



Invariant encoding of spatial stimulus topology in the temporal domain

Reto Wyss*, Peter König, Paul F.M.J. Verschure

*Institute of Neuroinformatics, Winterthurerstr. 190, University/ETH Zürich, CH-8057 Zürich,
Switzerland*

Abstract

Invariant representations emerge as a central topic in vision research. Here we investigate a transformation of visual spatial patterns by an ensemble of laterally coupled neurons into the temporal domain. The central property of the network leading to this transformation is the monotonic relationship of transmission delays between neurons with increasing distance. We show that stimuli become encoded in the temporal population activity and that this representation is invariant with respect to translations, rotations and small distortions. Furthermore, the proposed model provides a rapid encoding in accordance with physiological results. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Invariant pattern recognition; Temporal coding; Population code

1. Introduction

Primates excel at tasks like visual object recognition, tolerating considerable changes in images, for instance, due to different viewing angles and deformations. Elucidating the mechanisms of such *invariant pattern recognition* is an active field of research in neuroscience [10,12,6,4]. However, still very little is known about the underlying algorithms and mechanisms. A number of models have been proposed which aim to reproduce capabilities of the biological visual system, such as invariance to shifts in position, rotation, scaling and distortion [9,19,7].

Recently, however, the importance of the temporal dynamics of neuronal activity in representing visual stimuli has shifted into the focus of neuroscience [2,11,8]. Several

* Corresponding author. Tel.: +41-1-635-3072; fax: +41-1-635-3053.

E-mail addresses: rwyss@ini.unizh.ch (R. Wyss), peterk@ini.unizh.ch (P. König), pfmjv@ini.unizh.ch (P.F.M.J. Verschure).

modeling studies have addressed the properties of temporal codes [1,17,5]. For instance, Buonomano and Merzenich [3] proposed a model for position-invariant pattern recognition, using temporal coding. Here, we build on these concepts and investigate the formation of invariant representations by the dynamics of activity of neuronal populations.

We investigate a simplified model of primary visual cortex consisting of a map of integrate-and-fire neurons with local excitatory interactions. A central feature of this model is the monotonic relationship between transmission delays in this lateral coupling and the distance between pre- and post-synaptic neurons. We hypothesize that this network property induces dynamics of neuronal activity which is specific to the geometrical shape of a stimulus and invariant with respect to translations and rotations. Such a representation would emerge naturally without the need for training a stimulus repeatedly at different positions or orientations. In order to investigate the validity of this approach we determine the amount of information contained in the temporal population responses of this network for different parameters and stimulus sets. This is done statistically, using a clustering algorithm combined with temporal correlation as a similarity measure. The resulting temporal structure of network activity provides a representation which is position as well as rotation invariant. Moreover, it is robust with respect to synaptic noise and under local and global stimulus distortions. These results suggest that for invariant pattern recognition temporal coding is important, however, not at the level of single neurons, but at the population level.

2. Methods

The investigated network consists of a two-dimensional array of 40×40 standard conductance-based leaky integrate-and-fire neurons, which are modeled with a spike-triggered potassium conductance yielding frequency adaptation. Each neuron connects to a circular neighborhood of fixed size, such that only neurons with Euclidean distance ≤ 9 cells were connected. The synapses are of equal strength and modeled as instantaneous excitatory conductances, while transmission delays are proportional to the Euclidean distance between the positions of the pre- and post-synaptic neurons in the array with 1 ms/cell. Synaptic noise was modeled by perturbing the individual conductances dynamically. Therefore, each conductance was multiplied by a random factor f , drawn from a normal distribution with mean one and variance 0.05. Stimuli presented to the network first pass through an edge-detection stage and the resulting contours are projected topographically onto the array of neurons using a constant excitatory input conductance. After frequency adaptation, the stimulated neurons spike at a frequency of approximately 42 Hz, given the temporal resolution of the discrete time simulations $\Delta t = 1$ ms.

The algorithm for clustering the network's activity-traces is adapted from Victor and Purpura [18], using pairwise temporal correlation as a similarity measure.

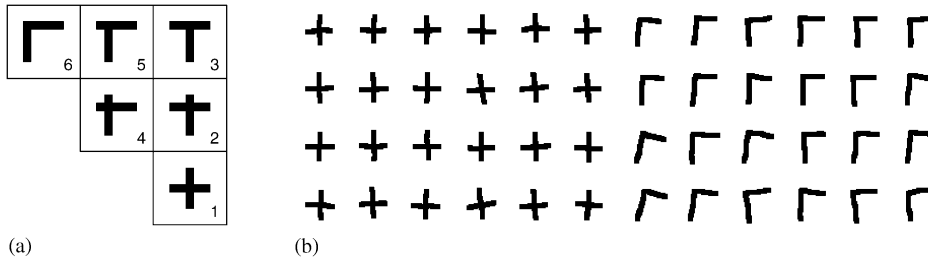


Fig. 1. The stimulus classes. (a) Each stimulus class consists of a horizontal and a vertical bar of equal length, but intersecting at different locations for the individual classes. The arrangement reflects an intuitive notion of topology in a sense that visually similar stimulus classes are closer than dissimilar ones. (b) From each stimulus class, 24 hand-drawn samples were presented to the network.

3. Results

In the first experiment, the network's performance in encoding stimuli invariant to distortion was investigated by presenting the six stimulus classes shown in Fig. 1a. The arrangement reflects an intuitive notion of topology, i.e. class 1 is visually more similar to class 2 or 4 than to class 5 or 6. For each stimulus class, 24 hand-drawn examples were presented to the network (Fig. 1b). For a quantitative analysis, the average responses of the network were clustered into six classes using pairwise temporal correlation over an interval of 100 ms as a similarity measure. The resulting hit-matrix reveals that most stimuli are classified correctly (85%), while the misclassifications (15%) do show some regularity. They only occur among visually similar stimuli. In order to investigate the relationship between the strength σ of the lateral coupling on the encoding of the stimuli, we varied this parameter in the range of $0 < \sigma < 0.25$ nS, given that the leak conductance of the neurons was 20 nS. For each of these lateral coupling conditions we calculated the *mutual information* from the hit-matrix. We found that for $\sigma < 0.01$ nS the activity-traces contained no information about the input stimuli. For $0.01 < \sigma < 0.25$ nS, however, mutual information reached 1.77 ± 0.18 bit (mean \pm S.D., of $\log_2(6) = 2.6$ bit possible). In summary, the network successfully and reliably encodes stimuli in a large range of lateral coupling strength.

In a next step we investigated classification of rotated stimuli. A stimulus set was constructed by taking one sample out of each stimulus class (Fig. 1), and generating the complete sets by rotating each sample by 23 evenly spaced angles between 0° and 360° . As in the previous experiments, no information about the input was conveyed by the network's activity traces for $\sigma < 0.01$ nS. For $0.01 < \sigma < 0.25$ nS, mutual information reached 1.56 ± 0.29 bit. Maximal information of 2.1 ± 0.04 bit is attained for $\sigma = 0.1$ nS. The corresponding hit-matrix reveals the same tendency as for the distortion stimulus set: 87% of the stimuli are classified correctly, while 12% are confused with nearest neighbors and only 1% with non-nearest neighbors. Hence, the population activity-trace reliably encodes the six stimulus classes invariant to rotation.

Subsequently, we analyze an important aspect in visual processing, the speed of information flow. We use the distortion stimulus set, with $\sigma = 0.13$ nS. So far, the

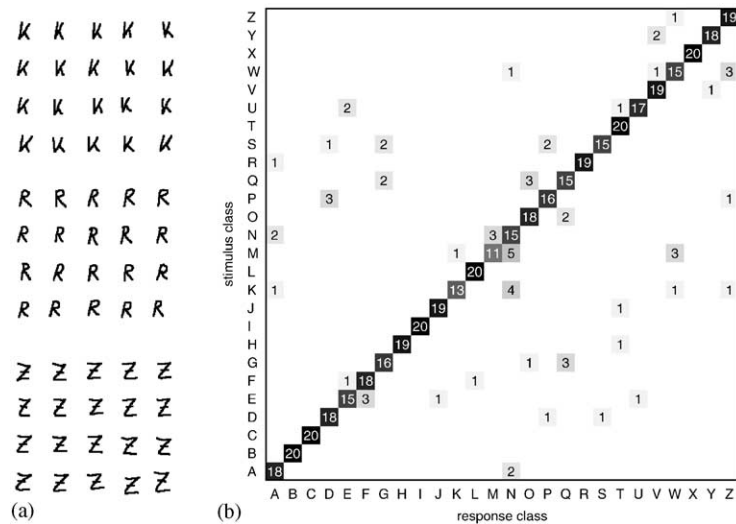


Fig. 2. Classification of handwritten characters. (a) Stimuli of the classes K, R and Z as examples from a single writer presented to the network. Each stimulus class contains 20 samples. (b) The resulting hit-matrix for $\sigma = 0.1$ nS (single trial).

temporal correlation between the responses was determined over an interval of 100 ms. By varying this interval length from 2 to 100 ms, we determine how much information about the present stimulus has been conveyed by the network's activity trace at different times after stimulus onset. We observe that 75% of the information about the present input already becomes available after 17 ms. Thus, the properties of our encoding scheme are compatible with the impressive speed of processing found in the mammalian visual system [13].

How the present model scales with an increasing number of stimulus classes was the subject of the following experiment. A handwritten (single writer) uppercase Roman alphabet was presented to the network, with 26 stimulus classes with 20 samples each (Fig. 2a). For $0.02 < \sigma < 0.25$ nS, information reaches 3.92 ± 0.12 bit (of $\log_2(26) = 4.7$ bit possible). Maximal information of 4.16 ± 0.02 bit is attained for $\sigma = 0.1$ nS, where 88% of the stimuli are classified correctly. The corresponding hit-matrix is shown in Fig. 2b. Thus, the proposed encoding scheme scales to problems of interesting complexity.

4. Discussion

We have shown that in a model of a cortical network, the interaction of network- and stimulus-topology induces stimulus-specific but transformation-invariant temporal dynamics. Thus, stimuli are represented by a temporal population code. This representation is position- and rotation-invariant as well as invariant to moderate distortions.

The stimulus encoding preserves an intuitive notion of visual similarity; classification errors predominantly occur by confusing visually similar stimuli. In addition, without tuning of parameters, the network is able to reliably encode stimulus sets of non-trivial size. These properties apply in a range of the strength of the lateral coupling which spans more than one order of magnitude.

The speed of information flow found in biological visual systems poses severe constraints upon computational models of pattern recognition. Experimental studies have shown that the analysis and classification of complex visual patterns can be carried out by humans in not more than 200 ms [14]. Considering that a minimum of 10 areas are involved to reach the relevant processing areas from the retina, little time is left for intra-areal processing. This view is supported by studies in macaque monkeys, which show that single visual areas can process significant amounts of information in just 20–30 ms [15,16]. Thus, our result, that 75% of information about a stimulus becomes encoded within 17 ms is compatible with these physiological findings.

Acknowledgements

The authors thank Manuel Sanchez-Montanes, Mark Blanchard and Arik Zucker for valuable discussions and contributions to this study. This work was supported by SPP-SNF.

References

- [1] H.R. Arnoldi, K. Englmeier, W. Brauer, Translation-invariant pattern recognition based on synfire chains, *Biol. Cybernet.* 80 (1999) 433–447.
- [2] W. Bair, Spike timing in the mammalian visual system, *Curr. Opin. Neurobiol.* 9 (4) (1999) 447–453.
- [3] D.V. Buonomano, M. Merzenich, A neural network model of temporal code generation and position-invariant pattern recognition, *Neural Comput.* 11 (1999) 103–116.
- [4] I. Fujita, K. Tanaka, M. Ito, K. Cheng, Columns for visual features of objects in monkey inferotemporal cortex, *Nature* 360 (1992) 343–346.
- [5] J.J. Hopfield, Pattern recognition computation using action potential timing for stimulus representations, *Nature* 376 (1995) 33–36.
- [6] M. Ito, H. Tamura, I. Fujita, K. Tanaka, Size and position invariance of neuronal responses in monkey inferotemporal cortex, *J. Neurophysiol.* 73 (1995) 218–226.
- [7] D.I. Perrett, M.W. Oram, Neurophysiology of shape processing, *Imaging Vis. Comput.* 11 (1993) 317–333.
- [8] B.J. Richmond, L.M. Optican, Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. II Quantification of response waveform, *J. Neurophysiol.* 57 (1) (1987) 147–161.
- [9] M. Riesenhuber, T. Poggio, Hierarchical models of object recognition in cortex, *Nat. Neurosci.* 2 (11) (1999) 1019–1025.
- [10] E.T. Rolls, Functions of the primate temporal lobe cortical visual areas in invariant visual object and face recognition, *Neuron* 27 (2) (2000) 205–218.
- [11] W. Singer, Time as coding space? *Curr. Opin. Neurobiol.* 9 (2) (1999) 189–194.
- [12] K. Tanaka, Mechanisms of visual object recognition: monkey and human studies, *Curr. Opin. Neurobiol.* 7 (1997) 523–529.
- [13] S. Thorpe, D. Fize, C. Marlot, Speed of processing in the human visual system, *Nature* 381 (1996) 520–522.

- [14] S.J. Thorpe, M. Imbert, Biological constraints on connectionist modeling, in: R. Pfeifer, Z. Schreter, F. Fogelman-Soulié (Eds.), *Connectionism in Perspective*, Elsevier, New York, 1989, pp. 63–92.
- [15] M.J. Tovee, E.T. Rolls, A. Treves, R.P. Bellis, Information encoding and the responses of single neurons in the primate temporal visual cortex, *J. Neurophysiol.* 70 (2) (1993) 640–654.
- [16] M.J. Tovee, E.T. Rolls, Information encoding in short firing rate epochs by single neurons in the primate temporal visual cortex, *Vis. Cogn.* 2 (1995) 25–58.
- [17] R. van Rullen, J. Gautrais, A. Delmore, S. Thorpe, Face processing using one spike per neuron, *BioSystems* 48 (1–3) (1998) 229–239.
- [18] J.D. Victor, K.P. Purpura, Metric-space analysis of spike trains: theory, algorithms and applications, *Network: Comput. Neural. System* 8 (1997) 127–164.
- [19] G. Wallis, E.T. Rolls, A model of invariant object recognition in the visual system, *Prog. Neurobiol.* 51 (1997) 167–194.