Fitting of spatio-temporal receptive fields by sums of Gaussian components

Thomas Wennekers[1] and Katrin Suder[2]

[1] *Max Planck Institute for Mathematics in the Sciences, Inselstr. 22–26, D-04103 Leipzig, Germany*

[2] *McKinsey&Company, Inc., Magnusstr. 11, 50672 Köln, Germany*

Abstract

In previous work we argued that spatio-temporal tuning functions can be envisaged as a superposition of individual amplitude modulated spatial components, where each component results from a particular projection pathway between the cell classes shaping the response. In the present work we investigate this possibility by fitting sums of Gaussians to simulated responses from neural field models. The time-course of fitted amplitudes and component tunings is shown to provide information about the contribution of input and recurrent synaptic connections to cortical tuning.

Key words: Receptive fields; Spatio-temporal tuning; Field Model; Fitting method

1 Introduction

Visual cortical simple cells have been experimentally shown to reveal dynamic receptive fields (RFs) comprising phases of specifically tuned enhanced and suppressed activity [2,3,8]. A recently developed analytical method for describing such dynamic responses suggests that the space-time profile of a simple cell should be approximately separable into a sum of temporally amplitude modulated Gaussian spatial components [7]. The theory relates each single component to a particular anatomical projection between the underlying cell classes that shape the response (e.g., LGN cells, excitatory and inhibitory cortical neurons). In the present work, we investigate this possibility by means of numerical fits of sums of Gaussians to response functions observed in computer

[1] Tel +49-341-9959-527, Fax +49-341-9959-555, Email: wenneker@mis.mpg.de

simulations. This way the relative contribution of feedforward and cortex-intrinsic excitatory and inhibitory feedback mechanisms to single cell tuning could be quantified in experimental data.

2 Neural field model

We describe the state of a d-dimensional topographically arranged collection of cells by a field $\phi_i(x,t),\ x \in R^d,\ d \geq 1$. The field $\phi_i(x,t)$ reflects average membrane potentials of neurons of type i at location x and time t. Depending on the application the variable x may represent cortical location, visual space, stimulus orientation, or some other feature domain. Physiologically different cell classes or cortical layers can be modeled using several (n) fields as

$$D_i\phi_i(x,t) = -\phi_i(x,t) + I_i(x,t) + \sum_{j=1}^{n} k_{ij}(x) * f_j(\phi_j(x,t)) \ , \qquad (1)$$

where f_i denote the rate-functions of cells in layer i and the D_i are linear differential operators in the time variable t, which contain only first or higher order derivatives in t. They define the dynamics of the membrane potentials, e.g., as first order low-pass filters with time constants τ_i, $D_i = \tau_i\partial/\partial t$, corresponding with decaying exponentials as temporal impulse response functions (which are roughly equivalent to unitary PSPs). The connectivity kernels k_{ij} in (1) represent distance-dependent synaptic densities from cells in layers j to those in layer i (where 'distance' refers to similarity in the respective feature domain described by x). For simplicity, we assume translation invariant densities and denote spatial convolution by the symbol '*' in (1). We take the coupling kernels k_{ij} in (1) as Gaussians with non-singular variance matrix σ_{ij}^2

$$k_{ij}(x) = \frac{K_{ij}}{(2\pi)^{d/2}} \exp\left[-\frac{1}{2}x^T\sigma_{ij}^{-2}x\right] \ . \qquad (2)$$

Gaussian kernels reflect that neurons encoding similar stimulus features are often coupled more densely than those with differing preferences. $I_i(x,t)$ in (1) is some external input into the i-th layer which we take as a Gaussian with amplitude $I_{i0}T_i(t)$ and variance matrix σ_{i0}^2. The matrix σ_{i0}^2 defines the tuning of the input into layer i, I_{i0} and $T_i(t)$ its absolute strength and time-course. Rate functions f_i in (1) are kept general in section 3. Typical choices are semilinear functions with gain β_i and threshold ϑ_i: $f_i(\phi_i) = \beta_i[\phi_i - \vartheta_i]_+ := \beta_i \max(0, \phi_i - \vartheta_i)$ or the logistic function $f_i(\phi_i) = 1/[1 + \exp(-\beta_i(\phi_i - \vartheta_i))]$.

3 Nonlinear superposition principle

In response to spatially localized (i.e. tuned) input, networks of the type (1) can reveal localized responses where significant neuronal firing activity is restricted to a constraint region in space, cf., e.g., Fig. 1A. Other modes of behavior like plane waves or solitary travelling pulses may also be possible, but are not considered in the present work. Since activity localized in x reflects feature selectivity of cells, neural field models can be used to study cortical tuning mechanisms. Moreover, the network activity usually reveals spatio-temporal dynamics, because either the input is already temporally modulated or due to network intrinsic processes. It should be expected that features of such spatio-temporal response functions provide information about the mechanisms underlying cortical tuning phenomena. Therefore, a characterization of the spatial and temporal properties of a response in terms of the underlying anatomical network structure and membrane dynamics would be desirable.

For linear rate-functions, Equation (1) can be readily solved using Fourier-methods. In contrast, if the f_i are nonlinear, explicit calculations of solutions of (1) are almost always impossible. Therefore, in previous work we developed methods to study dynamic localized solutions of (1) approximatively. The approximation basically replaces the localized, i.e., peaked or bell-shaped, firing rate profiles in the model layers by effective Gaussian profiles and derives a closed system of ordinary differential equations (ODEs) for the amplitudes $f_i(\phi_i(0,t))$ and effective tunings $\sigma_{pi}(t)$ per layer. These are exactly the most interesting variables in applications. Details and some applications of the approximation method can be found in [7] and are not repeated here.

A special outcome of the proposed method is that the potential profiles can be approximated by sums of amplitude modulated spatial Gaussian components:

$$\phi_i(x,t) \approx \sum_{j=0}^{n} a_{ij}(t) \exp\left[-\frac{1}{2} x^T \Sigma_{ij}^{-2}(t) x\right] . \tag{3}$$

In (3) each particular component results from an individual synaptic projection from layer j to layer i corresponding with the coupling kernels k_{ij} in (1). For notational simplicity we have included the external input —which is also Gaussian in x— as component $j = 0$. Observe that the component tunings Σ_{ij} as the amplitudes a_{ij} are in general functions of time.

Most importantly, Equation (3) provides some kind of "structure function relationship" because it relates the physiological response of the model to anatomical properties of the participating cell classes and their mutual connectivity. Each anatomical projection k_{ij} (as well as the input) evoke a separate tuned component in the spatio-temporal network response. In nonstationary states

the mixture of weight coefficents a_{ij} is dynamic, such that large amplitude components during certain response phases may reveal the predominating impact of particular projections k_{ij} in instantaneous potential profiles. The superposition in (3) furthermore is additive, because the membrane dynamics are assumed linear. The total network dynamic is of course nonlinear, since it depends on the rate functions f_i. Therefore, we termed (3) "nonlinear superposition principle". It provides a characterization of tuned spatio-temporal field responses in terms of network connecitivies in an intuitive way based on a theoretical analysis of localized solutions in neural field models.

Spatio-temporal response profiles of receptive fields or more general tuning functions can be measured experimentally using correlation methods [3,2]. To the degree that the various assumptions made in section 2 are satisfied, the nonlinear superposition principle suggests that it should be possible to fit these localized but dynamic profiles by a sum of amplitude modulated Gaussians. The resulting time-course of amplitudes and tuning widths then should provide information about the anatomical substrate generating the response.

The next section provides an example for this procedure using artifically generated data, where all parameters are well controlable. Reference [5] presents an application of the approach to real data. There, certain dynamic receptive field effects in cortical simple cells reported in [8] were tracked down by Gaussian fits as resulting from the feedforward projection from visual thalamus to cortex and distinct firing patterns of thalamic relay cells.

4 Numerical example

We considered a network of two one-dimensional layers of cells comprising excitatory and inhibitory neurons, respectively (i.e., $n = 2,\ d = 1$ in (1)). Membranes were first order low-pass filters and rate functions were logistic functions in the excitatory layer and semilinear in the inhibitory layer. Parameters used were $\tau_1 = 2,\ \tau_2 = 5,\ I_{10} = 0.3,\ I_{20} = 0,\ \beta_1 = 4,\ \theta_1 = 0.5,\ \beta_2 = 1,\ \theta_2 = 0,\ K_{11} = 1,\ K_{12} = -0.25,\ K_{21} = .047,\ K_{22} = 0,\ \sigma_{11} = 2.5,\ \sigma_{10} = 15,\ \sigma_{12} = \sigma_{21} = 30$. The above parameters are quite arbitrary choices. For a general discussion of the model behavior and its relation to previous models for cortical orientation tuning cf. [7]. The simulations employed a simple Euler integration scheme with temporal stepsize of 0.1 and a spatial discretization interval of 1 (433 grid points or "units" in total). Equidistributed noise in the interval $[-.075, .075]$ was added to the potentials of each unit in every timestep. For every 8th unit every 20th temporal sample was stored for subsequent fitting using the Levenberg-Marquardt method.

Figure 1A displays the excitatory potential profile when the network was stim-

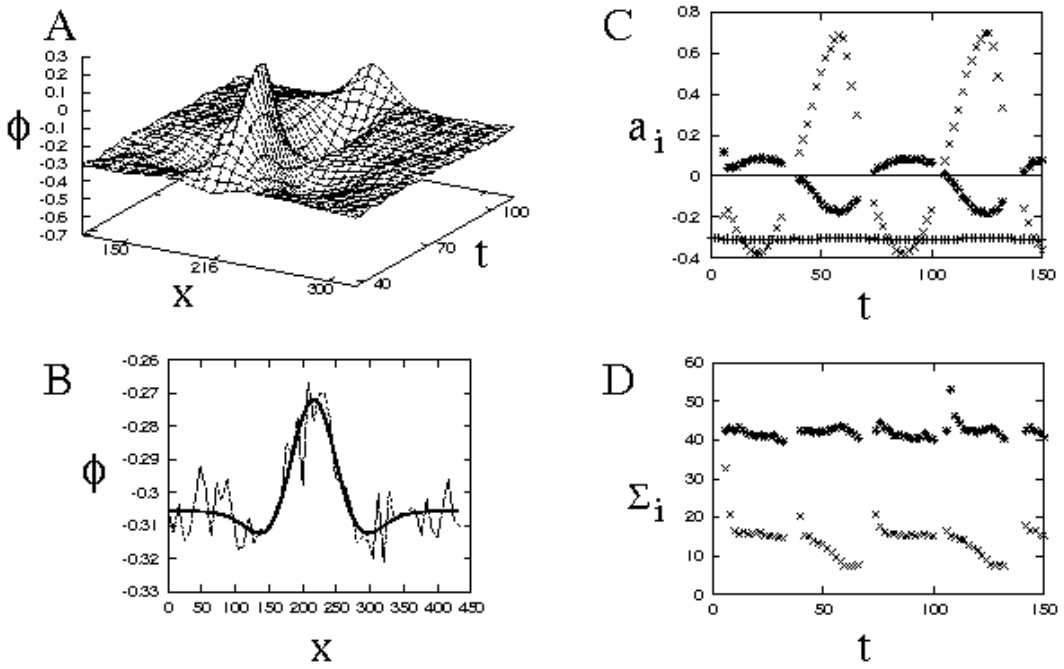

Fig. 1. A) A simulated spatio-temporal response in a lateral inhibition type neural field (counterphase input, 1 period shown, arbitrary units). B) Example fit of data in A at $t = 40$ to a sum of 2 Gaussians plus baseline. C,D) Time-course of fitted component amplitudes (C) and tunings (D).

ulated by a spatial Gaussian centered at $x = 216$ with amplitude $I_{10} = 0.3$ and width $\sigma_{10} = 15$. The input was modulated sinusodially in time at a frequency of $\omega/(2\pi) = 0.015$ corresponding with a period of 67 units of time. Figure 1B shows an example profile for $t = 40$ superimposed with a best fit comprising two Gaussian components and a baseline. Figures C and D reveal fitted component amplitudes and tunings over time. During most of the response, fits as in B comprised a sharp central component (crosses in C and D) and a weaker and broad peripheral one (stars). Not surprisingly, fits often failed where amplitudes were small compared to the noise level, cf. the gaps in the fitted amplitudes near zero crossings in Fig. 1C. Observe also that the baseline in C is virtually constant as it should be since it reflects resting potential in the non-excited regions in space.

In principle we know from the model equations, that the excitatory response profile should consist of *three* tuned components, one for the input, and two for k_{11} and k_{12}, respectively. We have nonetheless chosen only two tuned components because visual inspection of the data seem to reveal only two, cf. Fig. 1B. But in fact, the sharp central component actually consists of two sub-components due to the input and the recurrent excitation. The latter is strong only during the positive phase of the stimulus where firing activity becomes high and is further amplified by the relatively strong recurrent excitation via k_{11}. For that reason, the amplitude time-course of the sharply

tuned central component is not sinusoidal as one would expect for the pure input component alone (Fig. 1C, crosses). Also, as Fig. 1D shows, the width of the central component is roughly $\Sigma_1 = 16$, almost constant, and near to the input tuning during the negative stimulus phase, whereas it changes to small values around $\Sigma_1 \approx 8$ (nearer to the width of the excitatory coupling kernel) during the positive half-wave. This change in tuning again indicates that the nature of the central response carries over from input driven to recurrently driven during the positive stimulus half-wave, but is mainly stimulus driven during the negative half-wave. The second fitted broadly tuned component on the other hand reflects lateral inhibition (positive stimulus half-wave) or disinhibition (negative half-wave) due to the excitatory-inhibitory loop.

5 Discussion

The example in the previous section shows that it is in principle possible to fit spatio-temporal response profiles to sums of amplitude modulated spatial Gaussians as predicted theoretically. The example also points at possible difficulties if the method is applied to real data. First, the number of components is usually not known in experiments. Visual inspection of the data or theoretical reasoning about mechanisms underlying the observed responses may, however, give hints for a proper choice of the component number. This has been successfully demonstrated in [5] and inspection of spatio-temporal orientation tuning functions recorded by Ringach et al. [3,2] also suggest only one or two components. Second, noise in experimental data may complicate the detection of small components. This again asks for a low number of components. Third, some components may have a similar tuning such that their superposition is hardly distinguishable from a single component, cf., e.g., Troyer et al.'s model for orientation tuning [6], where excitatory and inhibitory kernels are assumed equally tuned. In all these cases the number of fitted components would supposedly be lower than the number of actually present ones. Nonetheless, as in the presented example the precise time course of amplitudes and tunings of incompletely separated components may still give hints about the underlying dynamic mechanisms. It should also be possible to design special stimulus time-courses, $T_i(t)$, which enable a separation of otherwise unresolvable components during different stimulus phases, as, e.g., visible in the response differences during positive and negative half-waves in the example. Finally, we should mention that the observed sharpening of tuning during the positive stimulus half-wave should be present also in real data *if* sharply tuned strong excitatory recurrent connections contribute to simple cell tuning as it has been proposed by several authors [1,4]. Such data could be obtained experimentally by presenting counterphase graitings at different orientations and phase-locked averaging of the respective simple cell responses.

References

[1] Ben-Yishai R, Bar-Or RL, Sompolinsky H (1995) Theory of orientation tuning in visual cortex. Proc. Natl. Acad. Sci. USA 92:3844–3848.

[2] Pugh MC, Ringach DL, Shapley R, Shelley MJ (2000) Computational Modeling of Orientation Tuning Dynamics in Monkey Primary Visual Cortex. J.Comput.Neurosci. 8:143–159.

[3] Ringach DL, Hawken ML, Shapley R (1997) The dynamics of orientation tuning in macaque V1. Nature 387:281–284.

[4] Somers DC, Nelson SB, Sur M (1995) An emergent model of orientation selectivity in cat visual cortex simple cells. J. Neurosci. 15:5448–5465.

[5] Suder K, Wörgötter F, Wennekers T (2001) Neural field model of receptive field restructuring in primary visual cortex. Neural Comput. 13:139–159.

[6] Troyer TW, Krukowski AE, Priebe NJ, Miller KD (1998) Contrast-Invariant Orientation Tuning in Cat Visual Cortex: Thalamocortical Input Tuning and Correlation-Based Intracortical Connectivity. J. Neurosci. 18:5908–5927.

[7] Wennekers T (2002) Dynamic approximation of spatio-temporal receptive fields in nonlinear neural field models. Neural Comput. 14:1801–1825.

[8] Wörgötter F, Suder K, Zhao Y, Kerscher N, Eysel UT, Funke K (1998) State-dependent receptive-field restructuring in the visual cortex. Nature 396:165–167.