Network: Computation in Neural Systems
June/September 2005; 16(2/3): 223–238

Taylor & Francis
Taylor & Francis Group

Learning of somatosensory representations for texture discrimination using a temporal coherence principle

JOERG HIPP,[1] WOLFGANG EINHÄUSER,[1,2] JÖRG CONRADT,[1] & PETER KÖNIG[3]

[1]*Institute of Neuroinformatics, University of Zürich & Swiss Federal Institute of Technology (ETH), Zürich, Switzerland,* [2]*California Institute of Technology, Division of Biology, Pasadena, USA, and* [3]*Institute of Cognitive Science, Department of Neurobiopsychology, University of Osnabrück, Osnabrück, Germany*

Abstract
In order to perform appropriate actions, animals need to quickly and reliably classify their sensory input. How can representations suitable for classification be acquired from statistical properties of the animal's natural environment? Akin to behavioural studies in rats, we investigate this question using texture discrimination by the vibrissae system as a model. To account for the rat's active sensing behaviour, we record whisker movements in a hardware model. Based on these signals, we determine the response of primary neurons, modelled as spatio-temporal filters. Using their output, we train a second layer of neurons to optimise a temporal coherence objective function. The performance in classifying textures using a single cell strongly correlates with the cell's temporal coherence; hence output cells outperform primary cells. Using a simple, unsupervised classifier, the performance on the output cell population is same as if using a sophisticated supervised classifier on the primary cells. Our results demonstrate that the optimisation of temporal coherence yields a representation that facilitates subsequent classification by selectively conveying relevant information.

Keywords: *Temporal coherence, somatosensory, whisker, vibrissal, slow feature analysis*

Introduction

The rodent vibrissae (whisker) system has been established as a standard model to study sensory processing (Woolsey & Van der Loos 1970; Gibson & Welker 1983a, 1983b). For rats, whiskers are an important source of sensory information (Metha & Kleinfeld 2004). Using solely their whiskers, rats can judge distances (Krupa et al. 2001) and discriminate objects (Brecht et al. 1997; Harvey et al. 2001) as well as textures (Guic-Roble et al. 1989; Carvell & Simons 1990). Of these capabilities, texture discrimination has recently received particular attention (Arabzadeh et al. 2003, 2004; Andermann et al. 2004; Moore 2004). In order to probe textures, rats actively move their whiskers across them. The whisker movements are detected by a population of primary sensory neurons, which are sensitive to the location and velocity of the whisker (Gibson & Welker 1983a, 1983b; Shoykhet et al. 2000; Deschenes et al. 2003). Based on these signals, rats perform quick and reliable discrimination at a level similar

Correspondence: J. Hipp, Institute of Neuroinformatics, University of Zürich & Swiss Federal Institute of Technology (ETH), Winterthurerstr. 190, 8057 Zürich, Switzerland. E-mail: joerg@ini.phys.ethz.ch

ISSN: 0954-898X print / ISSN 1361-6536 online
DOI: 10.1080/09548980500361582

to that of humans using their fingertips (Carvell & Simons 1990). Consequently, the sensory signals have to be transformed to allow the whisker system such a robust discrimination. Here we investigate whether this transformation, similar to representations of other sensory modalities, can be derived from a general coding principle.

In the context of visual processing, several general coding principles have been extensively studied (see Kayser et al. 2004 for a review). The principle of temporal coherence is based on the observation that different features of a natural stimulus vary on different time-scales. It uses the intuition that "relevant" stimulus features vary on a slower timescale than "irrelevant" components of the stimulus. By using this implicit information on the temporal structure of the input rather than an explicit teaching signal, temporal coherence forms an unsupervised (or rather "self-supervised") learning principle. This principle has been implemented in various ways, such as the trace rule (Földiak 1991), slow feature analysis (Wiskott & Sejnowski 2002) or "stability" (Kayser et al. 2001). Applying those implementations to natural scenes, model neurons replicate properties of V1 simple cells (Hurri & Hyvärinnen 2003), complex cells (Kayser et al. 2001; Einhäuser et al. 2002; Berkes & Wiskott 2005; Hashimoto 2003; Körding et al. 2004), colour selective cells (Einhäuser et al. 2003), and also face or object specific cells that are invariant to various transformations and thus resemble cell properties of higher visual areas (Wallis & Rolls 1997; Rolls & Milward 2000; Stringer & Rolls 2002). Consequently, the principle of temporal coherence has been demonstrated to be applicable to various levels of the visual hierarchy.

In the present study, we generalize the principle of temporal coherence beyond the visual modality. Since it has been previously demonstrated that temporal coherence can be implemented in a physiologically realistic framework using local learning rules (Földiak 1991; Wallis & Rolls 1997; Körding & König 2001), we here employ an abstraction of this principle. Following Kayser et al. (2001), we define a global objective function ("stability") that characterises the temporal coherence of cells as well as their mutual interaction. We train a population of cells on data that is obtained from a physical simulator of active whisker behaviour and compare the results to recent physiological findings. We demonstrate that, based on natural input statistics, neurons trained to optimise temporal coherence form a representation that is well-suited for somatosensory classification.

Methods

Input

Recording set-up. In order to probe objects in their environment, rats actively move their whiskers across them. We modelled this process using steel whiskers (steel wire, Small-Parts Inc., Miami Lakes, FL, USA), fixated with one end to a sensor head, which was mounted perpendicular to the rotation axis (Figure 1a). The steel whisker had a length of 7.30 cm and a diameter of 0.305 mm. By using a motor (Digital Servo S9251, Futaba, Huntsville, AL, USA) the whisker was rotated back and forth. The motor was controlled by a microprocessor (Atmega163 Atmel, San Jose, CA, USA) interfaced via the RS232 port to a PC. The whisking frequency was set to 1 Hz. To probe textures, we placed them, comparable to natural conditions, in a plane orthogonal in front of the whisker. We attached a tiny magnet near the base of the whisker. A magnetic field sensor (KMZ51 Philips-Semiconductors, Eindhoven Netherlands) recorded the distortions of the magnetic field induced by the movements of the whisker. The motor position was measured via a potentiometer. The motor position and the deflection of the whisker were digitalised via a data acquisition card (DAQCard-6036E, National Instruments, Austin, TX, USA) at a sampling rate of 4000 Hz.

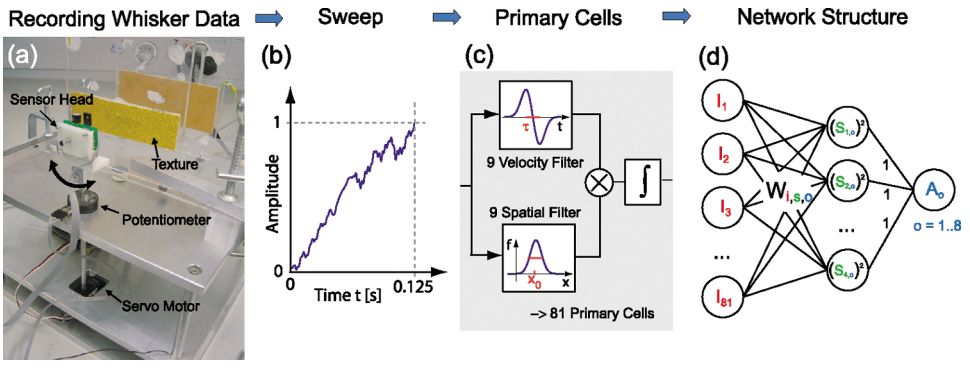

Figure 1. (a) Set-up for recording artificial whisker data; (b) portion of whisker signal used for further analysis ("sweep"); (c) schematic of filters representing primary cells; (d) network structure, I_i primary cell's activity, S_{so} subunit activity, A_o output cell activity.

Textures. We recorded whisker movements on eight sandpapers of different roughness, as defined by their so called *P*-values, which is a standard measure for roughness of sandpaper (http://www.fepa-abrasives.org/): we used *P*-values of 40, 60, 80, 100, 120, 150, 180 and 240. The average grain diameter of the roughest sandpaper (*P*-value of 40) was 425 μm, while the smoothest sandpaper (*P*-value of 240) consistsed of grains with average grain diameter of 59 μm.

Processing of the whisker signals. The magnetic sensor was mounted on the sensor head, which rotated to imitate the active whisking movement (Figure 1a). The movement led to a superposition of the deflection signal and the earth's magnetic field. To correct for this effect, we measured the contribution of the earth's magnetic field at each motor position and subtracted it from the signal. In one whisking cycle, a whisker moved across the surface, back and forth. Based on the motor position, the segments when the whisker touched the texture were selected. We refer to these segments as sweeps. A sweep consisted of 451 data-points, corresponding to 0.11 s. In total, we recorded 200 sweeps for each texture. Each sweep was then normalised to range from 0 to 1 in amplitude (Figure 1b). These signals served as input to simulated primary neurons.

Temporal properties of the input. When rats explore the environment using their whiskers, they repeatedly move them across objects of interest. Hence, it is most likely that two consecutive sweeps result from the same texture. In our main simulations, consecutive sweeps are taken from the same texture class. As a control, we additionally varied the probability that consecutive sweeps fall in the same class. The interval tested ranges from chance level (12.5%) to 100%. (100% can be reached only artificially when learning is switched off between class changes, of course.)

Input representation. The primary sensory neurons in the whisker system code for velocity and position of whiskers (Gibson & Welker 1983a, b; Shoykhet et al. 2000). We modelled the response of a population of primary cells to the whisker movements (Figure 1c). We filtered the signal with nine different spatial filters ("position"-filters) and nine different temporal band-pass filters ("velocity"-filters).

- The position filters are Gaussians with standard deviation 0.1 and mean (x_0) ranging from 0.1 to 0.9 in 0.1 steps (in units of the aforementioned normalised position).
- The velocity filters are formed by the difference of two Gaussians shifted by twice their standard deviation τ. τ ranges from 1.25 ms to 11.25 ms in steps of 1.25 ms. These filters are band-pass filters with different preferred frequencies (from 24 Hz to 219 Hz) and thus resemble low-pass filtered velocity sensors.

Each primary cell corresponds to the product of a position and a velocity filter, yielding 81 (N_i) different primary cells. Each sweep is filtered and integrated over the time of the sweep, such that each sweep results in one value per primary cell. These activities serve as input to the network. Although this is an abstract model of the primary sensory cells, it captures the main properties of spatial and velocity selectivity (see also the discussion).

Network

Neuronal model. As in previous studies of the visual system, we use a 2-layer network (Figure 1d). The activity of primary cells I_i ($i = [1,..,N_i]$) is used to train output cells. Each output cell receives input from several subunits (S_{os}). The activity of output cells A_o is given by

$$A_o = \sqrt{\sum_s S_{os}^2}, \quad S_{os} = \sum_i W_{iso} \cdot I_i \tag{1}$$

where S_s are the subunits' activities and W is the weight matrix between primary cells and subunits. Unless otherwise stated we use output cells with $N_s = 4$ subunits ($s = [1,..,N_s]$) and simulate $N_O = 8$ or $N_O = 81$ output cells.

Stability implementation of temporal coherence. In our implementation of temporal coherence, we follow the "stability" objective function approach (Kayser et al. 2001). The objective function for a single neuron is given by the mean squared temporal derivative of its activity normalized by the variance over time:

$$\Psi_{\text{Stability}} = -\frac{1}{N_o}\sum_{o=1}^{N_o} \psi_{\text{Stability},o} = -\frac{1}{N_o}\sum_{o=1}^{N_o} \frac{\left\langle \left(\frac{d}{dt} A_o\right)^2 \right\rangle_t}{\text{var}(A_o)} \tag{2}$$

where $\langle . \rangle_t$ denotes temporal averaging and var the variance over time. The temporal derivative is implemented as finite difference. This part of the objective function favours neurons, whose activity varies rarely (dA_o/dt is small), and at the same time avoids the trivial solution of constant activity by penalizing low variance. Hence, it presents one (of several) possibilities to formalize the concept of temporal coherence by a global objective rather than a local learning rule. To introduce a lateral interaction, we add a de-correlation term to this function that prevents neurons from obtaining similar response properties.

$$\Psi_{\text{Decorr}} = -\frac{2}{N_o(N_o-1)} \sum_{o_1} \sum_{o_2 > o_1} \frac{\text{cov}(A_{o_1}, A_{o_2})^2}{\text{var}(A_{o_1}) \cdot \text{var}(A_{o_2})} \tag{3}$$

where cov denotes the covariance.

The two objectives are added linearly with a weighting factor:

$$\Psi = \Psi_{\text{Stability}} + \beta \Psi_{\text{Decorr}} \tag{4}$$

Unless otherwise stated we weigh both objectives equally, i.e., $\beta = 1$.

We analytically determine the gradient for the objective function. The complete objective function is then optimised with respect to the weights by using the "RPROP" algorithm (Riedmiller & Braun 1993).

Data analysis

Optimising the stability objective transforms the representation given by the primary cells in a representation given by the output cells ("stability transformation"). When analysing such a transformation, two issues need to be addressed. First, does the transformation extract the information that is relevant for classification such that the performance of a simple classifier improves? Second, does the transformation destroy any information, such that the performance of an ideal classifier worsens? Consequently, we assess the classification performance of primary and output cells using classifiers of varied complexity. We compare the performance of the output cells to various other transformations of the primary cell representation.

To test the generalisation of the transformations, we split the dataset of 200 sweeps per class into a training set (2/3, 133 sweeps/class) and a test set (1/3, 67 sweeps/class). Using the training set, we train the output cells and determine the parameter for the supervised classifiers. Using the test set, we evaluate the classification performance. To increase the training set, we randomly draw 500 pairs of sweeps for each class. In additional simulations, we verified that this amount of re-sampling is sufficient and further re-sampling does not markedly improve performance on the test-set (data not shown).

"Separability index". In order to quantify how well a cell can separate different texture classes, we define a separability index (SI) for each cell. We measure the variance of a neuron's response within each texture class and average over classes. We divide this measure by the total variance of the neuron's activity. We define the separability index as 1 minus this ratio. This measure approaches 0 if classes are indistinguishable and 1 for perfect separability.

"Fisher transform". We compare the transformation based on stability to other common transformations, such as principal component analysis (PCA) and a generalisation of Fisher's linear discriminant to multi dimensions (Bishop 1995), which we will refer to as "Fisher transform". While PCA is an unsupervised transform, the Fisher transform is supervised in that it incorporates the class information. Given a data set with m classes $(c = 1 \ldots m)$ and n_c samples in class c, we compute the class specific means μ_c and the overall mean μ. In the following notation we use column vectors. Next, we determine the within class covariance matrix,

$$S_W = \frac{1}{m} \sum_{c=1}^{m} \frac{1}{n_t} \sum_{s=1}^{n_{ct}} \left(x_c^s - \mu_c\right) \cdot \left(x_c^s - \mu_c\right)^T \tag{5}$$

and the between class covariance matrix.

$$S_B = \sum_{c=1}^{m} (\mu_c - \mu) \cdot (\mu_c - \mu)^T \tag{6}$$

By taking the quotient of the between class covariance matrix to the within class covariance matrix yields the following eigen-problem.

$$\lambda_i \cdot S_W^{-1} \cdot S_B = \lambda_i \cdot \omega_i \tag{7}$$

The eigenvectors ω_i build a new set of basis vectors. Arranging the basis vectors ω_i as a transformation matrix W, the solution maximises the Fisher criterion *trace* $(W\, S_W^{-1}\, S_B\, W^T)$. The importance of the new basis vectors scales with the corresponding eigenvalue. The vector with the largest eigenvalue points in the direction where the ratio of variance between class centres to the variance within classes is maximal. The maximal number of basis vectors ("Fisher components") is the number of classes minus one, in our case, 7.

Classification. To determine the classification performance on different representations of the whisker input, we here use three different classifiers; plain Euclidian distance ("Euclidian"), Gaussian density estimation ("Gaussian") and K-means clustering ("K-means").

- For the "Euclidian" classifier, we first compute the centres of all classes based on the labelled data set. Then we attribute each sample to the class with the closest mean using the Euclidian distance measure.
- For the "Gaussian" classifier (i.e., Gaussian density estimation), we approximate the class probability distributions by a multi dimensional Gaussian. Given a set of feature vectors ϕ_t^i, e.g. the vector components after the Fisher transform, relating to classes c $(c = 1 \,..\, m)$ and sweep s $(s = 1 \,..\, n_c)$, we compute the class specific means μ_c and covariance matrices S_c. Note, that—even after the Fisher transform—the covariance matrix is not necessarily diagonal.
  This yields an estimation of the probability distribution of the feature vectors for each class.

$$p_c\left(x_c^s\right) = \frac{1}{(2\pi)^{(m-1)/2} \cdot \det(S_c)^{1/2}} \cdot e^{-\frac{1}{2}(\phi_c^s - \mu_c) S_c^{-1} (\phi_c^s - \mu_c)^T} \tag{8}$$

  For classification we attribute each sample to the class that has the highest value of its probability density function at the sample's position.
- For "K-means" clustering, we use the Matlab implementation of the algorithm with the Euclidian distance measure. The results given are the optimal classification results over 10 different random initial positions of cluster centres.

The Euclidian and Gaussian classifiers are supervised methods; first the distributions are estimated based on the samples. Then the samples are attributed to a class based on the estimated distribution. The K-means clustering, in contrast, is unsupervised since no learning on labelled data takes place. As this study focuses on the unsupervised generation of representations suited for classification, but not on the classification as such, we, however, specified the amount of cluster centres to be equal to the number of classes (eight).

Performance measure. The classification result can be represented as a matrix (H), that contains for each sweep from texture class $t = i$ the probability to be assigned to a class $c = j$.

$$H_{i,j} = p(c = j \mid t = i) \tag{9}$$

To quantify the *classification performance* as a single value we compute the fraction of correct classified patches (CC).

$$CC(H) = \frac{\sum_i H_{i,i}}{\sum_i \sum_j H_{i,j}} * 100\% \tag{10}$$

Response profiles. In order to analyse the responses of a single cell in velocity–position space, we generate test stimuli akin to the filters used for primary cells. The covered range is identical to the one spanned by the primary cells, while the sampling is about four times more dense (32 samples for each dimension). The generated test stimuli are applied to the converged network in the same manner as the training stimuli.

Results

Single cell analysis

We recorded the responses of simulated primary cells when stimulated by signals recorded with a set-up that mimiced rats probing textures with their whiskers. The responses of these primary cells exhibit a high variability within the same texture (Figure 2a, b). We quantify how well distinct texture classes can be separated by a separability index (SI, see Methods). On the training set, we find a low SI for the primary cells (0.35 ± 0.15, Figure 2a). Using these cells as input, we train eight output cells to optimise the stability objective. These output cells show a comparably low within class variability and thus a significantly higher SI (0.78 ± 0.02, $p < 10^{-12}$, t-test, Figure 2c). More importantly, this significant difference is conserved for the test-set, i.e., for sweeps never seen by the network during training. Here

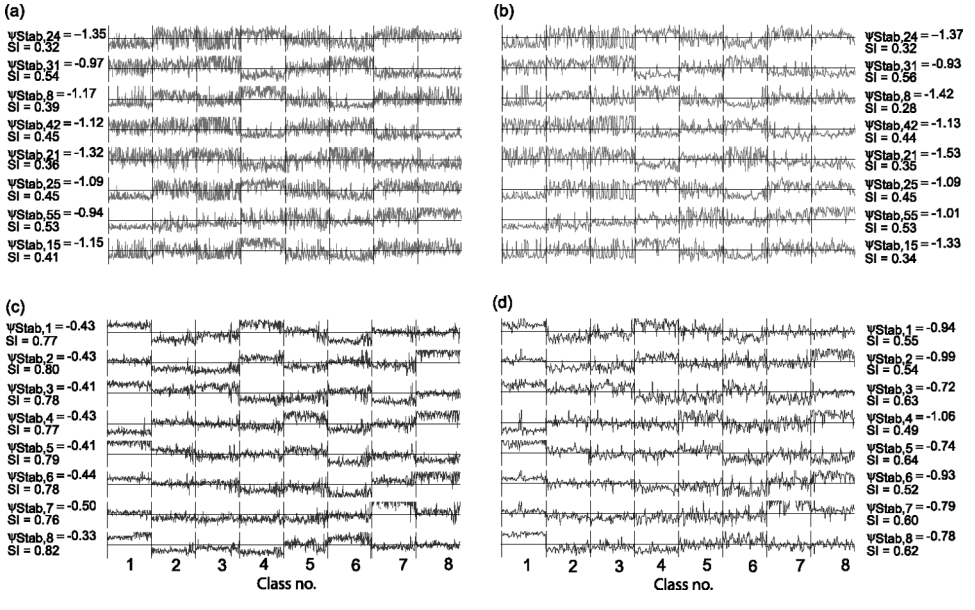

Figure 2. Activity traces of 8 randomly chosen primary cells of (a) training and (b) test data set. All eight output cell activities of the default simulation with $N_o = 8$ for (c) training and (d) test set. Ordinates of each cell's activity trace are scaled to ± 3 standard deviations around the cell's mean activity. Values for separability index and individual stability are given adjacent to each activity trace. Vertical lines indicate change of stimulus class.

the SI reaches 0.34 ± 0.14 for the primary cells (Figure 2b), but 0.57 ± 0.06 for the output cells (Figure 2d). This difference between primary and output cells is highly significant ($p = 1.3 \times 10^{-5}$, *t*-test). Consequently, optimising stability does not only increase the separability index for the training data, but also for previously unseen test data. Since such generalisation is a desired property of any learning algorithm, all analysis mentioned hereafter will refer to the unseen test set only. Since the difference between primary cells and output cells arises from optimising stability, we test whether the individual stability ψ_i of a cell is correlated to its SI. Indeed, we find a highly significant correlation for primary cells and output cells ($r = 0.973$ and $r = 0.966$ respectively, $p < 10^{-4}$). The observed difference in SI thus results from optimising stability and suggests that cells optimised for stability may be better suited for texture classification than primary cells.

For simulations with eight output cells ($N_O = 8$) and 81 output cells ($N_O = 81$), we quantify the single cell classification performance of primary cells and output cells by a simple classifier; the Euclidian distance in the space of activities. We find classification performance of a cell to be correlated to its stability, with a single regression line for primary cells and both output cell populations ($r = 0.87$, $p < 10^{-50}$, Figure 3). This correlation is also significant for each of the populations with 81 cells alone (primary cells: $r = 0.78$, $p < 10^{-17}$; output cells: $r = 0.68$, $p < 10^{-11}$). This result shows that cells with higher stability are better suited for classification. The classification performance of output cells ($34.9 \pm 1.0\%$, $N_O = 8$ and $34.1 \pm 4.2\%$, $N_O = 81$), which are trained to optimise stability, is indeed significantly larger than classification performance of primary cells ($24.7 \pm 3.9\%$; $p < 10^{-10}$, t-tests for $N_O = 8$ and $N_O = 81$). This result confirms that cells trained to optimise stability are well suited for simple classification.

Optimising the stability objective provides one specific transformation of the primary cells into output cells. How do the obtained output cells compare to other transformations of the primary cells? A transformation commonly applied to pre-process data for classification is principal component analysis (PCA). PCA decomposes the input into de-correlated dimensions, which are sorted according to the amount of variance explained. We apply PCA to the primary cells' responses. The obtained principal components also show the correlation between single cell classification performance and stability ($r = 0.82$, $p < 10^{-19}$, Figure 4,

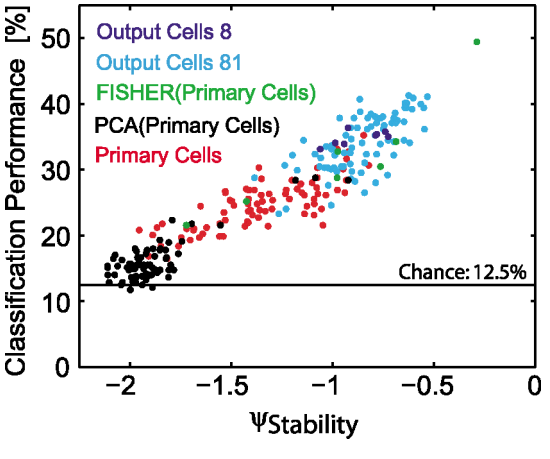

Figure 3. Single cell classification performance versus individual stability of each cell. Red: primary cells; blue: output cells ($N_O = 8$); cyan: output cells ($N_O = 81$); black: first 75 PCA components of primary cells; green: Fisher components of primary cells.

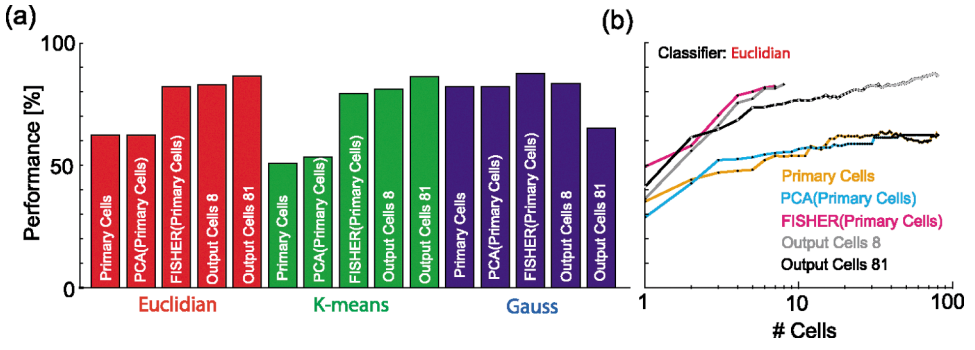

Figure 4. (a) Total classification performance of cell populations for different combinations of classifiers and transformations. Different colours represent different classifiers. (b) Cumulative classification performance (Euclidian classifier) for increasing number of dimensions. Dimensions are sorted by decreasing single cell classification performance. Different colours represent different transformations.

black). Stability and classification performance decrease for higher order components with the first three principal components being the three best classifying and three most stable cells. Even the first principal component, however, is worse than the average performance of output cells. This indicates that optimising stability is better suited to achieve good classification performance with a small number of cells than PCA. As a more sophisticated transformation, we apply the multi-dimensional generalisation of Fisher's linear discriminate ("Fisher transformation", see methods). Unlike optimising stability or PCA, the Fisher transformation requires supervised knowledge of the class structure of the input. Not surprisingly, the first Fisher component is more informative (49.4% classification performance, Euclidian classifier) than individual output cells. However, the next four Fisher components yield a similar performance as the output cells (34.3%, 30.5%, 32.8%, 28.8%) and the remaining two components perform remarkably worse (25.2%, 21.6%). This suggests that the optimisation of stability, which does not employ any explicit teaching signal, forms a representation that is comparable in classification performance to the supervised Fisher transform.

Population analysis

As a next step, we move from an investigation of single cell classification performance to the population level. We now investigate whether output cells still outperform primary cells and transformations of primary cells on the population level. We first use the same measure as above, classification based on Euclidian distance. When comparing the classification performance of 81 output cells to the 81 primary cells, we find performance to be dramatically increased from 62.3% to 86.6% (Figure 4a, red bars). Already the population with eight output cells, however, nearly reach this high performance (83.0%). This demonstrates that a small number of output cells is sufficient for good classification. Next, we test the effect of the aforementioned transformations on population performance. Since principal component analysis does not affect Euclidian distances, classification performance is identical for the principal components of the primary cells. Applying a Fisher transformation, we obtain increased performance (82.2%), which is still slightly smaller than the performance observed for output cells. These results demonstrate that optimising stability forms a representation that is well suited for classification with a simple classifier.

Since no transformation can generate information, no transformation can increase the performance of an ideal classifier. In order to test whether optimising stability loses relevant information, we use a more sophisticated classifier (Gaussian density estimation). This supervised classifier, which takes the input's covariance structure into account, reaches 87.7% performance on the Fisher transformed primary cell population (Figure 4a, blue bars). This is comparable to the performance of 86.6% using the simple Euclidian classifier on the large output cell population. This argues that optimising stability does not discard relevant information.

Using the sophisticated classifier on the large output cell population leads to a strong decrease in performance (down to 65.2%). This can be understood as an effect of overfitting. The small output cell population is not affected since the small number of cells acts as regularisation. However, using a simple classifier prevents overfitting as shown above. So far, we have shown that optimising stability allows comparable good performance of supervised classification algorithms.

To test whether the obtained representation is also suitable for classification without knowledge of the underlying class structure, we next investigate the performance of an unsupervised classifier. Closely related to the supervised Euclidian distance classifier, we perform unsupervised K-means clustering with the Euclidian distance measure. While performance for primary cells further drops as compared to the supervised measure (51.5%, Figure 5a, green bars), the performance of output cells stays nearly unaffected (81.3% for $N_O = 8$ and 86.4 % for $N_O = 81$). Also principal component analysis (51.9%) and Fisher transformation (79.4%) perform worse than output cell populations. Hence, optimising stability also facilitates unsupervised classification.

A goal of transformations like PCA or the Fisher transformation is to carry a large amount of the available information in a small number of ordered dimensions. Hence, we here compare the cumulative performance of increasing numbers of primary cells, output cells, principal components and Fisher components using the Euclidian classifier (Figure 4b). To obtain a common sorting principle, we sort the cells by their individual classification performance. This measure is closely correlated to the measures of individual stability (see above). It also correlates to the amount of variance explained by a principal component ($r = 0.76$, $p < 10^{-13}$) and it decreases for Fisher components of increasing order. Hence, sorting by single cell classification performance is in all cases similar to using the "intrinsic" order of the respective transformation. We find that the Fisher components and the output cells reach a high performance with just a few cells. The performance of the larger output cell populations, however, has a slightly smaller slope, reaching the same performance as the small output cell population with approx. 30 cells. This can be understood as cells in a large population mutually prevent each other from becoming optimal with respect to the stability objective, while as a population they still keep all the information. Thus, stable populations succeed to represent the input in a compact way comparable to that of the supervised Fisher transform.

In summary, we observe that unsupervised and supervised classifiers work well on the representation formed by the output cells. Unlike for the primary cells and the other transformations tested, this performance is optimal for the simple supervised Euclidian and even the unsupervised K-means classifier. This indicates that the representation obtained by optimising stability is indeed well-suited for classification by downstream areas.

Parameter dependence

The principle of temporal coherence is based on the assumption that the probability of two successive trials (e.g., whisker sweeps) being performed on the same class, texture or

object is higher than chance. Here, we test the dependence of the stability transformation on this property of natural input. We systematically vary the probability of successive sweeps being from the same class, i.e., the temporal coherence of the input signal, from chance (12.5%) to 100%. We find that classification performance ($N_O = 8$) increases monotonically with the temporal coherence of the input (Figure 5a) and saturates at about 80%. This demonstrates that—while the stability objective indeed uses the temporal coherence of the input—it is nevertheless robust to changes between classes occurring reasonably frequent.

To investigate how the choice of input filters influences the results we repeated our simulations with a smaller primary cell population. Instead of 9 spatial $\times$ 9 temporal filters we here use 16 filters (4×4). To obtain similar input space coverage as in the main simulations, we double the width of the filters' tuning. For the Euclidian distance classifier, optimising stability increases the performance from 55.5% (primary cells) to 67.2% (output cells, $N_O = 8$), which is comparable to the increase observed for 81 primary cells above. A similar increase is observed for the K-means clustering algorithm (primary cells: 47.7%; output cells: 64.6%). These results demonstrate that the gain for simple classifiers obtained by optimising stability is not critically dependent on the number and tuning of primary cells.

To test the influence of the network complexity on our results, we vary the number of subunits (N_S) and the number of output cells. For our default choice of parameters ($N_S = 4$, $N_O = 8$), the performance is nearly saturated (Figure 5b). This result is qualitatively independent of the classifier used (data not shown). These findings justify the choice of our default parameters.

The objective function consists of 2 competing terms; the stability term and the decorrelation term. In our default simulation, both terms are weighted equally. Here, we investigate the dependence of the result on this weighting factor (β). We find that the performance peaks at about $\beta = 1$ but is stable for about four orders of magnitude (Figure 5c, top). The peak coincides with both parts of the objective function taking a high value after convergence (Figure 5c, bottom). In contrast, the stability term dominates for too small β, which results in stable but too similar cells. This has an effect akin to reducing the number of cells and consequently lowers the performance. On the other hand, for too high values of β, the decorrelation term dominates, which results in dissimilar cells of low stability. As we have shown above, stability is correlated to the performance and thus a high β also decreases performance. Consequently, our default relative weighting of the objective functions' terms is in the optimal regime, but a precise tuning is not critical.

Response profiles

In order to further characterise the cells' properties, we probe their responses in position–velocity space. By construction, primary cells have a single peak in this space (Figure 6a). In contrast, output cells show a comparably complex response pattern: each output cell's response has multiple peaks and troughs, which cover the whole stimulus space (Figure 6b). This demonstrates that cells trained to optimise stability integrate information over a variety of positions and velocities to achieve their robust performance.

Discussion

To make a decision requires a chain of operations that starts with encoding the stimulus and extracting relevant features, continues with a comparison to prior knowledge and eventually

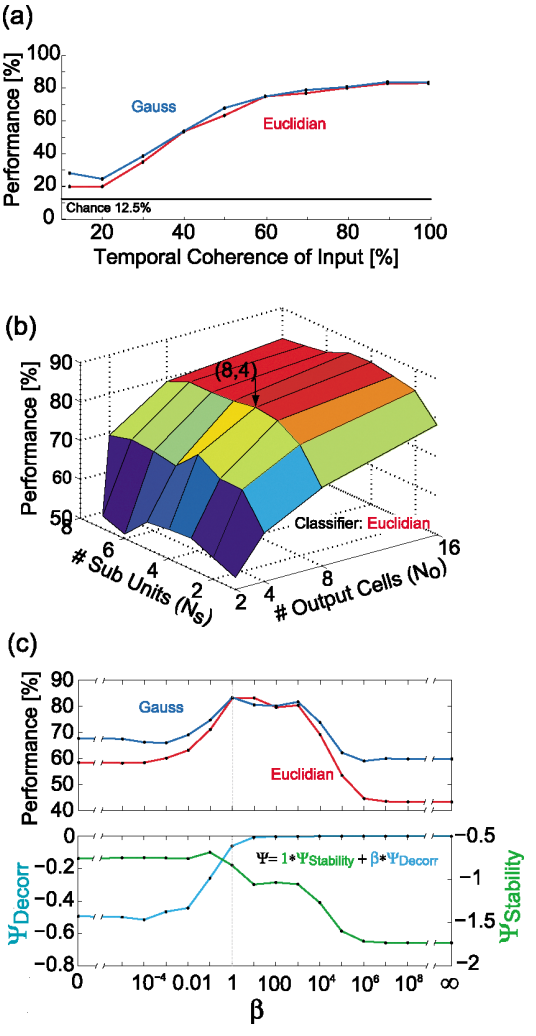

Figure 5. (a) Classification performance for varying temporal coherence (probability of successive sweeps being from the same class) of the input. (b) Classification performance (Euclidian classifier) for varying network parameters N_O and N_S. Default values indicated by arrow. (c) Dependence of classification performance (top) and objective function values (bottom) on the relative weight β of both parts of the objective functions.

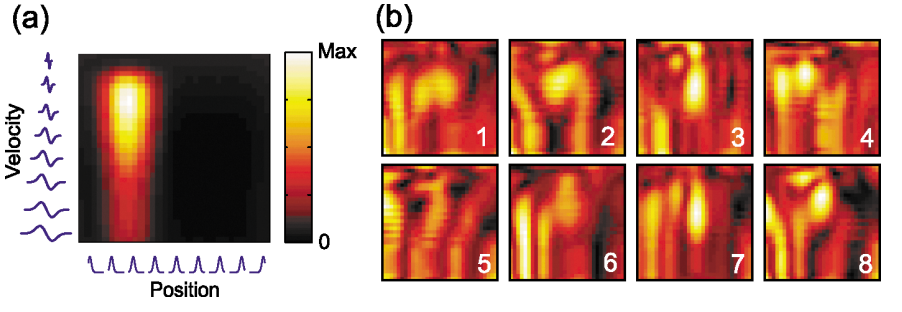

Figure 6. (a) Response profile of a single primary cell. Plots at the margin represent corresponding filters. (b) Response profiles of all 8 output cells of the default simulation with $N_O = 8$.

leads to the appropriate motor output (Romo & Salinas 2001). Apparently, the extraction of relevant features is a crucial link in this chain. Here, we show that the principle of temporal coherence allows a neuronal population to extract features meaningful for discrimination by solely exploiting the natural statistics of the input, unsupervised. The sensory input is represented in a way that differences between discriminanda (textures) are pronounced, which facilitates subsequent fast and robust classification. Hence, the general coding principle of temporal coherence can account for a robust basis of decision-making.

So far, the principle of temporal coherence had been applied almost exclusively to visual processing. A large number of studies have addressed how various implementations of temporal coherence relate to neuronal properties in primary visual cortex (Földiak 1991; Stone 1996; Einhäuser et al. 2002; Wiskott & Sejnowski 2002; Hurri & Hyvärinnen 2003; Hashimoto 2003; Körding et al. 2004; Berkes & Wiskott 2005) and the inferotemporal cortex (Wallis & Rolls 1997; Rolls & Milward 2000). In addition, the temporal coherence principle can be employed in unsupervised object recognition systems (Stringer & Rolls 2002; Einhäuser et al. 2005), and under some conditions even outperforms supervised learning schemes (Wallis 1996). We here demonstrate that the principle of temporal coherence is not restricted to the visual domain, but can be generalized to another sensory modality.

The prerequisite for any temporal coherence algorithm is that the information to be extracted varies on a slower timescale than the irrelevant transformations that should not interfere with the classification task. In the visual domain, Wallis and Baddeley (1997) derived optimal parameters for the trace rule implementation of temporal coherence to optimally adapt to the temporal coherence of the input. In the present case, we measured the probability needed for subsequent sweeps to be from the same class for temporal coherence to improve classification. We find a monotonic increase of performance with temporal coherence, which saturates at about 80%. This value corresponds to a rat "whisking" on average eight successive times across the same texture before moving on. At an average "whisking" frequency of 8 Hz (Carvell & Simons 1990) this corresponds to 1s and is thus compatible with rat behaviour. This result demonstrates that the optimisation of stability is indeed applicable to realistic learning scenarios in rodents.

To obtain realistic somatosensory input, we used a hardware model to mimic a rat that actively employs its whiskers across different textures. Although the model whisker was made of steel and the whisking frequency was lower (1 Hz) than in the rat (5 to 15 Hz, Carvell & Simons 1990), the artificial system captured the basic properties of natural whisker behaviour. Especially the whisker's amplitude and velocity were modulated by the texture surface while sweeping across, which induced a texture specific pattern. Indeed, we found in a different study that the relevant properties of real whiskers are identical to those of the model (Hipp et al. unpublished observations). In addition, the crucial result, that the required temporal coherence in the input is consistent with rat behaviour, is unaffected by the precise implementation. Hence, we feel confident that our input signals simulated natural whisker behaviour sufficiently well for the purpose of the present study.

Primary sensory neurons in the rat vibrissae system code velocity and position of the whisker (Gibson & Welker 1983a, b). To capture the essence of this information without specific assumptions, we use generic spatio-temporal filters as input to our network. The cell model used for the output neurons resembles a multi-subunit energy detector akin to the two-subunit energy-detector (Adelson & Bergen 1985), which was previously applied to the learning of complex cell properties by using the same stability objective function (Kayser et al. 2001; Körding et al. 2004). In the context of the visual system, it has been demonstrated that the choice of the non-linearity is not critical and that it can even be learnt simultaneously to the weights (Kayser et al. 2003). Here, we demonstrate that the number of

subunits is not critical either and the results are robust to its modification. In a physiological context, the subunits might be interpreted as distinct dendritic compartments or as a set of afferent cells. This property makes the model applicable to a vast number of physiological processes. The learning rule itself may also be implemented in physiological hardware. The de-correlation part of the objective can be mediated by strong lateral inhibition. The principle of temporal coherence itself can also be implemented by a physiologically realistic learning rule (Körding & König 2001; Einhäuser et al. 2002). In summary, all stages of the present model are compatible with physiological mechanisms.

Most electrophysiological experiments addressing texture discrimination in rats were performed in anaesthetised animals. Only recently have experiments in awake rats performing discrimination tasks been conducted (Prigg et al. 2002; Krupa et al. 2004). Krupa et al. (2004) show that there are striking differences in signal processing in an active discrimination task as compared to similar passive stimulation of whiskers. These results emphasise the importance of active whisking and of top-down signals to somatosensory processing. Our physical simulator of whisker movement accounts for the active whisking component, while cutting sweeps from the input and using them as basic units of processing may correspond to using top-down information. Hence our model is compatible with the recent experimental results in behaving rats.

In contrast to rats, tactile discrimination has been studied in great detail in monkeys (see Romo & Salinas 2003 for review). In these experiments, monkeys are subject to high frequency stimulation of their fingertips. These so-called “flutter vibrations” (Talbot et al. 1968) are in some respects similar to the high frequent whisker stimulations occurring when rats whisk across textures. Salinas et al. (2000) find that the precise timing of rapidly adapting primary neurons in the somatosensory system of monkeys is conserved in SI but nearly completely absent in SII. Properties of SII cells are insensitive to the exact timing and provide a robust stereotyped response. This property is reminiscent of our model’s output cells. Romo et al. (2002) investigate response properties in monkey SII of animals that compare two mechanical vibrations applied sequentially to the fingertips. They find that the firing rate in the first stimulus period reflects the stimulus frequency, whereas the activity in the second period is a function of the first and the second stimulus. This shows that, in the monkey somatosensory system, signals from successive contacts, which may correspond to whisker sweeps, are present. This property allows comparison of successive sweeps, which is an essential prerequisite for the stability learning rule. Despite the obvious differences in species and paradigm, these correspondences further support the use of temporal coherence for learning somatosensory representations.

Acknowledgements

This work was financially supported by the EU/BBW “AMOUSE” project (IST-2000-28127, 01.0208-1), Honda RI Europe and the Swiss National Science Foundation (PK, 31-61415.01; WE, PBEZ2–107367).

References

Adelson EH, Bergen JR. 1985. Spatiotemporal energy models for the perception of motion. J Opt Soc Am A 2:284–299

Andermann ML, Ritt J, Neimark MA, Moore CI. 2004. Neural correlates of vibrissa resonance; band-pass and somatotopic representation of high-frequency stimuli. Neuron 42:451–463.

Arabzadeh E, Petersen RS, Diamond ME. 2003. Encoding of whisker vibration by rat barrel cortex neurons: Implications for texture discrimination. J Neurosci 23:9146–9154.

Arabzadeh E, Panzeri S, Diamond ME. 2004. Whisker vibration information carried by rat barrel cortex neurons. J Neurosci 24:6011–6020.

Bishop M. 1995. Neural networks for pattern recognition. Oxford: Oxford University Press.

Berkes P, Wiskott L. 2005. Slow feature analysis yields a rich repertoire of complex cell properties. J Vis 5:579–602.

Brecht M, Preilowski B, Merzenich MM. 1997. Functional architecture of the mystacial vibrissae. Behav Brain Res 84:81–97.

Carvell GE, Simons DJ. 1990. Biometric analyses of vibrissal tactile discrimination in the rat. J Neurosci 10:2638–2648.

Deschenes M, Timofeeva E, Lavallee P. 2003. The relay of high-frequency sensory signals in the whisker-to-barreloid pathway. J Neurosci 23:6778–6787.

Einhäuser W, Kayser C, König P, Körding KP. 2002. Learning the invariance properties of complex cells from their responses to natural stimuli. Eur J Neurosci 15:475–486.

Einhäuser W, Kayser C, Körding KP, König P. 2003. Learning distinct and complementary feature-selectivities from natural colour videos. Rev Neurosci 14:43–52.

Einhäuser W, Hipp J, Eggert J, Körner E, König P. 2005. Learning viewpoint invariant object representations using a temporal coherence principle. Biol Cybern 93:79–90.

Földiak P. 1991. Learning invariance from transformation sequences. Neural Comput 3:194–200.

Gibson JM, Welker WI. 1983a. Quantitative studies of stimulus coding in first-order vibrissa afferents of rats. 1. Receptive field properties and threshold distributions. Somatosens Res 1:51–67.

Gibson JM, Welker WI. 1983b. Quantitative studies of stimulus coding in first-order vibrissa afferents of rats. 2. Adaptation and coding of stimulus parameters. Somatosens Res 1:95–117.

Guic-Robles E, Valdivieso C, Guajardo G. 1989. Rats can learn a roughness discrimination using only their vibrissal system. Behav Brain Res 31:285–289.

Hashimoto W. 2003. Quadratic forms in natural images. Network: Comput Neural Syst 14:765–88.

Harvey MA, Bermejo R, Zeigler. 2001. HP Discriminative whisking in the head-fixed rat: Optoelectronic monitoring during tactile detection and discrimination tasks. Somatosens Mot Res 18:211–222.

Hurri J, Hyvärinen A. 2003. Simple-cell-like receptive fields maximize temporal coherence in natural video. Neural Comput 15:663–691.

Kayser C, Einhäuser W, Dümmer O, König P, Körding KP. 2001. Extracting slow subspaces from natural videos leads to complex cells. In: Dorffner G, Bischoff H, Hornik K editors. Artificial neural networks–(ICANN) LNCS 2130, Springer-Verlag. pp 1075–1080.

Kayser C, Körding KP, König P. 2003. Learning the nonlinearity of neurons from natural visual stimuli. Neural Comput 15:1751–1759.

Kayser C, Körding KP, König P. 2004. Processing of complex stimuli and natural scenes in the visual cortex. Curr Opin Neurobiol 14:468–473.

Körding KP, König P. 2001. Neurons with two sites of synaptic integration learn invariant representations. Neural Comput 13:2823–2849.

Körding KP, Kayser C, Einhäuser W, König P. 2004. How are complex cell properties adapted to the statistics of natural stimuli? J Neurophysiol 91:206–212.

Krupa DJ, Matell MS, Brisben AJ, Oliveira LM, Nicolelis MA. 2001. Behavioral properties of the trigeminal somatosensory system in rats performing whisker-dependent tactile discriminations. J Neurosci 21:5752–5763.

Krupa DJ, Wiest MC, Shuler MG, Laubach M, Nicolelis MA. 2004. Layer-specific somatosensory cortical activation during active tactile discrimination. Science 304:1989–1992.

Metha SB, Kleinfeld D. 2004. Frisking the whiskers: Patterned sensory input in the Rat vibrissa system. Neuron 41:181–184.

Moore CI. 2004. Frequency-dependent processing in the vibrissa sensory system. J Neurophysiol 91:2390–2399.

Prigg T, Goldreich D, Carvell GE, Simons DJ. 2002. Texture discrimination and unit recordings in the rat whisker/barrel system. Physiol Behav 77:671–675.

Riedmiller M, Braun H. 1993. A direct adaptive method for faster backpropagation learning: The RPROP algorithm. Proceedings of the IEEE International Conference on Neural Networks, San Francisco, CA.

Rolls ET, Milward T. 2000. A model of invariant object recognition in the visual system: Learning rules, activation functions, lateral inhibition, and information-based performance measures. Neural Comput 12:2547–2572.

Romo R, Salinas E. 2001. Touch and go: Decision-making mechanisms in somatosensation. Ann Rev Neurosci 24:107–137.

Romo R, Hernandez A, Zainos A, Lemus L, Brody CD. 2002. Neuronal correlates of decision-making in secondary somatosensory cortex. Nat Neurosci 5:1217–1225.

Romo R, Salinas E. 2003. Flutter discrimination: neural codes, perception, memory and decision making. Nat Rev Neurosci 4:203–218.

Salinas E, Hernandez A, Zainos A, Romo R. 2000. Periodicity and firing rate as candidate neural codes for the frequency of vibrotactile stimuli. J Neurosci 20:5503–515.

Shoykhet M, Doherty D, Simons DJ. 2000. Coding of deflection velocity and amplitude by whisker primary afferent neurons: Implications for higher level processing. Somatosens Mot Res 17:171–180.

Stone JV. 1996. Learning perceptually salient visual parameters using spatiotemporal smoothness constraints. Neural Comput 8:1463–1492.

Stringer SM, Rolls ET. 2002. Invariant object recognition in the visual system with novel views of 3D objects. Neural Comput 14:2585–2596.

Talbot WH, Darian-Smith I, Kornhuber HH, Mountcastle VB. 1968. The sense of flutter-vibration: Comparison of the human capacity with response patterns of mechanoreceptive afferents from the monkey hand. J Neurophysiol 31:301–334.

Wallis G. 1996. Using spatio-temporal correlations to learn invariant object recognition. Neural Networks 9:1513–1519.

Wallis G, Baddeley R. 1997. Optimal, unsupervised learning in invariant object recognition. Neural Comput 9:883–894.

Wallis G, Rolls ET. 1997. Invariant face and object recognition in the visual systems. Prog Neurobiol 51:167–194.

Wiskott L, Sejnowski T. 2002. Slow feature analysis: Unsupervised learning of invariances. Neural Comput 14:715–770.

Woolsey TA, Van der Loos H. 1970. The structural organization of layer IV in the somatosensory region (SI) of mouse cerebral cortex. The description of a cortical field composed of discrete cytoarchitectonic units. Brain Res 17:205–242.