with a disparity step size of 0.0125 deg. Thus, the largest disparities are approximately 1/4 of the wavelength to which the low-frequency filter is tuned, and more than two wavelengths of the frequency to which the high frequency filter is tuned. In the iterative training process, one of the 3100 stereograms was selected at random and a learning iteration was performed. Learning was stopped after 3×10^6 iterations. Seventy-five independent nets were trained for each of the three network types (position, phase, and hybrid).

The teacher was the value of a Gaussian function of the input's disparity. The Gaussian was centered at the "desired" preferred disparity and its peak value was 1. Preferred disparities for the different nets varied from -0.1875 deg to 0.1875 deg in steps of 0.0125 deg. The standard deviation of the Gaussian was 0.063 deg, similar to the widths of physiologically measured tuning curves in monkeys (Poggio 1995).

2.3 Analysis after training

All nets were tested with both the training set and two other data sets that were not used in the training process. These additional stereograms were created in the same way as those used for training. In one of the test sets, the range of disparities was equal to the range used in training. In the other test set, it was four times larger than the training range, i.e., -0.9375 deg to 0.9375 deg. The responses of the nets were plotted as a function of disparity (called "disparity tuning curve") as are responses from physiological recordings. We determined the disparity that produced the largest response, as well as the width (at half-height) of these response curves. All hybrid nets were also tested with anti-correlated random dot stereograms (aRDS) like those used by Cumming and Parker (1997) to study single cells in V1. In particular, while aRDSs do not produce a stereo percept in humans, they do elicit significant responses in primate V1 cells. We generated aRDS by inverting the contrast of one of the half-images of a normal stereo pair. We plotted response curves for both correlated and anti-correlated noise for each net. Histograms of the response modulation, i.e., the difference between maximum and minimum responses, in the RDS and the aRDS cases, were generated.

We also measured the responses of the nets to binocular sine wave gratings. Sinusoidal gratings are used in electrophysiology (Yin and Kuwada 1983; Wagner and Frost 1993) and theory (Fleet et al. 1996a) to characterize neuronal behavior as a function of stimulus frequency. We use the same stimuli to probe our trained units and to compare their responses to real cells. As is customary in physiological studies (e.g., Wagner and Frost 1994), stimuli were drifted through the receptive fields (in steps of 1/320 deg), and responses were integrated over one spatial period of the stimulus. We used gratings of 27 different wavelengths λ_i between 0.0625 deg and 1.5 deg. Gratings at each frequency were presented at 41 disparities ranging from $-1/2\lambda_i$ to $+1/2\lambda_i$. Responses were plotted as disparity tuning curves.

The network responses to sinusoidal inputs were analyzed in several ways that allowed a comparison to electrophysiological data and to theoretical considerations. First, we determined the maximum response to all frequencies. These data allowed us to plot iso-intensity rate functions, a measure of the frequency selectivity of the entire network. Then, we calculated disparity selectivity using the vector strength (VS) measure at each grating frequency; VS is a common measure of the strength of modulation of cyclic phenomena (Batschelet 1981). We plotted the VS as a function of the grating frequency for each network. A Rayleigh test ($P < 0.05$, Batschelet 1981) was then used with the VS data to decide whether the network was disparity-tuned at a given frequency. For each frequency at which a network exhibited significant disparity tuning, we then calculated the mean binocular phase, or in short, the mean phase of the disparity tuning (MP, Batschelet 1981); the MP in circular statistics is regarded as an equivalent of the familiar arithmetic mean in non-circular statistics. The MP gives the preferred interocular phase at each grating frequency, which can be viewed as the analogue to preferred disparity with noise stimuli.

Plotting the MPs of all tuned disparity-response curves, as a function of the stimulus frequency, ω, yields phase-frequency plots (Yin and Kuwada 1983; Fleet et al. 1996a). The energy model predicts different linear-phase frequency relations for the three types of models. The slope and y-intercept of the plot determine the position-shift (i.e., the characteristic disparity, CD) and phase-shift (i.e. the characteristic phase, CP) of the monocular receptive subfields (Fleet et al. 1996a).

Hybrid-type model: $\mathrm{MP} = \mathrm{CD} \times \omega + \mathrm{CP}$, $\qquad$ (1)

Phase-type model: $\mathrm{MP} = \mathrm{CP}$, $\qquad$ (2)

Position-type model: $\mathrm{MP} = \mathrm{CD} \times \omega$. $\qquad$ (3)

The phase-frequency plots were tested for linearity using the mean square error (MSE, Yin and Kuwada 1983) that measures the distance between each point and the regression line. The MSE is independent of the slope of the line, and thus better suited as a test for linearity than the regression coefficient. Values of significance for the

MSE at $P < 0.005$ were obtained from Yin and Kuwada (1983).

From those phase-frequency plots that satisfied the MSE linearity criterion, we computed the slope of the regression line, which is equivalent to the CD (Yin and Kuwada 1983), and the intercept of the line, which represents CP. The distributions of CD and CP helped to decide whether phase shifts or position shifts were responsible for the disparity tuning.

A Runs test (Siegel 1956) applied to the residuals of the phase-frequency plots suggested the existence of two subclasses of plots. The first class showed significantly ($P < 0.005$) fewer changes in the signs of their residuals than would be expected from a linear relation with random scatter; the second class was not significant according to the Runs test.

In order to study the effects of individual frequency bands, an extra test was applied to the phase-type networks. We computed phase-frequency plots, CD and CP for each frequency band with responses from all other bands set to zero.

3 Results

The training of all networks was successful in that the networks did learn disparity selectivity. Moreover, many of the properties of the artificial net were very similar to those found from electrophysiological studies. Differences were found in responses to aRDS and in the width of frequency tuning (for details Sect. 4). Predictions of the energy neuron model were found to be valid, though some very important deviations could be observed.

3.1 Responses to noise patterns

After training, to ensure that the output units achieved the desired disparity tuning, we tested the disparity selectivity of the units with images from the training set and with a new set of images. All output neurons had acquired disparity selectivity much like that of their respective teachers. Amplitudes at the preferred disparity were distributed between 0.3 and 0.9, with hybrid-type (mean amplitude: 0.73) and position-type (mean amplitude: 0.78) models reaching higher amplitudes (Fig. 3d,e) than phase-type models (mean amplitude: 0.65, Fig. 3c; $P < 0.0001$, Mann-Whitney U-test). Differences between the hybrid-type and position-type models were also significant ($P < 0.0001$). Most nets also generalized to new test stimuli, exhibiting tuned responses (Fig. 3a). Several nets, however, did not show an excitation to test patterns (phase-type: six nets; position-type: four nets; hybrid-type: seven nets; Fig. 3c–e). The mean responses under these conditions, however, dropped to significantly lower values for all three network types (Fig. 3c–e; $P < 0.0001$ Mann-Whitney U-test for all three types). The mean amplitudes reached were 0.60, 0.53, and 0.50 in the position-, hybrid-, and phase-type networks, respectively. These
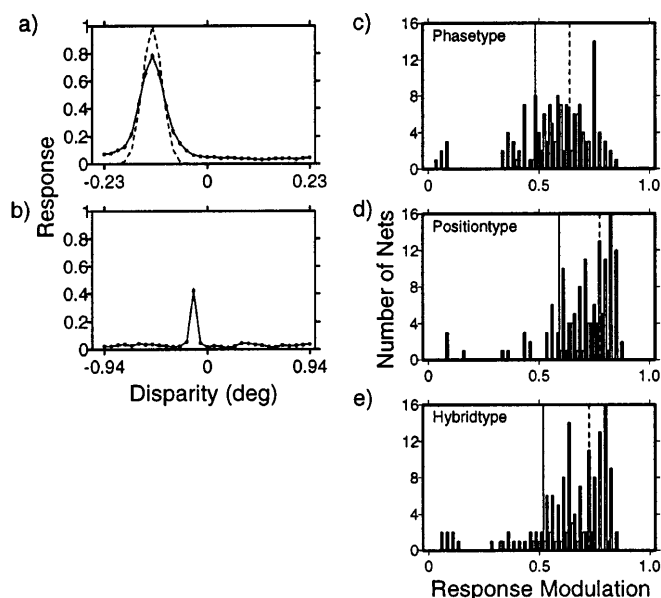

Fig. 3a–e. Learning success. **a** The output function in response to test stimuli (*dots* with standard deviations) of this hybrid net resembled the teacher function (*dashed line*). The response did not increase for disparities outside the training range (**b**). **c–e** Distributions of the amplitude of the responses to training stimuli (*filled bars*) and test stimuli (*open bars*) for the three model types. Note that the responses in the phase-type model are smaller (**c**) than those of the position-type (**d**) and the hybrid-type (**e**) nets. Means for training and test stimuli are plotted as *vertical lines*

differences were again significant (pos/ph: $P < 0.0001$, pos/hyb: $P < 0.001$, ph/hyb: $P < 0.01$).

In addition, generalization to disparities outside the range presented in training was observed (Fig. 3b).

The values of the response maxima were independent of the preferred disparity of the teacher in the position-type and hybrid-type nets (Fig. 4b,d). The phase-type nets, in contrast, showed a decline of maxima from 0.8 to 0.5 when the absolute value of disparity of the teacher increased from 0 to 0.19 deg (Fig. 4f). This reveals a characteristic property of phase-type coding; namely, disparities greater than 50% of the wavelength of the monocular receptive field cannot be coded by energy neurons of the phase-type (Fleet et al. 1996a). This effect must also hamper the coding in our phase-type networks. While the response maxima were smaller, the width of the disparity tuning curves of the nets showed no significant differences. The disparity at which the tuning curves are maximal were highly correlated with the disparity of the teacher (Fig. 4a–c). The maximum difference between the disparity of the teacher and the disparity of the output of a net were 0.016 deg for all nets. Thus, we conclude that the training of these nets was effective.

While the mean responses were closely related to the teacher function (Fig. 3a), the responses of the nets to individual images showed greater variations. In some instances, the disparity curves exhibited more than one response maximum (Fig. 5b), and the width and height of the tuning curves sometimes varied (Fig. 5a–c). The summary plot (Fig. 5d) demonstrates that nearly all
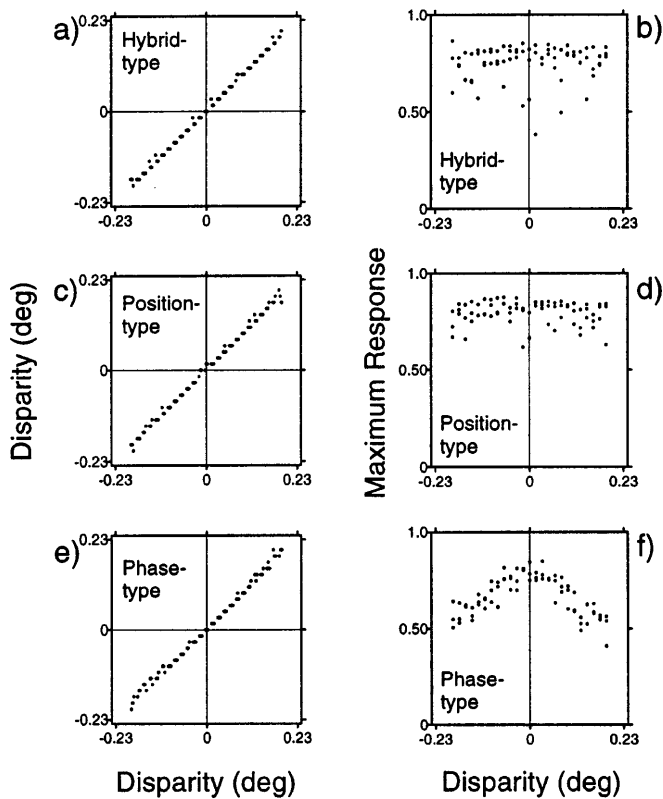
Fig. 4a–f. Maximum response of the networks. The results were obtained with test stimuli. The disparity producing a maximal responses (**a, c, e**) was closely related to the best disparity of the teacher. The maximum responses were evenly distributed for the hybrid-type (**b**) and the position-type nets (**d**). In the phase-type nets (**f**), the responses declined as absolute disparity values increased

Fig. 5a–d. Disparity tuning curves in response to different test stimuli. Individual curves may resemble the teacher closely (**a**, see also Fig. 3a), or they may vary widely and exhibit smaller peaks (**c**) or peaks at wrong disparities (**b**). The responses were obtained with a hybrid-type net. The response profile of the network to all 100 test stimuli is shown in (**d**). Note the different amplitude values in the responses

images caused a high response at the disparity corresponding to the maximum of the teacher function. However, responses to some patterns showed lower amplitudes at the best disparity. Disparities far away from the teacher's preferred disparity usually produced small responses, but they occasionally exceeded 0.4.

As was done with V1 cells (Cumming and Parker 1997) we compared the responses of the output units to correlated RDS and aRDS. In aRDS, a black dot in one eye's input pattern is matched by a white dot in the other eye. While the response of the output unit to correlated RDS was high and exhibited a clear maximum at the preferred disparity, the response of the same output unit to aRDS was constantly weak (Fig. 6a) and showed no modulation. Moreover, the histogram of all responses to aRDS demonstrated, in contrast to the results of Cumming and Parker (1997), that the responses to these stimuli were much lower than those obtained with correlated RDS (Fig. 6b). The mean ratio between responses to aRDS and RDS was 10.5%.

3.2 Responses to sine gratings

To learn more about the mechanism, we examined the disparity tuning to sinusoidal grating inputs (Fig. 7). Spatial frequency tuning curves were broader than in
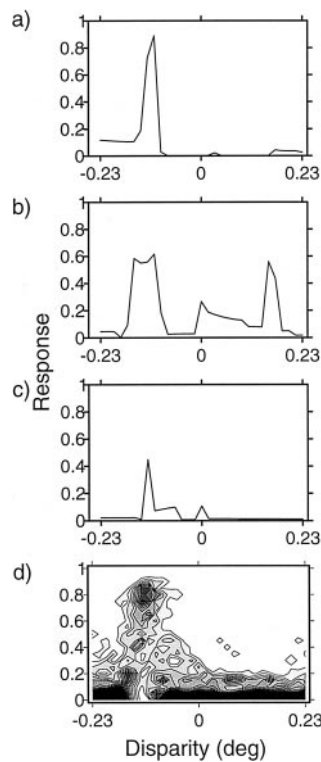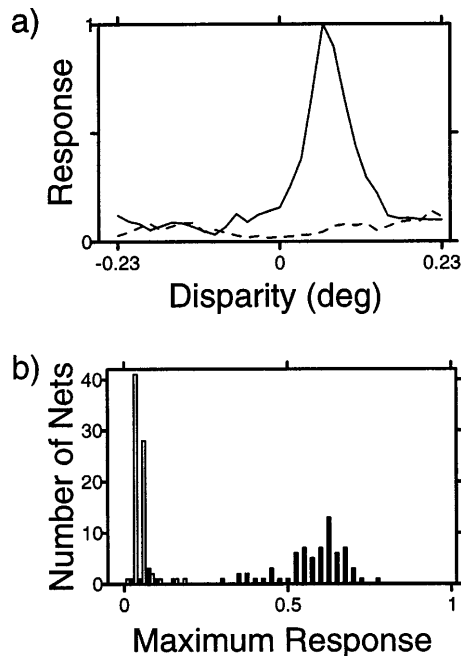

Fig. 6a,b. Responses to anti-correlated random dot stereograms. **a** Example of a hybrid network's response to RDS and anti-correlated RDS (aRDS) (*dashed lines*). **b** Histogram of the maximum responses to RDS (*filled bars*) and aRDS (*open bars*) of the hybrid nets
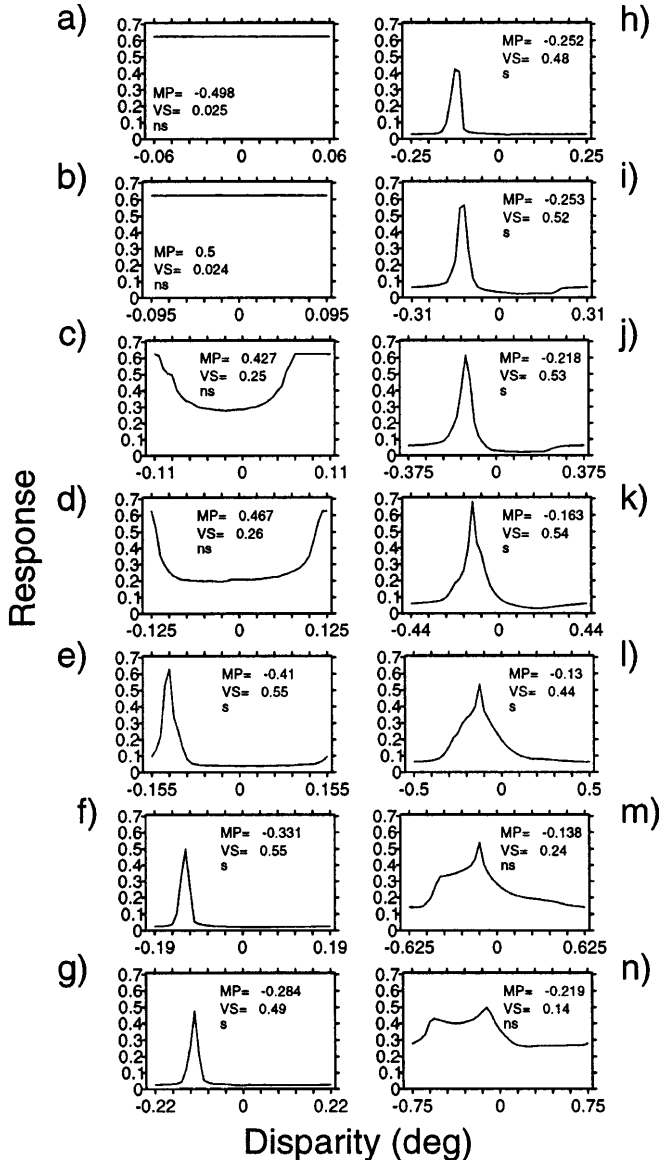
Fig. 7a–n. Disparity-tuning curves obtained with sinusoidal stimuli. Responses were obtained from one hybrid-type net. The wavelength of the stimulus varied from 0.12 deg (**a**) to 1.5 deg (**n**). The testing range covered one period of the stimulus ($\pm 0.5/v_i$). The inset shows the mean phase (*MP*) and the vector strength (*VS*) as derived from circular statistics. The result of the Rayleigh test for tuning is indicated by *s* (significant) or *ns* (not significant). Note that the maximum of the response shifts as frequency is changed

Fig. 8a–f. Examples of frequency-tuning curves of hybrid nets. *Solid lines* show iso-intensity tuning curves; half-height is indicated by the *horizontal line*. Peaked responses (**c**) resemble physiological response curves. The majority of the tuning curves were, however, low pass (**a, b**), complex (**f**), untuned (**d**) or trough (**e**). *Dashed lines* show vector strength tuning; half-height is indicated by the *dashed horizontal line*. We observed low pass (**a**) but also peaked (**b–e**) and complex tunings (**f**)

iso-intensity functions and not threshold-tuning curves (see Sect. 2, Fig. 8). They measure how the activity of the output unit changes with frequency. The resulting frequency tuning curves were divided into six classes (Table 1):

Class 1: untuned (responses at all frequencies at least 50% of the maximum, Fig. 8d).

Class 2: peaked (one peak, responses at highest and lowest frequencies below 50%, Fig. 8c).

Class 3: low pass (responses at low frequencies are high, and those at high frequencies are below 50% of the maximum, Fig. 8a,b).

Class 4: high pass (responses at high frequencies are high, and those at low frequencies are below 50% of the maximum).

Class 5: trough (response below 50% in middle range, and high at the highest and lowest frequencies, Fig. 8e).

Class 6: complex (tuning curves having several peaks and troughs, Fig. 8f).

All six classes could be found in the tuning curves (Table 1). While the high pass and peaked class were rare, the other four classes were quite abundant in all three net types.

3.2.2 Disparity tuning and vector strength tuning curves. The disparity tuning to sinusoidal gratings of different frequencies provides further insight into the mechanism of disparity tuning. To measure the degree of disparity tuning, we computed the VS of the modulation of disparity response curves for each output unit. Disparity response curves for sinusoidal stimulation were generally

real cells. Disparity tuning over frequencies, i.e., also scale, examined by phase-frequency plots showed deviations from energy neuron behavior.

Responses to high spatial frequencies showed weaker disparity tuning than responses to low frequencies (compare Fig. 7a–d to e–l). Also, as shown in the examples in Fig. 7, the disparity of the response maximum was sometimes frequency dependent. This observation will be analyzed in greater detail below.

3.2.1 Frequency-tuning curves. Frequency tuning was determined from the response magnitudes to sine gratings of the same amplitude, and thus represented
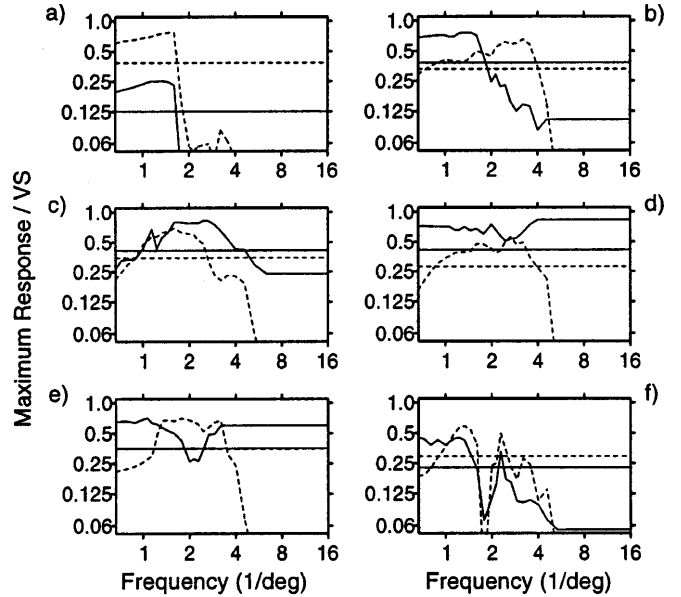
Table 1. Classification of frequency-tuning curves. Maximum and vector strength (*VS*) tuning curves of all three network types were classified as untuned, low pass, high pass, trough, peaked or complex

Net type/class			Untuned	Low pass	Trough	High pass	Peaked	Complex
Hybrid type	Max	N	18	15	17	2	4	19
	Max	%	24	20	22.7	2.7	5.3	25.3
	VS	N	0	12	0	0	39	24
	VS	%	0	16	0	0	52	32
Position type	Max	N	27	17	7	0	7	17
	Max	%	36	22.7	9.3	0	9.3	22.7
	VS	N	0	31	0	0	14	30
	VS	%	0	41.3	0	0	18.7	40
Phase type	Max	N	24	20	16	1	2	12
	Max	%	32	26.7	21.3	1.3	2.7	16
	VS	N	0	54	0	0	10	11
	VS	%	0	72	0	0	13.3	14.7
Total	Max	N	69	53	41	2	13	47
	Max	%	30.7	23.6	18.2	0.9	5.8	20.8
	VS	N	0	97	0	0	64	64
	VS	%	0	43.1	0	0	28.4	28.4

tuned. However, especially if the stimulus frequency was high, disparity tuning was sometimes weak or even absent (Fig. 7a,b). We used $P < 0.05$ in the Rayleigh test as a criterion for disparity tuning, and measured tuning curves at 27 different wavelengths ranging from 0.065 to 1.5 degrees (Fig. 7). Averaged over all networks, 16.9 (ph: 18.33, pos: 16.8; hyb: 15.56, with medians ph: 17, pos: 17, hyb: 19) of these curves exhibited tuning. In some nets, however, less (min: ph 13, pos 8, hyb 4) disparity response curves were tuned, while in others more (max: ph 21, pos 22, hyb 21) frequencies lead to tuned disparity curves. In those disparity curves that were tuned, the shape of the tuning curve varied. A typical tuning curve had one peak, the width of which was frequency dependent. The response flanks to the sides of the peaks tended to be asymmetrical (Fig. 7m,n). Maxima in the medium-frequency range were often very pronounced, and widths were smaller than those of the sinusoidal inputs (Fig. 7e–k). At high and low frequencies, the opposite effect was found (Fig. 7a–d and Fig. 7m,n).

By plotting VS as a function of stimulus frequency (Fig. 8) we could classify the observed VS-tuning curves according to the same classification scheme as used for the frequency-tuning curves. For VS-tuning only the low-pass, peaked and complex class occurred (Table 1). The low-pass class was the most prominent (Figs. 7, 8). As illustrated in Fig. 8, even those nets showing a tuning of the complex or peaked class had a clear cut-off frequency above which the VS monotonically decreased to zero. The cut-off frequency is a measure for the highest frequency that still has an effect on the disparity tuning of the output. Higher frequencies could only produce a baseline effect.

The correlation between the cut-off frequency, defined as the highest tested frequency at which the VS first reaches 50% of the maximum VS, and the absolute value of the preferred disparity of a network's teacher was weak (hyb: $r = -0.33$, pos: $r = 0.00$; ph: $r = -0.2$).

This was counterintuitive because narrow filters with relatively high best frequencies produce uncorrelated signals at large disparities and should, therefore, be eliminated during the training. To answer the question whether the cut-off frequency depended on the properties of the teacher function, we conducted some extra tests. We varied the widths and the center of the Gaussian teacher (i.e., the preferred disparity), and determined the cut-off frequency.

We trained a total of 72 hybrid-type networks to disparities between 0 deg and 1.13 deg (0, 0.25, 0.5, 0.75, 0.94, 1.13 deg, 12 networks for each disparity value) and determined the cut-off frequency. For this purpose, we generated training noise stimuli as above, but we produced a set of stereograms with a wider range of disparities and with bigger disparity steps (maximum disparity: 1.875 deg, disparity step: 0.063 deg). We found a negative correlation (with $r = -0.42$) between the disparity of the teacher and the cut-off frequency of the net. The cut-off frequency declined from 5.4 cyc/deg at 0 deg disparity to 3.0 cyc/deg at 1.13 deg with a slope of 2.2 cyc/deg^2.

Secondly, to check for a correlation between teacher width and cut-off frequency, hybrid nets were trained with teachers having 0.13 deg preferred disparity and widths of 0.00078, 0.0078, 0.016, 0.031, 0.063 and 0.13 deg. The training involved eight nets at each width, a total of 48 nets. The result was a negative correlation with a $r = -0.73$ between the teacher width and the cut-off frequency of a net. The cut-off frequency declined from 4.9 cyc/deg at 0.00078 deg teacher width to 1.0 cyc/deg at 0.125 deg with a slope of 31 cyc/deg^2.

3.2.3 Phase-frequency plots. Phase-frequency plots are a means to analyze the way disparity is represented in the nets. As described in Sect. 2.3, different CD and CP values result from different combinations of monocular RFs. We assembled plots from the MP in the disparity curves and the corresponding stimulus frequencies

(insets in Fig. 7). Phase-frequency plots were analyzed separately for each type of network.

Testing for linearity (MSE test) rendered significance in all nets with the exception of one hybrid-type net that showed no tuned responses to sinusoidal stimuli. A step-like appearance of some plots, however, suggested that the MSE test was not sensitive enough to discriminate between different classes of phase-frequency relations. Within limited frequency ranges, phase sometimes even remained almost constant (Fig. 9b,d,f). Between such frequency regions, there was a change of CD or there were jumps to higher or lower phases following the network's overall CD. In these cases, the jumps were responsible for the overall slope. To obtain an objective criterion for the classification of these types of networks and phase-frequency relations, we applied a Runs test to the signs of the residuals of the plots. The test results afforded discrimination between two classes of phase-frequency plots. Those that exhibited significantly ($P < 0.005$) less changes in sign than expected for a random distribution were termed complex type, because of the appearance of their plots, those having no significant deviation from a random scatter were termed simple-linear type (Fig. 9a,c,e).

The fewest complex-type networks were found for the hybrid model ($P < 0.005$: 19%). Of the position-type nets, 32% were classified as complex type. The greatest number of complex-type plots were found for the phase-model nets (41%).
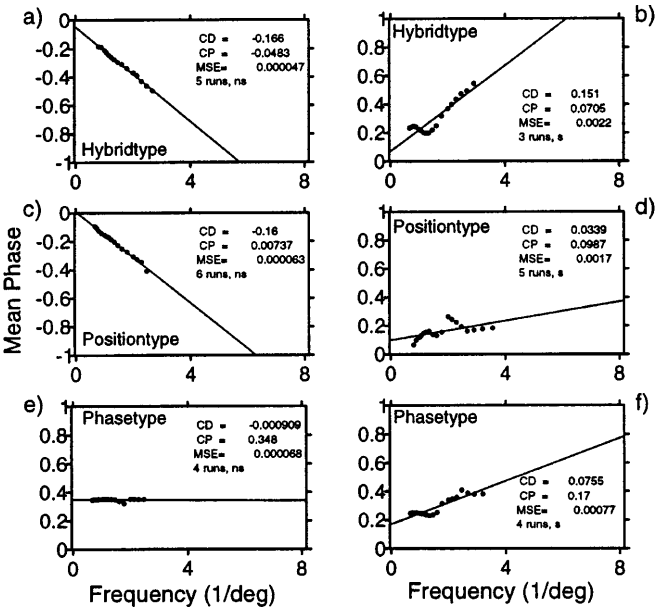
One can hypothesize that the frequency regions with nearly constant phase frequency are due to the behavior of units at the hidden layer that receive input from units tuned to a narrow range of spatial frequencies. To examine this, we tested all phase-type networks for which the inputs from the hidden layer to the output unit were set to zero for all but one frequency band. This was done for all bands. The resulting phase-frequency plots again exhibited the two subclasses simple linear and complex (data not shown).

3.2.4 Distributions of characteristic disparity and characteristic phase. To measure the contribution of position and phase shifts to the disparity tuning of the output units, we calculated their CD and CP. Although predictions concerning CD and CP were originally made using the energy model, they provide information about position shifts and phase shifts for a broader class of binocular models. They also provide another way that we can compare the behavior of the neural nets with that of real V1 cells.

CD and CP were calculated from all phase-frequency plots except the one hybrid-type net that was not disparity tuned when tested with sinusoids. As outlined in Sect. 1, the predictions for the three models regarding CD and CP are different. Figure 9a,c and e show selected examples of plots of the three network types. Both the position-type and the phase-type examples (Fig. 9c,e) comply with the theoretically derived expectations for the phase-frequency relation. The plot of the position net in Fig. 9c has a zero CP but a non-zero CD and the plot of the phase type net in Fig. 9e shows zero CD but non-zero CP. A typical example of a hybrid-type net is shown in Fig. 9a. It shows, as do the majority of the hybrid-type nets, a phase-frequency plot very much like those from position nets.

When the distributions of CD and CP (Fig. 10) were analyzed for each type of model, we observed a distribution around zero for both parameters (Fig. 10). The CP distribution in the phase-type nets was broader than the CP distributions of the position-type and hybrid-type nets, which proved significant in the Levene test ($P < 0.0001$; Fig. 10a,c,e; Levene 1960). The distributions of CP in the nets representing the latter two types were similar (Fig. 10c,e). In contrast, the CDs were more or less evenly distributed within the range of disparities used in the learning process in the position-type and hybrid-type models. Again, no statistical difference was found between hybrid-type and position-type nets (Fig. 10d,f), while the distribution of CDs was smaller in the phase-type nets (Fig. 10b, $P < 0.0001$ Levene test).

Testing for correlations between CP/CD and the preferred disparities of the teacher, we found a positive correlation between the preferred disparity of the teacher of a net and its CP only for the phase-type nets (hyb: $r = 0.26$, pos: $r = 0.14$, ph: $r = 0.83$). For all three net types, the CD showed a positive correlation with the preferred disparity of the teacher (hyb: $r = 0.79$, pos: $r = 0.89$, ph: $r = 0.79$). However, applying a linear regression to our data, we did not only find a slope differing from unity for the phase and the
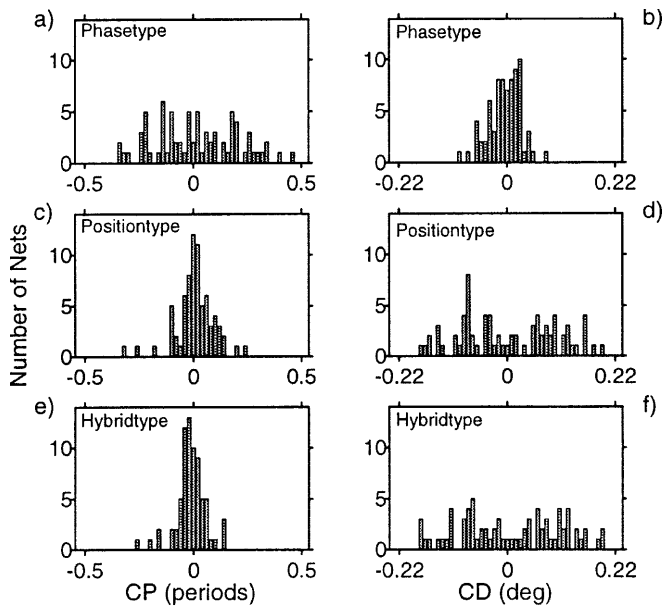

Fig. 9a–f. Examples of phase-frequency plots for three network types. All plots exhibited a significant relation between mean phase (MP) and frequency according to the MSE test. *Insets* show the characteristic disparity (*CD*), the characteristic phase (*CP*), the MSE, and the number of runs. The results of the Runs test are indicated by *s* (significant) or *ns* (not significant). **a** Typical hybrid-type simple-linear plot. **b** hybrid-type complex plot. **c** Typical position-type simple-linear example plot. **d** position-type complex plot. **e** Typical phase-type simple-linear example plot. **f** Phase-type complex plot

Fig. 10a–f. Distributions of characteristic phases and characteristic disparities for the three different net types. **a, c, e** Distributions of the characteristic phase (*CP*) for the phase- position- and hybrid-type networks. **b, d, f** Distributions of the characteristic disparity (*CD*), respectively

hybrid nets but also for the position-type nets, where the theory of the energy model predicts a slope of 1 (hyb: slope = 0.77, pos: slope = 0.77, ph: slope = 0.22). In contrast, when we restricted this analysis to those nets that were trained on the smaller range of teacher disparities of −0.09 deg to 0.09 deg, we found slopes much closer to the predictions of the energy model (hyb: slope = 0.97, r = 0.89; pos: slope = 0.95, r = 0.79; ph: slope = 0.27, r = 0.82).

The distributions of CD and CP resulting for phase-type networks under conditions where input to the output unit from all but one frequency band was set to zero were similar to those found for the whole networks. While the CP was evenly distributed over the whole range of possible phases, the CDs were concentrated around zero (data not shown).

4 Discussion

We trained our neural networks to respond to, and thereby represent, disparities just as real tuned excitatory neurons do (Pettigrew et al. 1968; Pettigrew and Konishi 1976; Poggio and Fischer 1977; Wagner and Frost 1993, 1994; Poggio 1995; Ohzawa et al. 1997; Anzai et al. 1997, 1999). Three types of networks were used to study the position and phase shift encoding.

A comparison with V1 cells and the energy neuron model revealed some differences that give insights into the mechanism of disparity coding.

The results of our simulations will be discussed with respect to (1) the representation of disparity, (2) the frequency tuning, (3) the relation between phase and frequency, and (4) the relation to biological neurons.

4.1 The representation of disparity

It is known that neural network implementations can extract disparities from stereograms (Becker and Hinton 1992). Neural networks have been used to investigate several aspects of the extraction of depth from disparities (Pouget and Sejnowski 1994; Gray et al. 1998). However, our study is the first to show that disparity selectivity based on Gabor filters can be achieved within a simple backpropagation paradigm. The structure of the energy model suggested this result for all three types of networks.

We interpreted the amplitude of the response of the output unit as a straightforward measure of the quality of the representation of disparity in our nets. As baseline responses of our networks are all very low, the response amplitude carries the same information as a signal-to-noise ratio. The mean amplitude was highest in the position-type networks. Differences between the hybrid-type and the position-type networks were less pronounced than differences between the phase type and the other two types (Fig. 3c–e). This was mainly a consequence of the phase type's low amplitudes for larger disparities (Fig. 4f). At zero disparity, where the three energy neurons are equivalent, no difference was found in our network types.

Although the hybrid-type networks had access to five times the number input units, and hence five times as many connections between the input and the hidden layers, position-type nets showed slightly higher responses than hybrid networks. This may have happened if hybrid nets had been hampered by irrelevant information that, first, had to be eliminated. Extra inputs would not be an advantage if all the relevant information was already contained in a subset of the input. The good results of the position type suggest that this was the case. Despite its overall lower performance, the phase type also solved the task by extracting most of the relevant information contained in the input.

All nets exhibited significant responses at wrong disparities when single input patterns were tested. This observation is consistent with results by Fleet et al. (1996a) and Qian and Zhu (1997). These authors suggested the use of spatial pooling to solve the problem of false responses (ghost matches). Besides the pooling over the spatial scale (frequency) done by the networks themselves, we have not pooled over cells at different positions, but we have effectively pooled over many input patterns. Our pooling is equivalent to pooling over a spatial position within one frequency band as long as the spatial pooling takes place over a sufficiently large spatial extent as to include statistically independent measurements. Thus, our results are in agreement with the earlier studies.

4.2 Frequency tuning

The frequency tuning of the output units in our nets (Fig. 8) was often broader than the mean bandwidths measured in neurons of the visual cortex of monkeys

Table 2. Main differences between network features and real cells and energy neurons. *aRDS* Anti-correlated random dot stereogram, *CD* characteristic/disparity, *CP*, characteristic phase

Feature	Neural network	Real neurons	Energy neurons
Disparity tuning at high disparities	Worse for phase-type networks	Tuning to disparities of more than 1 deg	Impossible for phase-type energy neurons
Responses to aRDS	No response	Responses invert, amplitude of response modulation can decrease	Responses invert, amplitude of response modulation does not decrease
Frequency tuning	Broad, sometimes complex	Smaller, normally band-pass-like	Band-pass-like, width depending on Gabor filter
Phase-frequency plots	Step-like plots, some phase-type networks show non-zero CD	No data	Linear, zero CD for phase-type neurons
Hybrid-type coding	Preference for non-zero CD versus non-zero CP	Cat data suggest a mixture; owl data favor CD	All possible combinations

and cats (Movshon et al. 1978; DeValois et al. 1982; Foster et al. 1985; Hammond and Fothergill 1994). As a consequence of the broad frequency tuning, our output units did not respond to aRDS (Fig. 9, Table 2). This could be expected for broadly tuned neurons with a non-linear output function (i.e., a threshold function) and demonstrates a difference between our nets and the recordings from neurons in the visual forebrain (Cumming and Parker 1997; Ohzawa et al. 1997).

The VS tuning of all networks was best described by low-pass or band-pass filters. A broader teacher function and higher preferred disparities resulted in lower cut-off frequencies. The latter can easily be understood. The spatial width of the input filters was related to the frequency of the filters by the width of the Gaussian envelope ($\sigma_i = 0.5/v_i$). At high disparities, the responses of two high frequency filters from the two eyes are essentially uncorrelated. This should lead to a suppression of higher frequencies when the nets were trained with larger disparities. The effect of the teacher width can also be explained. High frequency filters produce, as a result of the periodicity of their response, several peaks within the range of the maximum of a broad teacher. This also tends to make their responses irrelevant for the training task.

4.3 Relation between phase and frequency – differences from the energy model

The MSE test demonstrated the general linearity of the phase-frequency plots. However, using the MSE test over a large frequency range has some shortcomings. We found that nets that exhibited a complex relation of phase and frequency were still judged to be linear ($P < 0.005$, Fig. 10). Therefore, we analyzed all plots with the Runs test and, indeed, two subclasses, simple linear and complex, could be distinguished. While the linear phase-frequency plots can be interpreted as the result of a network that is homogenous over the entire frequency range, the complex phase-frequency plots require more discussion. One possible explanation for complex phase-frequency curves lies in the frequency tuning of the input. The distribution of the filters was

such that, at most frequencies, only one frequency band was dominant. Thus, this band should have dominated the response and, consequently, the disparity tuning and mean phase. Since all training procedures started from random weight values, complex plots can be either interpreted as an intermediate state, when training has not yet reached a possible final solution, or as a more complex solution to the demanded task. A phase-type network, for example, could use differences in CP between bands to imitate non-zero CD, and indeed, complex phase-frequency plots were mainly seen in the phase-type nets. To test this hypothesis we carried out some extra tests with the phase-type networks. We assembled phase-frequency plots, along with CD and CP distributions, when only one frequency band was connected to the output unit. However, the resulting phase-frequency plots were not linear with zero CD. Rather, again we observed two classes of phase-frequency curves. We conclude that complex plots might represent an alternative way of disparity coding that cannot be explained with the energy neuron model.

The distributions of CP and CD were different between the position-type and phase-type nets (Fig. 10). These differences, broad CP and peaked CD distributions and vice versa in the phase nets and the position networks, respectively, correspond to the theoretical expectations of the energy neuron model. Due to their generality, the behavior of CD and CP is not tightly constrained by the hybrid energy model. It was therefore surprising to see that the distributions of CD and CP obtained from the hybrid nets were similar to those of the position model. We interpret this finding as evidence for the more effective coding by position shifts, which was also seen in the analysis of Fleet et al. (1996a).

The results of our simulations are consistent with most predictions of the energy model. Yet, there are two important differences (see Table 2). First, energy neurons respond to aRDS, while the output units of our networks do not. Secondly, the energy model predicts a linear phase-frequency relation. Our networks agree with the latter property but only for restricted subranges of spatial frequency. This implies the existence of a mechanism for the encoding of disparities that differs from the energy mechanism.

4.4 Relation to biological neurons

One main difference from real cells is given by the broad frequency tuning of our networks. Together with the output non-linearity of our units, the frequency tuning can account for the absence of responses to aRDS, which is the second important difference to real cell responses in monkey V1. Interestingly, disparity-sensitive neurons in the owl with high response latencies respond also very weak to aRDS (Nieder and Wagner, personal communication).

There has been some discussion as to whether disparity is represented in the visual system of birds and mammals according to the phase-type, the position-type or the hybrid-type model (Wagner and Frost 1993, 1994; Anzai et al. 1997, 1999; Ohzawa et al. 1997). Although our simulations cannot replace experiments, and although we did not implement the energy model directly, our data might be helpful in the discussion of the effectiveness of the three types of energy neurons. The elements of our nets were biologically motivated. We restricted the model parameters concerning the input and the teacher of the nets to those stimulus qualities that are relevant for disparity sensitivity. This is an impoverished representation of the stimuli that animals are exposed to during development. Nevertheless, we assert that our model reveals general principles of disparity coding from bandpass monocular signals that are also valid for biological systems. Although there is some experimental evidence for the existence of backpropagation in biological neural networks (Fitzsimonds et al. 1997), the biological relevance of the backpropagation algorithm that we used here has been questioned. Still, the algorithm is a simple but very powerful tool for learning solutions to certain problems. The preference of hybrid-type nets to encode their best disparities mainly with non-zero CDs and to a lesser extent with non-zero CPs and the higher output amplitudes for position-type networks suggest that CDs are more robust than CPs in the encoding of disparities.

Nevertheless, hybrid-type coding with a prominent phase component seems to be realized in the cat visual system (Anzai et al. 1997, 1999). One could argue that the multitude of possible tasks of neurons in V1 and V2 could lead to constraints under which the stereo task only plays a subordinate role. This could force disparity sensitive neurons to develop a suboptimal solution for the stereo task.

References

Anzai A, Ohzawa I, De Angelis GC (1997) Neural mechanisms underlying binocular fusion and stereopsis: position vs. phase. Proc Natl Acad Sci USA 94: 5438–5443

Anzai A, Ohzawa I, Freeman RD (1999) Neural mechanisms for encoding binocular disparity: receptive field position vs. phase. J Neurophysiol 82: 874–890

Batschelet E (1981) Circular statistics in biology. Mathematics in biology. Academic Press, London

Becker S, Hinton GE (1992) Self-organizing neural network that discovers surfaces in random dot stereograms. Nature 355: 161–163

Campbell FW, Robson JG (1968) Application of Fourier analysis to the visibility of gratings. J Physiol (Lond) 197: 551–566

Cumming B, Parker A (1997) Responses of primary visual cortical neurons to binocular disparity without depth perception. Nature 389: 280–283

De Angelis GC, Ohzawa I, Freeman RD (1991) Depth is encoded in the visual cortex by a specialized receptive field structure. Nature 352: 156–159

DeValois RL, Albrecht DG, Thorell LG (1982) Spatial frequency selectivity of cells in macaque visual cortex. Vision Res 22: 545–559

Fitzsimonds RM, Song H, Poo M (1997) Propagation of activity-dependent synaptic depression in simple neural networks. Nature 388: 439–448

Fleet DJ, Wagner H, Heeger DJ (1996a) Neural encoding of binocular disparity: energy models, position shifts and phase shifts. Vision Res 36: 1839–1857

Fleet DJ, Heeger DJ, Wagner H (1996b) Modeling binocular neurons in primary visual cortex. In: Jenkin M, Harris L (eds) Computational and biological mechanisms of visual coding. Cambridge University Press, Cambridge, pp 103–130

Foster K, Gaska J, Nagler M, Pollen D (1985) Spatial and temporal frequency selectivity of neurons in the visual cortical areas v1 and v2 of the macaque monkey. J Physiol (Lond) 385: 331–363

Gray MS, Pouget A, Zemel RS, Nowlan SJ, Sejnowski TJ (1998) Reliable disparity estimation through selective integration. Vis Neurosci 15: 511–528

Hammond P, Fothergill L (1994) Cat striate cortex: monocular and interocular comparison of spatial-frequency selectivity. Ann Acad Bras Sci 66: 95–113

Levene H (1960) Robust tests for equality of variances. In: Olkin I et al. (eds) Contributions to probability and statistics. Essays in honor of Harold Hotelling. Stanford Univ. Press, Stanford, pp 278–292

Movshon J, Thompson I, Tolhurst D (1978) Spatial and temporal contrast sensitivity of neurons in areas 17 and 18 of the cat's visual cortex. J Physiol (Lond) 283: 101–120

Ohzawa I, De Angelis GC, Freeman RD (1990) Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors. Science 249: 1037–1041

Ohzawa I, De Angelis GC, Freeman RD (1997) Encoding of binocular disparity by complex cells in the cat's visual cortex. J Neurophysiol 77: 2879–2909

Pettigrew J, Konishi M (1976) Neurons selective for orientation and binocular disparity in the visual Wulst of the barn owl (*Tyto alba*). Science 193: 675–678

Pettigrew J, Nikara T, Bishop P (1968) Binocular interaction of single units in cat striate cortex: simultaneous stimulation by single moving slit with receptive fields in correspondence. Exp Brain Res 6: 391–410

Poggio GF (1995) Mechanisms of stereopsis in monkey visual cortex. Cereb Cortex 3: 193–204

Poggio G, Fischer B (1977) Binocular interaction and depth sensitivity in striate and prestriate cortex of behaving rhesus monkeys. J Neurophysiol 40: 1392–1405

Pouget A, Sejnowski TJ (1994) A neural model of the cortical representation of egocentric distance. Cereb Cortex 4: 314–329

Qian N (1994) Computing stereo disparity and motion with known binocular cell properties. Neural Comput 6: 390–404

Qian N, Zhu Y (1997) Physiological computation of binocular disparity. Vision Res 37: 1811–1827

Rumelhart DE, Hinton GE, Williams RJ (1986) Learning representations by back-propagating errors. Nature 323: 533–536

Siegel S (1956) Nonparametric statistics for the behavioral sciences. McGraw Hill. New York

Wagner H, Frost B (1993) Disparity-sensitive cells in the owl have a characteristic disparity. Nature 364: 796–798

Wagner H, Frost B (1994) Binocular responses of neurons in the barn owl's visual Wulst. J Comp Physiol A 174: 661–670

Yin TC, Kuwada S (1983) Binaural interaction in low-frequency neurons in inferior colliculus of the cat. III. Effects of changing frequency. J Neurophysiol 50: 1020–1042

Zhu Y, Qian N (1996) Binocular receptive field models, disparity tuning and characteristic disparity. Neural Comput 8: 1611–1641