

Controlling the magnification factor of self-organizing feature maps

H.-U. Bauer

Institut für Theoretische Physik and SFB Nichtlineare Dynamik,
Universität Frankfurt,
Robert-Mayer-Str. 8-10, 60054 Frankfurt/Main 11, Fed. Rep. of Germany

R. Der

Institut für Informatik, Universität Leipzig,
Augustusplatz 10/11, 04009 Leipzig, Fed. Rep. of Germany

M. Herrmann^{1 2}

Laboratory of Neural Modeling, RIKEN,
2-1 Hirosawa, Wako-shi, 351-01 Saitama, Japan

submitted to Neural Computation, Feb. 23, 1995
in print.

¹Corresponding author

²former address: NORDITA, DK-2200 Copenhagen, Denmark

Abstract

The magnification exponents μ occurring in adaptive map formation algorithms like Kohonen's self-organizing feature map deviate from the information theoretically optimal value $\mu = 1$ as well as from the values which optimize, e.g., the mean square distortion error ($\mu = 1/3$ for one-dimensional maps). At the same time, models for categorical perception such as the "perceptual magnet" effect which are based on topographic maps require negative magnification exponents $\mu < 0$. We present an extension of the self-organizing feature map algorithm which utilizes adaptive local learning step sizes to actually control the magnification properties of the map. By change of a single parameter, maps with optimal information transfer, with various minimal reconstruction errors, or with an inverted magnification can be generated. Analytic results on this new algorithm are complemented by numerical simulations.

1. Introduction

The representation of information in topographic maps is a common property of many regions in the brain, including the visual, auditory and somatosensory areas of the cortex. Many of these maps are known to be generated or refined by adaptive self-organization processes. A first theoretical description of the self-organisation of orientation columns in the primary visual cortex has been presented by von der Malsburg (1973), later many more map formation models have been introduced. A particularly widespread algorithm is Kohonen's self-organizing feature map (Kohonen, 1995). It has not only been used to model the formation of maps in different sensory domains (see, e.g., Martinetz et al., 1988, Obermayer et al., 1990, Wolf et al, 1994), but has also found wide distribution in the technically oriented communities. Here, the self-organizing feature map is often utilized as a neighborhood preserving vector quantizer.

A general characteristic of neural maps in brains is the selective magnification of regions of interest. Regions of interest are usually those which are excited most often. Examples include the enlarged representation of the central visual field in visual areas, the enlarged representation of frequencies close to the echolot frequency in the auditory cortex of the bat (Suga, 1991), or the enlarged representation of the hand in areas 1 and 3a in somatosensory cortex (Kaas et al., 1981). It has been hypothesized that the selective magnification of maps is adjusted such that each region in the map is excited equally often. The map then transfers the maximum amount of information about

the stimulus ensemble. This property is not only approximately observed in biological maps (for retinotopic maps, see e.g. Wässle et al, 1989), but is also often regarded as a desirable design objective in technical contexts.

Maps which result from self-organization processes by and large also show an increased magnification in regions which are often stimulated. The detailed magnification properties of map formation algorithms, however, deserve further investigation. This claim is substantiated by the following three arguments. First, an analysis by Ritter and Schulten (1986) clarified that for Kohonen’s self-organizing feature map in the one-dimensional case, and in higher-dimensional cases which separate, the relation between the stimulus density $P(w)$ and the magnification factor $M(w)$ is governed by an exponent $\mu = 2/3$,

$$M(w) \propto P(w)^{\frac{2}{3}}. \tag{1.1}$$

So the self-organizing feature map does not yield a maximum entropy map which would correspond to $\mu = 1$ (at least not in those cases which were analytically accessible so far). A similar result holds for the elastic net algorithm (Durbin and Willshaw 1987), where Gruel and Schuster (1994) found an exponent $\mu_{elas} = 0.4$ for one-dimensional maps in the limit of soft string tension. Question now arises how neural maximum entropy maps could be self-organized.

A second, related, argument concerns the mean distortion error properties of self-organizing feature maps. It has been shown by Zador (1982) that the mean distortion error of a neural map is optimized, if the map has a particular magnification exponent. The value of the exponent depends on the order of the error as well as on the dimension of the map input space. As an example, for one-dimensional maps the mean square error is minimized by an exponent $\mu = 1/3$, which deviates from the value $2/3$ inherent to the SOFM-algorithm in this case. The worst case distortion error is optimized for maps (or vector quantizers) which have a flat magnification, $\mu = 0$. In conclusion, an optimization of neural maps with regard to various distortion error measures would require a control of the map magnification exponents.

In a third line of argument, we are concerned with modifications of maps which could allow for better categorization. Specifically, it might be appropriate to spent receptive fields at (rarely excited) class boundaries instead of at (often excited) class centers, if a classification (categorization) task has subsequently to be performed. Such an inverted magnification scheme was employed in a recent model (Herrmann et al., 1994) for a phenomenon related to categorical perception (Repp, 1985), namely the “perceptual magnet” effect. More details about this effect, and its relation to maps with

negative magnification exponents will be discussed in a subsequent chapter. For the purpose of this introduction it suffice to say that negative magnification exponents might also occur in topographic maps and that existing map self-organization algorithms do not allow for such exponents.

In the present article we put forward a simple extension of Kohonen’s self-organizing feature map algorithm which addresses the problems raised by the above arguments. Solely by adaptively adjusting the local learning rate while keeping all other parts of the algorithm unchanged we can control the magnification properties of the resulting maps. The detailed form of the learning rate control, preceded by a brief description of the unmodified version of Kohonen’s self-organizing feature map algorithm (SOFM), is given in the next chapter. There, we also discuss how the previous analytic results on magnification exponents in SOFMs have to be modified to include the effects of the learning rate control. The third chapter is devoted to results of simulations of our algorithm, in particular with regard to the map distortions errors. In the fourth chapter we show results for two-dimensional maps and argue, how these could provide a neurobiological basis for categorical perception, as exemplified by the “perceptual magnet” effect. A discussion, which addresses the comparison of our algorithm to other optimization efforts in the context of competitive learning, as well as the relevance of our model to auditory recognition experiments, concludes the paper.

2. Self-organizing feature maps with node-dependent adaptability

As a basis for our magnification control algorithm, we use Kohonen’s self-organizing feature map algorithm (SOFM). Not only can it be regarded as a standard algorithm due to its wide distribution, but in addition there are analytical results on its magnification properties already known (Ritter and Schulten, 1986). A SOFM consists of neurons which are located at positions \mathbf{r} in an output space grid A , and which have receptive fields with centers $\mathbf{w}_{\mathbf{r}}$ in an input space V associated to them. A stimulus $\mathbf{v} \in V$ is mapped onto that neuron $\mathbf{s} \in A$, the receptive field center $\mathbf{w}_{\mathbf{s}}$ of which lies closest to \mathbf{v} ,

$$\mathbf{s} = \operatorname{argmin}_{\mathbf{r}} |\mathbf{w}_{\mathbf{r}} - \mathbf{v}|. \quad (2.2)$$

During an adaptation phase, the receptive field center positions $\mathbf{w}_{\mathbf{r}}$ are adjusted such that the resulting map spans the input space in a topographic fashion. A sequence of random stimuli is presented to the map, the respective best-matching neuron \mathbf{s} is determined. Then, the receptive field center

of \mathbf{s} plus its output space neighbors are shifted towards the stimulus,

$$\Delta \mathbf{w}_r = \epsilon h_{rs} (\mathbf{v} - \mathbf{w}_r), \quad (2.3)$$

where the property of being an output space neighbor is imposed by the (usually Gaussian) neighborhood function h_{rs} , Gaussian shape,

$$h_{rs} = \exp\left(-\frac{|\mathbf{r} - \mathbf{s}|^2}{2\sigma^2}\right). \quad (2.4)$$

In this way, the topography of the map is ensured, i.e. neighboring neurons in the output space are made to have neighboring receptive fields (the inverse relation—neighboring positions in stimulus space project onto neighboring neurons—does not necessarily hold). A comprehensive treatment of many theoretical and application related aspects of the SOFM can be found in Ritter et al. (1992) and Kohonen (1995).

A characteristic of neural maps is their areal magnification factor, first introduced by Daniel and Whitteridge (1961). The magnification factor is given by the density $M(\mathbf{w})$ of receptive field centers in the map input space. (In continuum approximation, the position of receptive field centers \mathbf{w} varies continuously in the input space, as does the position of stimuli \mathbf{v}). Often $M(\mathbf{w})$ is related to the stimulus density $P(\mathbf{w})$ via a magnification exponent μ ,

$$M(\mathbf{w}) \sim P(\mathbf{w})^\mu. \quad (2.5)$$

An analysis by Ritter and Schulten (1986) showed that such a relation with an exponent $\mu = 2/3$ also holds for one-dimensional SOFM's in the continuum limit (and, under certain conditions, for maps of higher output dimension (see section 3.4)). So one-dimensional SOFM's do magnify regions of high stimulation, but not sufficiently to be information-theoretically optimal (see introduction).

Therefore, and for the other reasons listed in the introduction, we now proceed to modify the standard SOFM to allow for a control of the magnification behaviour. To this purpose, we introduce adaptive node-dependent adaptabilities ϵ_r (Der and Herrmann, 1992) and replace (2.3) by the modified learning rule

$$\Delta \mathbf{w}_r = \epsilon_s h_{rs} (\mathbf{v} - \mathbf{w}_r). \quad (2.6)$$

Further we require the local adaptabilities ϵ_r to depend on the stimulus density P at the position of the receptive field center \mathbf{w}_r associated to \mathbf{r} ,

$$\langle \epsilon_r \rangle = \epsilon_0 P(\mathbf{w}_r)^m. \quad (2.7)$$

m is a new free parameter of the learning rule which will allow us to actually control the magnification exponents.

Question arises how we can enforce relation (2.7) during learning, when the stimulus density $P(\mathbf{w}_r)$ is not known. Here, we can rely on information acquired by the network already, and can exploit the relation

$$P(\mathbf{w}_r) \propto M(\mathbf{w}_r)\tilde{P}(\mathbf{r}) \quad (2.8)$$

between stimulus density $P(\mathbf{w}_r)$, receptive field center density $M(\mathbf{w}_r)$ and the probability $\tilde{P}(\mathbf{r})$ of the neuron at \mathbf{r} to be the bestmatching node. Assuming independence between successive stimuli, we approximate the mean values $M(\mathbf{w}_r)$ and $\tilde{P}(\mathbf{r})$ by quantities which can be computed at each individual learning step. The receptive field center density $M(\mathbf{w}_r)$ is inversely proportional to the volume of the respective Voronoi polygons, which in turn, in a d -dimensional input space, are proportional to the d -th power of the mean distance $|\mathbf{v} - \mathbf{w}_r|$ between receptive field center and stimulus. The probability $\tilde{P}(\mathbf{r})$ is, on average, related to the time interval Δ_r between successive such events. So we realize relation (2.7) by choosing as a learning step size in one learning step

$$\epsilon_s(t) = \epsilon_0 \left(\frac{1}{\Delta t_s} \left(\frac{1}{|\mathbf{v} - \mathbf{w}_s|^d} \right) \right)^m, \quad (2.9)$$

with \mathbf{s} being the best-matching neuron for this stimulus. Should the data be given in a d -dimensional space, but span only a d_{eff} -dimensional submanifold, then the effective dimension d_{eff} will have to be used in Eq. (2.9). In order to avoid exceedingly large values for $\epsilon_s(t)$ which might destabilize the learning process, we also bound the learning step size according to $\epsilon_s(t) \leq \epsilon_{max} = 0.9$. It should be noted at this point that the whole modification rests on applying the same learning step size ϵ_s associated to the winning neuron \mathbf{s} to all weight changes $\Delta\mathbf{w}_r$ in the learning step. Had we used the individual ϵ_r for the change $\Delta\mathbf{w}_r$, information about the individualized learning steps would not be transferred to the neighboring neurons. No change of the magnification would result, instead each \mathbf{w}_r would fluctuate on an individual scale about its equilibrium value.

How does the changed learning rule (2.6) and (2.7) affect the previously derived magnification exponent $\mu = 2/3$? By a calculation analogous to the original derivation by Ritter and Schulten the relation

$$M(w) = P(w)^{\mu'} = P(w)^{\frac{2}{3}(1+m)} \quad (2.10)$$

for the modified exponent could be established.³ In the following sections we

³Corrections to this exponent due to finite size neighborhood widths, or general neighborhood functions, were studied in recent contributions by Ritter (1991) and Dersch and Tavan (1995). Their results have no direct impact on our present arguments. We note, however, that their more general results could be combined with ours, amounting to a multiplication of their magnification exponents with a factor of $(1 + m)$.

will see in detail, how this relation can be exploited to induce the optimization of various map performance measures by a suitable choice of the control parameter m .

3. Results of Simulations

3.1 Magnification exponents in one-dimensional SOFMs

How does the analytical relation (2.10) based on Eqs. (2.6) and (2.7) compare to numerically obtained maps, which have to rely on Eq. (2.9) approximating Eq. (2.7)? Results for simulations of one-dimensional maps with a linearly increasing stimulus density are depicted in Fig. 1, they coincide very well with relation (2.7). Also for other stimulus distributions ($P(v) \propto \sqrt{v}$ and $P(v) \propto v^2$), the numerically obtained exponents were found to coincide very nicely with those given by Eq. (2.10).

3.2 Information transmission in SOFMs

As mentioned above, a map with optimal information transmission is characterized by $\mu = 1$, such that the resulting probability $\tilde{P}(\mathbf{r})$ for output nodes \mathbf{r} to be excited is a constant across the whole map. To demonstrate that our algorithm can deliver such maps we investigated our numerically obtained maps with regard to their information content

$$I = - \sum_{i=1}^N \tilde{P}(s) \log \tilde{P}(s). \quad (3.11)$$

As shown in Fig. 2a, I becomes maximal for $m = 0.5$, as should be expected for a map with magnification exponent $\mu = 1$.

3.3 Distortion errors in SOFMs

Apart from the maximum transfer of information, an important performance measure for maps used as neighborhood preserving vector quantizers is their mean distortion error

$$E_p = \int_{\mathbf{v}} | \mathbf{w}_s - \mathbf{v} |^p P(\mathbf{v}) d\mathbf{v}. \quad (3.12)$$

As was proven by Zador (1982), for a vector quantizer (or neural map in the present context) operating on d -dimensional data points E_p is minimized, if the map obeys Eq. (2.5) with an exponent

$$\mu = d/(d + p). \quad (3.13)$$

First we note that the unmodified one-dimensional SOFM optimizes the $E_{1/2}$ error, a rather exotic error measure. Our magnification control mechanism

now opens the possibility to optimize more standard distortion errors, like the mean square error, or the mean linear error. In a one-dimensional input space, these should be minimal for $\mu = 1/3$ ($m = -0.5$) and $\mu = 1/2$ ($m = -0.25$), resp.. Figs. 2b,c show that these distortion errors are indeed minimized for these values of m (with a slight deviation of $m = -0.2$ instead of $m = -0.25$ for the linear error.) In addition, the worst case error E_{max} can also be optimized. Minimization of E_{max} requires all receptive fields to be of identical size, i.e. $\mu = 0$. Fig. 2d shows, as can be expected from Eq. (2.10), that the choice of $m = -1$ indeed achieves a minimization of E_{max} . Analogous simulations showed that the above-mentioned distortion errors of two-dimensional SOFMs are also minimized for the resp. values of m resulting from Eqs. (3.13) and (2.10).

3.4 Inverted magnification and the “Perceptual Magnet Effect” in two-dimensional SOFMs

Finally we investigated the regime of negative magnification exponents $\mu < 0$ which we would like to suggest as a possible neurobiological basis for categorical perception, and in particular for the “perceptual magnet” effect observed by P. Kuhl et al. (1991, 1992). In several psychophysical experiments these authors established *a*) that some versions of synthetically generated vowels are perceived to be more typical than others (prototypicality), *b*) that the discrimination capability for vowel prototypes is smaller than for non-prototypes (“perceptual magnet” effect) and *c*) that the position of the prototypes in vowel space depends on the language surrounding, in which children grow up (adaptivity). In the latter experiment, the position of the prototypes were noticeably different at an age of six months already, well before language comprehension.

Let us make several assumptions on how this effect could be implemented in a neural system. First, guided by the abundance of topographic organization in all sensory modalities, we assume that a low-level representation of sounds perceived in these experiments (vowels) is also based on a topographic map. Second, as a consequence of the above-mentioned adaptivity, we assume that the map is self-organized by external stimulation. In other words, we suggest that the “perceptual magnet” effect ought to be discussed in the framework of map self-organization algorithms. Next, we assume that versions of vowels which are perceived as near-prototypical occur more often in a language environment, than versions perceived as non-prototypical. Finally, we assume that different, but similar vowels are easier to distinguish if their representations in the map are further apart.

The latter two assumptions are quite reasonable, yet they contain the chal-

lence for the map framework. “Perceptual magnet” in a map then means that regions of frequent stimulation have to be magnified to a smaller degree than regions of rare stimulation. This corresponds to a negative magnification exponent. In conclusion the “perceptual magnet” effect could nicely be interpreted as a map formation phenomenon, provided one can generate maps with inverted magnification ratios.

We performed numerical simulations to find this regime in two-dimensional maps. The choice of two-dimensional maps for the representation of vowels is plausible considering that vowels occur as clusters in the two-dimensional space spanned by the two formants (see, e.g., Morgan and Scofield 1991). Yet, our magnification control scheme could operate in output spaces of other dimensionality as well. As stimuli we chose points $\mathbf{v} = (v_x, v_y)$ in the unit square, i.e. with $0 < v_x, v_y < 1$, which were drawn according to the probability distribution

$$P_x(v_x) \propto \frac{1}{2} \left(1 + e^{\left(-\frac{(v_x - v_{x,0})^2}{2\sigma_v^2} \right)} \right), \quad P_y(v_y) \propto \frac{1}{2} \left(1 + e^{\left(-\frac{(v_y - v_{y,0})^2}{2\sigma_v^2} \right)} \right), \quad (3.14)$$

with $v_{x,0} = v_{y,0} = 1/3$, $\sigma_v = 8$. So the stimulus density was a Gaussian, located in the left lower center of the unit square, in front of a constant background (see Fig. 3a). In three simulations this input space was mapped onto quadratic output spaces (see Fig. 3b-g). Depending on the exponent m of the local adaptability, the resulting maps provided a higher resolution of the Gaussian peak (Fig. 3b,c), they equilibrated the resolution over the whole input space (Fig. 3d,e), or they decreased the resolution in the region of the peak (Fig. 3f,g).

At this point it should be noted, that the results on magnification for one-dimensional maps can under certain conditions be transferred to two-dimensional maps. The maps have to be organized on a rectangular lattice (which is not to be elongated to much, see Van Velzen (1994)), and the stimulus density has to separate ($P(v_x, v_y) = P_x(v_x)P_y(v_y)$). Then the magnification properties along the two directions also separate and can be treated as two one-dimensional problems (Ritter and Schulten, 1986).

5. Discussion

Some of the ideas presented in this paper relate to the problem of left-out codebook vectors, which can occur in adaptive vector quantization algorithms. Here the most common design objective is the minimization of the mean square distortion error E_2 . It is a plausible (though not exact, see

below) assumption that rarely used codebook vectors do not sufficiently contribute to the minimization of E_2 and, therefore, should be brought into play by an equilibration of excitation probabilities. To achieve this heuristic idea, several strategies have been developed, which can be categorized according to two criteria. The first criterion is the observable onto which the localized adaptation is based. Some algorithms rely on an evaluation of the individual excitation probabilities. These include DeSieno's conscience mechanism (1988) as well as Ahalt et al's frequency sensitive competitive learning (FSCL, 1990). Others exploit measures for the local reconstruction errors. The latter include Kim and Ra (1995) and Chinrungrueng and Séquin (1995). Our present algorithm relies on measures for both, the excitation probability and the local deviations. A second criterion is the way the equilibration is achieved in the algorithm. In many cases, including DeSieno's, Ahalt et al's and Chinrungrueng and Séquin's algorithms, a weighted distance measure is used, which depends on either the excitation probability (such that often excited nodes get a punishing factor for their distance measure) or the local deviations. Whereas the FSCL mechanism achieves magnification exponents relatively close to $\mu = 1$, the approach based on local deviations is considerably worsened by inhomogeneity of the input distribution. Moreover, since a distorted distance measure is not compatible with topology preservation, an implementation of such equilibration for topographic maps should rest in an adaptation of the local learning step sizes. This track is followed in the present paper, as well as (in a different context) in Kim and Ra's algorithm.

Depending on the details of the different implementations, analytic criteria on what is optimized by these different algorithms is often lost. Even though a numerical improvement of performance with regard to E_2 is generally observed, no mathematical reason can be given for this improvement (with the exception of Chinrungrueng and Séquin (1995)). Also, an equilibration of excitation probabilities corresponds to a minimization of E_2 only in the limit of large input dimensions. In contrast, the magnification exponents μ used in the present approach to parametrize the map behaviour can rigorously be related to the maximum of information transfer at $\mu = 1$ as well as to the minima of distortion errors E_p at $\mu = d/(d + p)$.

A second point we would like to comment on is the possible transfer of our results to other map formation algorithms, like, e.g., the elastic net. For the elastic net, Gruel and Schuster (1994) calculated a magnification exponent of $\mu = 0.4$ in one limiting case (soft string tension). The two quantities entering our scheme, namely the probability of each neuron to be best-matching, and the local density of rf's as indicated by the degree of match between rf's and stimulus, can also be evaluated in the elastic net algorithm. Then the

learning rate could be adjusted in an analogous fashion as in the present paper, and the magnification properties of the elastic net could also be made subject of a control.

The connection suggested in this paper between feature map self-organization and categorical perception and the “perceptual magnet” effect is also not specific for the SOFM. Essential for the “magnet” effect was that the representations of two sounds can be distinguished in the map. In the SOFM, with hard competition, two sounds can be distinguished whenever they are mapped onto two different neurons in the map. This seems to be a quite strict model assumption. In a map where the resulting excitation pattern is based on soft competition, the representations of two sounds could possibly overlap, making a distinction less straightforward. However, it is reasonable to assume that also in such maps the discrimination is improved with increasing distance between the centers of the excitation patterns. So our argument about an inverted magnification as the basis for the “perceptual magnets” is not influenced by the nature of the lateral competition in the map.

Finally, we want to discuss the relevance of our model to the auditory recognition experiments mentioned in the fourth chapter and to categorical perception. The proposed model owes to its simplicity several deficiencies. Since by the feature map model only one level of auditory perception was picked out the model cannot account for, e.g., preprocessing of stimuli or context effects. Besides, the collective action of those auditory moduls will have implications to the formulation of a categorization part, e.g. a generalization to maps with more than one center of activation per stimulus or with specific lateral connectivity. Whereas these problems are beyond the scope of the present investigation, we should briefly discuss an alternative setup used for classification in artificial neural systems.

If a single Gaussian unit is provided per prototype vowel its activation will change only slightly close to the prototype, but steeply at the flanks corresponding to non-prototypical stimuli (R. Baddeley, pers. comm, 1994). Another layer could perform the discrimination relying on the difference in activation of the vowel unit. Generally, the distinction between the SOFM model and a Gaussian unit or an effectively similar group of cells is a relative one. The neighborhood function in the Kohonen algorithm has also Gaussian shape and can be interpreted as the probability of activation of a neuron. On the other hand, the Gaussian units have to become spatially arranged. The latter aspect of the problem which relates to various experimental evidence (Kuhl 1991, Repp 1985) seemed to us the more interesting one since it can explain artificially produced “perceptual magnets” using synthetic stimuli, the occurrence of categorial perception in several types of non-speech stimuli,

the possibility to un-learn categorical perception in single auditory modalities, and the degradation from the continuous innate discrimination abilities. Further, the “grandmother cell” idea behind the single unit model renders it to the use in technical systems with a well-defined class of prototypes and non-noisy processing elements. A similar lack of incorporating the virtues of low-dimensional topological features is also present in models using attractor neural networks as a neuronal basis for “perceptual magnets” (Gupta, 1993).

To summarize, we have considered the virtues of a modified SOFM which allows to control the magnification factor of the resulting map. In this way the range of applicability of such self-organization models with regard to map formation in sensory systems is extended. Our results are also relevant for categorization in neural network models.

Acknowledgements

It is a pleasure to acknowledge interesting discussions with Zhao-Ping Li, which brought P. Kuhl’s experiments and the perceptual magnet effect to our attention.

The reported results are partially based on work done in the LADY project sponsored by the German Federal Ministry of Research and Technology under grant 01 IN 106B/3. HUB gratefully acknowledges support from the DFG through Sonderforschungsbereich 185 Nichtlineare Dynamik, TP E3. MH received support from the EC HCM network *Principles of optical Computation*.

References

- S.C. Ahalt, A.K. Krishnamurthy, P. Chen, D.E. Melton, 1990. Competitive Learning Algorithms for vector quantization, *Neur. Netw.* **3**, 277-290.
- C. Chinrungrueng, C.H. Séquin, 1995. Optimal adaptive k-means algorithm with dynamic adjustment of learning rate, *IEEE Trans. Neur. Netw.* **6**, 157-169.
- P.M. Daniel, D. Whitteridge, 1961. The representation of the visual field on

- the cerebral cortex in Monkeys, *J. Physiol.* **159**, 203-221.
- R. Der, M. Herrmann, 1992. Attention based partitioning, in: M. van der Meer, Status Seminar des BMFT Neuroinformatik, DLR (Berlin), 441-446.
- D.R. Dersch, P. Tavan, 1995. Asymptotic level density in topological feature maps, *IEEE TNN* **6**, 230-236.
- D. DeSieno, 1988. Adding a conscience to competitive learning, *IEEE Int'l. Conf. Neur. Netw.* Washington, 1988, I-118-124.
- R. Durbin, D. Willshaw, 1987. An analogue approach to the travelling salesman problem using an elastic net method, *Nature* **326**, 689-691
- J.C. Gruel, H.G. Schuster, 1994. Self organizing feature maps: magnification exponents of elastic net and winner relaxing Kohonen algorithms, submitted to *Neur. Comp.*
- P. Gupta, 1993. Investigating phonological representations: A modeling agenda. In: M. C. Mozer e.a. (eds.) *Proceedings of the 1993 Connectionist Models Summer School*, Hillsdale, NJ; Lawrence Erlbaum.
- M. Herrmann, H.-U. Bauer, R. Der, 1994. The “perceptual magnet” effect: A model based on self-organizing feature maps, In: L.S. Smith, P.J.B. Hancock (Eds.), *Neural Computation and Psychology*, Stirling, 1994, Springer, London, 107-116.
- J. Hertz, A. Krogh, R.G. Palmer, 1991. *Introduction to the theory of neural computation*, Addison Wesley.
- J.H. Kaas, R.J. Nelson, M. Sur, M.M. Merzenich, 1981. Organization of somatosensory cortex in primates, in: *The organization of the cerebral cortex*,

- eds. F.O. Schmitt, F.G. Worden, G. Adelman, S.G. Dennis, MIT Press, 237-261.
- Y.K. Kim, J.B. Ra, 1995. Adaptive learning method in self-organizing map for edge preserving vector quantization, *IEEE TNN* **6**, 278-280.
- T. Kohonen, 1995. *The Self-Organizing Map*, Springer.
- P. K. Kuhl, 1991. Human adults and human infants show a “perceptual magnet” effect for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics* **50** (2), 93-107.
- P. K. Kuhl, K. A. Williams, F. Lacerda, K. N. Stevens, B. Lindblom, 1992. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* **255**, 606-608.
- C. von der Malsburg 1973. Self-organization of orientation sensitive cells in the striate cortex, *Kybernetik* **14**, 81-100.
- T.M. Martinez, H. Ritter, K. Schulten, 1988. Kohonen’s self-organizing map for modeling the formation of the auditory cortex of a bat, *SGAICO–Proc. ”Connectionism in Perspective”*, (Zürich), 403-412.
- D.P. Morgan, C.L. Scofield, 1991. *Neural Networks and Speech Recognition*, Kluwer Academic, Boston.
- K. Obermayer, H. Ritter, K. Schulten, 1990. A principle for the formation of the spatial structure of cortical feature maps, *Proc. Nat. Acad. Sci.* **87**, 8345-8349.
- B. Repp, 1984 in: *Speech and Language*, Vol. 10, *Adv. in Basic Res. and Practice*, ed. N. Lass, Academic Press, 224-335.

- H. Ritter, K. Schulten, 1986. On the stationary state of Kohonen's self-organizing sensory mapping. *Biol. Cybern.* **54**, 99-106.
- H. Ritter, 1990. Asymptotic level density of a class of vector quantization processes, *IEEE TNN* **1**, 173-175 (1991).
- H. Ritter, T. Martinetz, K. Schulten, 1992. *Neural computation and self-organizing maps*. Addison Wesley, Reading, Mass.
- N. Suga, 1991. Cortical computational maps for auditory imaging, *Neural Netw.* **3**, 3-22.
- G. A. Van Velzen, 1994. Instabilities in Kohonen's self-organizing feature map, *J. Phys. A* **27**, 1665-1681.
- H. Wässle, U. Grünert, J. Röhrenbeck, B.B. Boycott, 1989. Cortical magnification factor and the ganglion cell density of the primate retina, *Nature* **341**, 643-646.
- F. Wolf, H.-U. Bauer, T. Geisel, 1994. Formation of field discontinuities and islands in visual cortical maps, *Biol. Cyb.* **70**, 525-531.
- P.L. Zador, 1982. Asymptotic quantization error of continuous signals and the quantization dimension, *IEEE Trans. Inf. Theory* **28**, 149-159.

Figure Captions

Fig. 1: Magnification exponent μ as a function of magnification parameter m . At each value of m , three one-dimensional maps were simulated ($0 < v < 1$, $P(v) = 2v$, $N = 50$ neurons). The resulting exponents μ_{num} are indicated by crosses, the line shows the analytic relation (2.10).

Fig. 2: Information content I , mean linear distortion E_1 , mean square distortion E_2 and worst case distortion E_{max} as a function of the magnification parameter m (one-dimensional SOFMs, $N = 50$ nodes, $\epsilon = 0.5 \rightarrow 0.001$, $\sigma = 10.0 \rightarrow 0.1$, each cross denotes an average of three maps). As expected from Eq. (2.10), I is maximized by $m = 0.5$ (the maximally possible value of $I_{max} = -\sum_{s=1}^{50} \frac{1}{50} \log(\frac{1}{50}) = 3.912$ for an ideal map is indicated by the dotted line). E_1 is minimized for $m = -0.2$, slightly off the theoretical value $m = -0.25$. The minima for E_2 and E_{max} are attained at the theoretical values $m = -0.5$ and $m = -1$, resp..

Fig. 3: Maps from the two-dimensional input space $0 \leq x, y \leq 1$ with a stimulus distribution $P(\mathbf{v})$ exhibiting a peak in front of a background (**a**) onto a grid of 25×25 neurons. (**b**, **d**, **f**) show the the receptive field center distribution of the maps in input space for $m = 0, -1, -2$, resp. (**c**, **e**, **g**) show the corresponding local generalization capabilities, as given by the size of the input space regions, which map onto the respective neuron (i.e. by the inverse of the local receptive field center density $M(\mathbf{w})$).

Fig. 1:

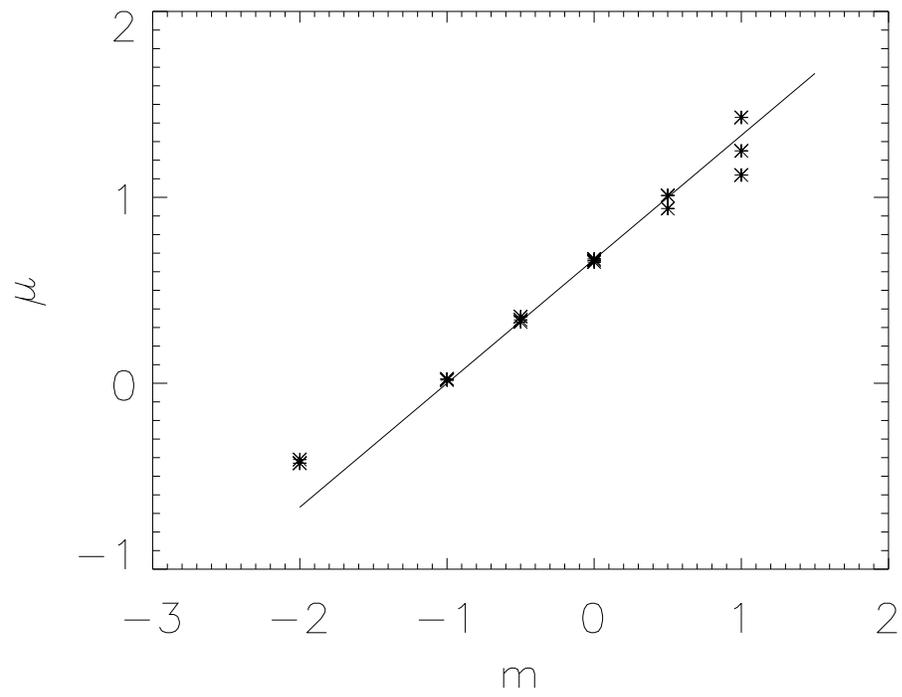


Fig. 2:

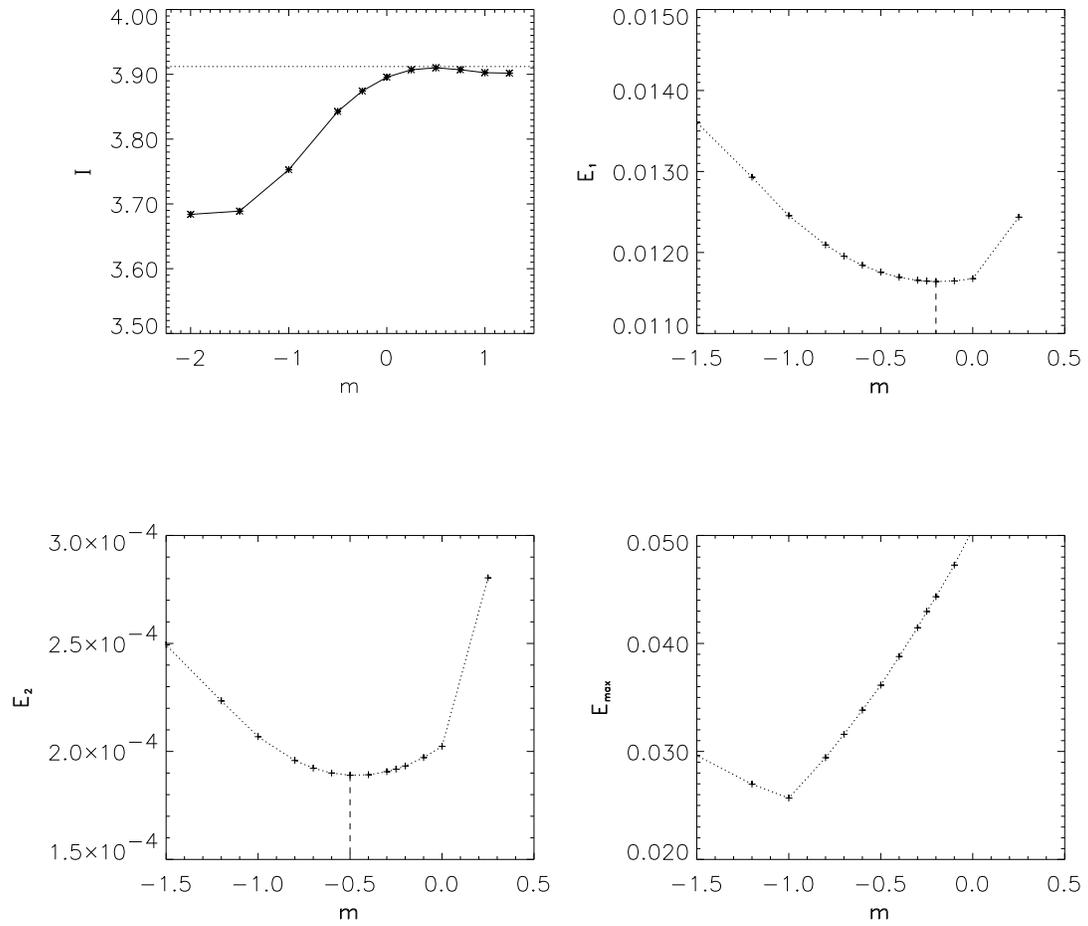


Fig. 3:

