

The distribution of neuronal population activation as a tool to study interaction and integration in cortical representations

Wolfram Erlhagen^{2,1}, Annette Bastian¹, Dirk Jancke²,
Alexa Riehle¹, Gregor Schöner¹

¹ Centre de Recherche en Neurosciences Cognitives, CNRS, Marseille, France

² Institut für Neuroinformatik, Ruhr-Universität Bochum, Germany

Correspondence:

Dr. Gregor Schöner
C.N.R.S–C.R.N.C
31, Chemin Joseph Aiguier
13402 Marseille Cédex 20
France

Tel: +33-491-164308

Fax: +33-491-774969

Email: gregor@LNF.CNRS-MRS.FR

Key words:

neural population representation; cat visual cortex; monkey motor cortex; neural field theory; cortical interaction; optimal linear estimator

Abstract

In many cortical areas, simple stimuli or task conditions activate large populations of neurons. We hypothesize that such populations support processes of interaction within parametric representations and integration of multiple sources of input and we propose to study these processes using distributions of population activation (DPAs) as a tool. Such distributions can be viewed as neuronal representations of continuous stimulus or task parameters. They are built from basis functions contributed by each neuron. These functions may be explicitly chosen based on tuning curves or receptive field profiles. Or they may be determined by minimizing the distance between chosen target distributions and the constructed DPAs. In both cases, construction of the DPA is based on a set of reference conditions in which the stimulus or task parameters are sampled experimentally. In a second step, basis functions are kept fixed, and the DPAs are used to explore time dependent processing, interaction and integration of information. For instance, stimuli which simultaneously specify multiple parameter values can be used to study interactions within the parametric representation. We review an experiment, in which the representation of retinal position is probed in this way, revealing fast excitatory interactions among neurons representing similar retinal positions and slower inhibitory interactions among neurons representing dissimilar retinal positions. Similarly, DPAs can be used to analyze different sources of input that are fused within a parametric representation. We review an experiment in which the representation of the direction of goal-directed arm movements in motor and premotor cortex is studied when prior and current information about upcoming movement tasks are integrated.

1 Introduction

Why are so many neurons activated in response even to very simple stimuli or during very simple movement tasks? Why are neurons broadly tuned in relation to narrowly defined values of task parameters? Do these neurons help to improve the precision of cortical representations of stimulus or task parameters? This idea suggests a strategy of coarse coding, in which stimulus or task parameters need not to be sampled very extensively (e.g., Erickson, 1974; Gielen, Hesselmans, Johannesma, 1988; Seung, Sompolinsky, 1993). Recent more formalized thinking about this issue of the precision of cortical representations has revealed that typically only few neurons are needed to estimate the values of parameters to which cortical neurons are broadly tuned, as long as neural networks are in place that implement optimal estimators (Salinas, Abott, 1994; Sanger, 1996; Zemel, Dayan, Pouget, 1998).

Our thinking about the role of population activity was guided by ideas from dynamical systems theory (Wilson, Cowan, 1973; Amari, 1977; Grossberg, 1980; Schöner, Kopecz, Erlhagen, 1997) which deemphasize input-output relationships of cortical representations as a way to characterize cortical function. Instead, in this conception, the main function of neural representations is the creation of stable states through interaction. Such states may reflect input-output relationships, but may also reflect interaction or may integrate information. Interaction takes place within a neural representation, so that information from multiple instances of the represented category is either fused or selected from. Globally speaking, interactions are responsible for deviations from the predictions of linear additive models when multiple instances of a representation are stimulated. Integration consists of fusing or selecting from multiple sources of sensory information. For instance, information from current sensory representations may be fused with memorized information to determine which action must be initiated, so that when little time is left to process sensory information the response is primarily determined by memorized information

while later during processing the response is primarily determined by current sensory information (Schöner, Kopecz, Erlhagen, 1997).

In this picture, the activation of broad populations of neurons supports the processes of interaction and integration, so that many neurons representing a wide spectrum of information may cast their vote and thus contribute to what is ultimately translated into behavior (e.g. Ben-Yishai, Bar-Or, Sompolinsky, 1995; Zhang, 1996, Jancke et al., 1996).

In this paper we describe how distributions of population activation (DPAs) constructed from populations of cortical neurons can be used as a tool for analyzing representations of simple task or stimulus parameters (cf. also Anderson, 1994; Zemel, Dayan, Pouget, 1998). A set of experimental conditions, called *reference* conditions, serves to define parametric population representations. The DPA construction is then *extrapolated* to a second set of experimental conditions, which probe interaction or integration. Signatures of interaction are sought in DPAs obtained when complex stimuli or tasks are presented to the system. Similarly, different sources of integrated information can be identified and characterized by varying the sensory or task context in which the parametric population representation is measured.

To help elucidate and to make explicit the concepts of interaction and integration we begin with a section on theoretical modelling ideas. Next, the concepts and specific methods of the DPA approach are laid out. A first experimental example explores interaction within the representation of retinal position in cat visual cortical area 17 (Jancke et al., 1999). The second experimental example shows how current visual information is integrated with prior information about upcoming goal-directed movements in monkey motor cortex (Bastian, Riehle, Erlhagen, Schöner, 1998).

2 Dynamic field theory

First we need to think about “what” is represented by populations of cortical neurons. Here we are concerned with parametric representations, in which a continuum

of stimulus or task situations is relevant. Continuous-valued parameters describe these situations. For instance, an ensemble of visual stimuli might be characterized by such parameters as retinal position, orientation, contrast, or direction and velocity of motion. Similarly, an ensemble of movement tasks might be characterized by such parameters as direction, extent, or peak velocity of the movement. Each specific stimulus or task is represented by a particular value for each pertinent parameter.

To be specific, consider now only a single parameter, x , whose cortical representation we are interested in. In the two experimental examples this will be the retinal position of a point of light along a horizontal axis in one case and the direction of a goal-directed movement in the other case. In order to express information about such a stimulus or task parameter, the abstract concept of “activation” is introduced. One activation variable, u , is associated with each possible value of the parameter, x . High levels of the activation variable, $u(x)$, indicate that information about the value, x , is present, while low levels of activation indicate that this particular value of the parameter is not currently involved in cortical processing. In the next section we shall provide an operational link between the measurable firing rates of cortical neurons and the corresponding level of these abstract activation variables.

Once these definitions have been made, it is clear that patterns of activation, $u(x)$, represent the information available simultaneously about all possible values of the parameter, x . For instance, a unique value of the parameter corresponds to a localized peak of activation. The spatial extent of such a peak reflects the range over which different values of the parameter are fused. Thus, two or more values are represented by a double- (or multi-) peaked distribution if the values are far from each other, or by a broad single-peaked distribution if the values are close to each other. The level of activation reflects amounts of information. For instance, a more probable value is represented by a higher level of activation than a less probable value. The distribution of activation levels over the parameter axis, x , is a *field* in

the sense of physics¹.

Information processing takes place when an activation field evolves in time under the influence of external or internal forms of input. Under some fairly general conditions, this evolution is gradual in time and can be described by a dynamical system

$$\tau \dot{u}(x, t) = -u(x, t) + S(x, t) \quad (1)$$

where $\dot{u}(x, t)$ is the rate of change of activation at field location x and time t . This equation describes how a time varying input sets a momentary stable state, $u(x, t) = S(x, t)$, toward which the field relaxes exponentially with a time scale, τ . If input changes sufficiently slowly, the field activation represents perfectly the input distribution low-pass filtered in time. Equation 1 thus amounts to setting the stimulus-response relationship of the field, and therefore defines what is being represented by the field. Conditions under which this definition is made will be called below *reference conditions*. What characterizes these conditions is the assumption that the different sites of the field evolve independently of each other.

Interactions come into play when more than one location of the field is stimulated. Interaction means that the evolution of the field at one location, x , depends on the level of activation at other field sites, x' :

$$\tau \dot{u}(x, t) = -u(x, t) + S(x, t) + h + \int dx' w(x - x') \sigma(u(x')). \quad (2)$$

Here, the integral sums over all field sites, x' , which contribute with strength $w(x - x')$, but only if activation at those sites is sufficiently large such that the sigmoid function $\sigma(u)$ is larger than zero. A typical form of interaction that we shall look at is local excitation and global inhibition: Activation representing similar values of the parameter, x , is mutually excitatory ($w(x - x') > 0$ for small distances $|x - x'|$) whereas activation representing dissimilar values of the parameter, x , is mutually inhibitory ($w(x - x') < 0$ for large distances $|x - x'|$). This form of

¹The potential field of electrodynamics, for example, associates a level of electrical potential with every location in three-dimensional space.

interaction generates a tendency for neighboring sites to co-evolve, whereas sites at larger distance compete. Thus, convergent information specifying similar values of the parameter, x , is fused, while conflicting information specifying incompatible values of the parameter may lead to decision making or local enhancement of contrast. The model parameter h defines the resting level of the field, obtained in the absence of input. This parameter also controls how much input is required before a field site can begin to contribute to interaction. (This particular mathematical formulation was analyzed by Amari, 1977. Related models were studied by a number of authors including Wilson, Cowan, 1973 and Grossberg, 1980).

While the input-output relationship is trivial in this dynamic field description (the field being defined through a one to one relationship between stimulus parameter value and field site), the interaction is not. From the point of view of information processing, interaction is the major contribution of a dynamic field of this type. What are observable signatures of such interactions? The most direct approach is to consider patterns of input that generate activation at more than one location. Such patterns of input may be thought of as superpositions of multiple inputs, each strictly local. These local inputs represent the reference conditions that define the field. Without interactions the field is predicted to evolve toward a state that is a superposition (i.e., linear addition) of the distributions obtained separately in each of the component reference conditions. Interaction can thus be observed as characteristic deviations of the field response to the complex stimulus or task from the superposition of the responses to the component localized stimuli or tasks.

Figure ?? illustrates this fact in a set of simulations of Equation 2. The field is defined over one dimension of visual space. Seven localized stimuli (arrows in top panel) are applied in separate trials, the ensemble of which forms the reference conditions. Each stimulus generates a different localized pattern of activation in the field centered on the stimulus location (dashed lines on top). When two of the seven localized stimuli (arrows on bottom) are applied simultaneously, the interaction-free

field is predicted to respond with a superposition of its two responses to the two component stimuli (dashed lines on bottom). The presence of interactions (local excitation — global inhibition) leads to a different pattern of activation (solid line on bottom). At the shown distance between the two component stimuli inhibitory interaction is dominant such that total activation is reduced. Moreover, the two peaks in the field are shifted away from the stimulus locations as the gradient from excitatory to inhibitory interaction generates a mutual repulsion between the peaks. In Section 4 we review evidence for interaction effects with such signatures in distributions of population activation obtained from populations of neurons in area 17 of cat visual cortex.

Integration is another aspect of information processing by dynamic fields. Integration takes place when more than one source of input impinge on a field. For instance, a dynamic field supporting the preparation of goal-directed movements integrates input, S_{spec} , from a current sensory stimulus specifying the upcoming movement with input, S_{pre} , representing prior information about possible movements (Schöner, Kopcecz, Erlhagen, 1997; Erlhagen, Schöner, 1999):

$$\tau \dot{u}(x, t) = -u(x, t) + S_{\text{spec}}(x, t) + S_{\text{pre}}(x, t) + h + \int dx' w(x - x') \sigma(u(x')). \quad (3)$$

One source of input (here: S_{spec}) may be used to define the field by applying individual localized patterns of input to the field (reference conditions). The other source of input may then be identified by observing the field under conditions in which the second source contributes. In an experiment, reviewed in Section 5, the direction of a goal-directed arm movement was represented by a distribution of population activation obtained from populations of neurons in motor and premotor cortex (Bastian, Riehle, Erlhagen, Schöner, 1998). In the reference conditions, information about a single direction of movement was provided by a response signal. The contribution of a second source of input, provided by a preparatory signal specifying either one, two or three possible upcoming movement directions, was then identified by observing

the distributions of population activation in the epoch prior to the presentation of the response signal.

3 Constructing Distributions of Population Activation (DPAs)

How can the concept of dynamic fields as parametric cortical representations of information be applied to real neural data? The existence of topographic maps in some cortical areas may at first suggest to search for signatures of activation fields defined directly over the cortical surface. Since changes in neural firing rate depend on many stimulus or task parameters, the appropriate fields might actually be defined over a high-dimensional parameter space, which cannot be described by the two dimensions of the cortical surface. Even for the description in cortical space of a single parameter like “retinal position“ a problem arises from the fact that the retinotopy is far from coming close to a clean representation of the visual space (Das and Gilbert, 1997). In addition, the application of very small stimuli results in a widespread pattern of cortical representation (Grinvald et al., 1994). This renders a similar analysis of the interaction effects studied within the framework of parametric representations nearly impossible.

A more powerful idea is, therefore, to construct lower dimensional activation fields directly over those parameter axes that are actually probed in a single set of experiments. Such a set might probe, for instance, retinal location by presenting stimuli in different positions on the retina, but might not probe such parameters as direction, shape, orientation, line length, etc (e.g., by using isotropic, point-like local brightness patterns). Similarly, only a limited set of voluntary movements is used to probe how the firing rate of neurons in motor cortex varies with movement task (described, for instance, by the spatial direction in which the target lies from the initial position) although those neurons may modulate their activity also when many other, unexplored parameters are varied (e.g., the configuration of the effector,

the amount of resistance to the movement, etc., see, for instance, Scott, Kalaska, 1995).

Construction of a DPA estimator from neural activity observed in the reference conditions. To define an estimator for the distributions of population activation (DPAs) of a given stimulus or task parameter (or multiple such parameters), reference conditions must be created in which the different values of this parameter are probed. This may be done by presenting only a single parameter value at a time. For instance, the parameter “retinal position” can be sampled by presenting localized spots of light on a grid of possible retinal position. The parameter “movement direction” is probed by presenting in different trials a sensory signal (“reponse signal”) that uniquely specifies a particular movement direction for each trial. Neuronal activity observed from many neurons in particular cortical areas is averaged over a particular, fixed time interval following the presentation of such elementary stimuli. The ensemble of such neuronal activity forms the basis of the construction of the DPA.

Mathematically, the construction is a scalar distribution function, $u_k(x)$, of the variable, x . It is defined as a linear combination of basis functions, $f_i(x)$, contributed by each neuron, i . The coefficients, $c_i(x_k)$, of each basis function are proportional to the activity of the neuron, i , in the reference condition, x_k . Thus, for each elementary stimulus or task condition, x_k , a different distribution, $u_k(x)$, of population activation (DPA) is defined as:

$$u_k(x) = \sum_{\text{neurons}, i} c_i(x_k) f_i(x) \quad (4)$$

Here, $c_i(x_k)$, is the neural firing rate of neuron, i , in the experimental condition, x_k .

We consider two approaches for defining the *basis functions*, $f_i(x)$, contributed by each cell, i . The first approach consists of choosing the basis functions directly from firing patterns of neurons in the reference conditions. For instance, appropriately normalized tuning curves or receptive field profiles, obtained from elementary stimuli

or task conditions within the set of reference conditions, may be used (cf. Anderson, 1994, for a related idea). This approach includes as a special case the original “population vector” idea of Georgopoulos and colleagues (Georgopoulos, Schwartz, Kettner, 1986). The direction of the population vector $P(x_k) = \sum_i c_i(x_k)x'_i$, where x'_i is a fixed vector (“preferred direction”) for each neuron, i , and x_k is the current movement direction, is identical to the peak position of the scalar distribution function Eq. 4 if the basis functions $f_i(x)$ are chosen as the cosine of the angle between the preferred direction x'_i of each cell and the movement direction x_k . The DPA method yields a scalar distribution function in the underlying space of movement directions, while the population vector method yields a single value of movement direction, at which the distribution is maximally activated.

A second approach consists of imposing a set of target DPAs that are defined for the set of elementary stimuli or tasks. For instance, at each elementary task condition, a DPA of a particular fixed functional form centered on the parameter value corresponding to that task condition may be imposed as the population representation (cf. Zemel, Dayan, Pouget, 1998, for a similar approach). By minimizing the difference between the DPA computed from actual firing rates and the imposed target DPA, the basis functions, $f_i(x)$, are determined. This minimization can be based on optimal linear estimation (OLE) (Salinas, Abbott, 1994; 1995) or on other procedures (Pouget et al., 1998; Gielen, Glasius, Komoda, 1999). In practice, the represented dimension, x , is discretely sampled with a fixed but arbitrary sampling interval, Δx . The minimization determines the basis functions, $f_i(x)$, on that grid of sample points.

Extrapolating the DPA estimators to other than the reference conditions. To actually learn something about the information processing that takes place within DPAs, these must be observed in conditions other than the reference conditions. These new conditions may involve complex stimuli probing interaction

or may involve multiple sources of specification probing integration of information. *Extrapolation* consists of keeping fixed the basis functions, $f_i(x)$, and using the current firing rates of each neuron, i , observed in the new conditions as the coefficients, $c_i(x_k)$. For instance, when visual stimuli are presented that illuminate more than one retinal location, the firing rates observed may be multiplied with the basis functions and summed over neurons to estimate the DPA representing such “complex” stimuli. Similarly, when incomplete information about an upcoming movement is given prior to a response signal that uniquely specifies that movement, then a DPA representing this second source of specification may be estimated by using the current firing rates observed before the response signal in the DPA formula Eq. 4.

Such extrapolated DPAs may be analyzed in a number of ways. Signatures of interaction may be sought by comparing the observed DPAs representing complex stimuli with DPAs predicted from the reference conditions by linear superposition. Overall, deviation of observed from predicted DPAs are evidence for interaction, but this is less interesting as such. More important is to characterize the way these deviations depend on the structure of the complex stimulus (e.g., the distance between the stimulated retinal locations) as well as how these deviations evolve in time (e.g., early excitation vs. late inhibition). Similarly, when multiple sources of input are played against each other, the total amount of activation, but also the shape and the temporal evolution of the DPAs may reveal properties of integration (e.g., fusion of similar information vs. suppression of dissimilar information).

Extrapolation in this sense has been used earlier to analyze information processing in parametric cortical representations when Georgopoulos and colleagues estimated the population vector of movement direction under conditions in which a monkey had to either withhold a response or to “mentally rotate” the required movement direction from the direction displayed (Georgopoulos et al., 1984; 1989). In both cases, the “preferred” movement directions (corresponding to our basis functions, see above), had been determined in other experimental conditions (corre-

sponding to our “reference conditions”) involving neither withholding of responses nor mental rotation. These preferred movement directions were then weighted with the current firing rate of each neuron during the waiting or rotation period.

4 Example 1: Using the DPA to study interaction in primary visual cortex

It is well established that the firing of single neurons is modulated in a highly nonlinear way by complex stimuli (Szuborski, Palmer, 1990; Tolhurst and Heeger, 1997). Much less is known about how populations of neurons in the visual cortex collectively represent a complex visual scene. Psychophysical evidence suggests that the perceived position of a stimulus is influenced by its visual surround (for recent review see, e.g., van der Heijden, Müsseler, and Bridgeman, 1999). While it is tempting to explore in ever greater detail the potential feed-forward neural circuitry that might be activated by complex stimuli, at some point an account for interaction within the activated neural populations must be provided. Our approach enables us to examine this issue through the quantitative analysis of the temporal evolution and spatial shape of DPAs representing complex stimuli. It provides a mechanistic understanding of how nonlinear interaction processes within the neural network shape the spatio-temporal properties of the reconstructed activity patterns. In this section we illustrate how interactions within the representation of retinal position in primary visual cortex can be uncovered and thus demonstrate how the DPA can serve as a tool for the analysis of interaction.

In the experiments of Jancke et al. (1999), reference conditions consisted of neural activity recorded in the foveal representation of cat area 17 (178 neurons) during the time interval from 40 to 65 msec following presentations of simple visual stimuli. Each stimulus was a small square (0.4 by 0.4 deg) of light presented at one of seven horizontally positioned contiguous locations, x_k ($k = 1, \dots, 7$). Two-dimensional receptive field profiles were measured separately (using a stimulus grid adjusted to

each cell’s receptive field). Interaction was probed by presenting “composite” stimuli, which consisted of two squares of light at varied horizontal spatial separation (from 0.4 to 2.4 deg visual angle).

DPA’s were constructed based on two different approaches. By restricting the representation of retinal position to a single dimension along the horizontal axis, probed by the reference conditions, an optimal estimation based DPA was constructed. For each elementary stimulus, the target function was a Gaussian curve, centered on the stimulus location, x_k , discretely sampled by a grid with 50 sample points. The basis functions that each cell contributes were determined by minimizing the distance between the target functions and the DPA, Eq. 4, in a least squares sense across all 7 reference conditions. Extrapolation consisted simply of replacing, for each cell, i , the coefficient, $c_i(x_k)$, in Eq. 4 by its current firing rate in the conditions involving composite stimuli, while keeping the basis function, $f_i(x)$, fixed.

It would be difficult to generate a DPA over the two dimensions of retinal position this way, as this would require reference conditions that sample retinal position on a two-dimensional grid, implying prohibitively long measurements on each cell. To obtain a two-dimensional parametric representation of retinal position, we used the “direct” approach, in which we chose the two-dimensional receptive field profile of each cell as its basis function, $f_i(\mathbf{x})$, where \mathbf{x} is now a two-dimensional position vector. Because receptive field profiles were measured on grids centered on each cell’s receptive field, these basis functions had to be resampled and padded with zeros in order to generate a common sample grid for all cells across the two-dimensional patch of retinal space, on which the composite stimuli were presented. The DPA estimator was slightly modified relatively to Eq. 4 by normalizing the raw firing rate, $C_i(x_k, t)$, measured in the reference conditions in each time bin, t , by

$$c_i(x_k, t) = \frac{C_i(x_k, t) - b_i}{m_i - b_i} \quad (5)$$

where b_i is the background and m_i the maximum firing rate of neuron i . Extrapolation

tion consisted of replacing the raw firing rate, $C_i(x_k, t)$, in this expression by the raw firing rate observed in conditions with composite stimuli, keeping basis functions as well as b_i and m_i fixed.

Both estimators provide DPAs that reliably represent the retinal position of simple stimuli in the reference conditions. The optimal one-dimensional DPA is arbitrarily precise in the reference conditions for our sample of 178 cells. The time-resolved version has unchangingly high precision throughout the response interval from ca. 30 to 80 ms after stimulus onset. The two-dimensional DPA represents the simple stimuli by transient localized patterns of activation, which are roughly circular in shape. These reach maximum level of activation after about 54 ± 4 msec for all stimulus positions tested in the reference conditions and then decay back to resting level. The peak location remains approximately constant during this evolution resulting in a constant reconstruction error of about half a stimulus width, that is, 0.2 deg (Jancke et al., 1999).

A first, interesting observation can be made by analyzing the shape of the DPAs during the response interval. The width of the localized peaks was characterized by computing a theoretical standard deviation (treating the DPA as a probability distribution that was normalized in each time slice; only times larger than 45 ms were used as total activation is too small and noisy earlier). Figure ?? shows how this measure of width evolves during the response interval from 45 to 80 ms. The mean over the 7 reference conditions is shown in each time slice. The widths of the localized peaks of activation increase continuously during the response interval. While the absolute level of width cannot be compared between one and two-dimensional distributions, this trend is clear in both estimators. This observation goes in the opposite direction of the typical argument, that posits that neuronal specificity increases during a response as input from the less specific feed-forward connections are over time more and more modified by input from recurrent connections (e.g., Orban, 1984).

The extrapolated DPAs observed following presentation of the composite stimuli reveal signatures of interaction effects. Figure ?? shows the temporal evolution of activation in the one dimensional DPAs. The left-most component stimulus is common to all composite stimuli. Activation was therefore integrated over a strip of width plus/minus 0.4 deg around that stimulus location. This activation measure is shown as a function of time for all six composite stimuli with increasing retinal distance between its two component squares of light (0.4 to 2.4 deg from top to bottom, solid line). For comparison, the DPA obtained when the left-most stimulus is applied alone (in one of the reference conditions) is also included (dashed line). Comparison between these two types of DPA reveals that at small stimulus separations, the early part of the response is enhanced (early excitatory interaction), whereas the later part of the response is reduced (late inhibitory interaction). At larger separations, only the inhibitory component remains (local excitation, global inhibition). More quantitative detail is reported in Jancke et al., 1999.

The interaction effects shown here lead also to changes in the position of the localized peaks. Figure ?? compares the one-dimensional DPAs observed for three composite stimuli (solid lines) to the linear superpositions of the DPAs obtained for the two reference conditions in which the corresponding squares of light are presented in isolation (dashed lines). At the three stimulus separations (separation 1.6, 2, 0, and 2.4 deg from left to right), the DPA is bimodal, with a peak near either of the two component stimuli, whose positions are marked by vertical lines. Note the overall reduced level of activation for the observed compared to the superposed DPAs, evidence for inhibitory interactions. The amount of repulsion (0.30, 0.27, 0.17 deg from left to right) has to be compared to the reconstruction error of the optimal DPA for a simple stimulus (0.02 deg, Jancke et al., 1999). The size of the deviations differs in more than one magnitude revealing the robustness of the repulsion. This repulsive effect is predicted from the observed spatial gradient from local excitatory to global inhibitory interaction (see Figure ??). Figure ?? shows

this repulsion effect for the two-dimensional DPAs based on receptive fields. The two largest stimulus separations are shown (2 deg in panel (a) and 2.4 in (b)). The observed DPAs (solid lines) are compared with the superpositions of the DPAs obtained for each individual simple stimulus (dashed lines). Spatial contours represent activation levels corresponding to 50% of peak activation of the DPAs averaged over the 30 – 80 msec time period after stimulus onset. The maxima of the DPA (marked by crosses) are repelled from each other as compared to the spatial position of the maxima of the superposed DPAs (triangles). The amount of repulsion is 0.16 deg at 2.0 deg stimulus separation and 0.23 deg at 2.4 deg stimulus separation.

Figure ?? illustrates that such repulsion effects are predicted from the interacting field dynamic discussed in Section 2. Eq. 2 was simulated for seven localized stimuli (arrows on top) and a composite stimulus (arrows on bottom). Note how activation is reduced by inhibitory interaction when a composite stimulus is applied compared to the elementary stimuli. Moreover, the two peaks in the field are repelled outward from the positions of the two component stimuli. This is due to the interaction profile in which excitatory interaction over short distances in the field cross over into inhibitory interaction over larger distances. All other interaction effects can likewise be modelled in detail within this theoretical framework (see Jancke et al., 1999).

To demonstrate that the DPA construction method using reference conditions and extrapolation correctly picks up such interaction effects, we analyzed the simulated time courses of field activation, $u(x, t)$, using the optimal DPA construction procedure. The seven stable stationary solutions obtained in response to seven localized inputs (top panel of Figure ??) constituted the reference conditions used to determine the basis functions, $f_i(x)$, for each “neuron” (field site), i . Only activity above the resting level (defined as the stable state of the dynamics in the absence of input) was used. For convenience, the field patterns were normalized. The target DPA for each stimulus was chosen as a Gaussian centered on the stimulus loca-

tion (top left of Figure ??). Because the positive portions of the field solutions do resemble gaussians near zero we subtracted a constant from these Gaussians and retained only the positive part, making target and field states more similar. Least squares optimization then led to good reproduction of the field activation patterns in the reference conditions (bottom left of Figure ??). The basis functions, $f_i(x)$, were then held fixed and the DPA estimator was extrapolated by using the field activation simulated at each field site, i , in response to composite stimuli as “firing rate” $c_i(x_k)$ in Eq. 4. The bottom right panel of Figure ?? illustrates that the signatures of interaction (overall inhibition and repulsion effect, cf. top right panel of Figure ??, which repeats what was shown in the bottom panel of Figure 1) are preserved in the estimated field DPA.

5 Example 2: Using the DPA to study integration in motor cortex

The observation that many neurons in motor and premotor cortex are broadly tuned to the direction of arm pointing movements has led Georgopoulos and colleagues (review, Georgopoulos, 1995) to the hypothesis, that populations of such neurons represent the movement parameter “direction”. This suggestion has led to intense research on the precision and generality of such population coding. Moreover, processes such as mentally rotating the specified movement direction or keeping in memory a prepared movement direction during a delay have been linked to the representation of movement direction by populations of neurons.

A dynamic field theory of movement preparation based on behavioral data was proposed by Erlhagen and Schöner (1999). Dynamic fields representing different movement parameters including “direction” integrate two sources of input (Eq. 3): (a) The “command” input fully specifies the upcoming movement and models the typical “response signal” in reaction time experiments. (b) The “preshaping” input reflects the range of possible movements in a particular task setting. Such infor-

mation may be provided from a memory trace of the previous movements, but also from sensory information (precued information, perceptual layout of work space). Because such information is available before the response signal, it preshapes the movement parameter field such that the final specification process triggered by the response signal departs from an already prestructured representation.

Bastian, Riehle, Erlhagen, and Schönér (1998) have experimentally observed the preshaping of DPAs. DPAs were constructed from populations of neurons in motor and premotor cortex as an estimate of a dynamic field representing movement direction. Macaque monkeys were trained to perform arm movements from a center target to one of six targets arranged equidistantly around it. The preparatory signal (PS), providing prior information, consisted in illuminating in green one, two, or three contiguous targets. One second later, one of the lighted targets turned red, any other being turned off. This served as response signal (RS), instructing the monkey to reach toward the lighted target and to touch it. In three blocked conditions, prior information about one, two, or three possible directions of the upcoming movement was provided. The original analysis consisted of determining, for each neuron, the tuning curve to movement direction on the basis of its firing rate integrated over the reaction time interval in the condition of complete prior information prespecifying a single movement direction. Tuning curves were normalized and used as basis functions of a DPA (Eq.4) which was then extrapolated to the conditions of incomplete prior information in which two or three targets remained possible. These DPAs revealed prestructuring of the representation that reflected the location and range of precued movement directions during the preparatory period.

Here, we look at these results afresh by analyzing data from another monkey (117 neurons measured in motor and premotor cortex) using the OLE approach. The reference condition, on which we based the construction of the optimal estimator for the DPA over the parameter “direction”, was again the integrated activity over the reaction time interval, but now averaged across all three precue conditions. This

is a conservative choice, since the tendency of the OLE to compensate for other structure in the data than that imposed by the target distributions would act in a sense opposite to preshape. The target function was chosen as a circular normal distribution

$$\hat{u}_k(x) = A \exp[\sigma_{cg}(\cos(x - x_k) - 1)] - B \quad (6)$$

representing the direction, x , measured in radians. To compensate for the average background firing rate, a constant $B = 0.15$ was used and the factor $A = 1.15$ made the target distribution amplitude equal to 1.0. The width was chosen as $\sigma_{cg} = 0.8$ rad (approximately 45 deg) and 36 equidistant directions were used to sample the full circle. The least squared difference between this target distribution and the DPA Eq. 4 summed over all six directions, x_k ($k = 1, \dots, 6$), was minimized by computing the sampled basis functions, $f_i(x)$ (Salinas, Abbott, 1994). The basis functions, obtained from this optimization, were held fixed and the DPA estimator was extrapolated by inserting the current firing rate, $c_i(t)$, computed in each 100 msec time window, t , separately for each precue condition into Eq. 4.

Figure 7 gives a survey over the evolution of the DPA by comparing the DPA obtained when prior information specifying one movement direction was provided (on top, target 1 precued) with the DPA obtained when two targets (1 and 2) were precued. In both cases, a movement to target 1 was performed following the response signal. In response to complete prior information, a peak of activation appears that is centered over the cued target. After a first phasic component, activation builds up in anticipation of the response signal. When two targets are precued, very little activation is visible immediately following the precue, but a broad peak centered over the two cued targets builds up as the response signal approaches.

Figure 8 provides a more quantitative description. DPAs were constructed for three epochs, the first half of the preparatory period, the second half of that period, and the reaction time period (between RS and movement onset). DPAs constructed in the three conditions, that is, prior information specifying one movement direction

(target 1), two movement directions (targets 1 and 2) and three movement directions (targets 6,1 and 2), are compared. During the first half of the preparatory period, the DPA is preshaped only when a single movement direction is precued. At the end of the preparatory period, two and three target precues lead to clear prestructure of the DPA centered correctly over the range of possible targets. They have lower activation and are broader, the broader the range of precued targets. During the reaction time period, signatures of these preshapes remain visible as the DPA gradually shifts to the correct location and shape.

Figure 9 provides quantitative evidence for the specificity of preshaping. The circular concentration (Fisher, 1993) is computed as a measure of width of the DPAs. Higher values of concentration indicate sharper distributions of activation. During the reaction time period (RT), all DPAs are sharper than during the second half of the preparatory period (PP2) (and much sharper than during the first half of the preparatory period, not shown). With increasing range of precued directions from one to two and three precued contiguous targets (increasing uncertainty about movement direction), concentration decreases (the DPA becomes less specific) in both epochs. The preshaping of the DPAs thus reflects, through its shape, the amount and metrics of prior information about upcoming movements.

These results closely match the earlier results obtained for another monkey using the direct tuning curve based DPA construction (Bastian et al., 1998). This demonstrates that the integration of prior and current sensory information within populations of cortical neurons can be analysed using the method of distributions of population activation. That analysis is largely independent of details of the estimation procedure used.

6 Conclusions

Distributions of population activation (DPAs) were constructed as functions of a represented parameter by combining the responses of an ensemble of neurons in a

set of reference conditions. The basis function that each neuron contributes was either directly chosen based on tuning functions or determined by minimizing the difference between a set of chosen target distributions and the corresponding DPAs. By extrapolating the DPA estimator, keeping the basis functions fixed, into conditions other than the reference conditions, signatures of interactions within the neural population and of integration of multiple sources of input to the population could be identified. These signatures were largely independent of the precise method that was used to construct the DPA.

The goal of this method of analysis is the discovery and description of cortical function in a language that emphasizes interaction and integration while the concepts of coding and input-output mapping play more of an auxiliary role. Deviations from linear superposition of DPAs obtained in response to superpositions of stimuli or task demands are therefore analyzed for their dependence on time and on distances in the space of the represented parameters. The temporal evolution of DPAs generated in response to sensory signals is characterized not only in terms of total activation, but also with respect to their shape and the localization of peaks of activation.

The DPA may be conceived of as a quantitative characterization of the cortical representation of continuous stimulus or task parameters. It may thus serve as a tool to the neuroscientist for finding out about processes taking place within such parametric representations. It is another issue then to characterize in which way the brain “reads” or “uses” population representations. One way this can be done is by deactivating parts of populations of broadly tuned neurons and observing whether the parameter represented by these populations changes. David Sparks and colleagues (Lee, Rohrer, Sparks, 1988) have done this for the parameter “end-point of a saccadic eye movement” by exploiting the topographic ordering of receptive fields of neurons in colliculus superior. Saccadic end-point shifted as predicted from vector-averaging over the broad population of activated neurons when small patches of colliculus were reversibly deactivated. Similar conclusions were drawn by New-

some and colleagues from experiments in MT (Groh, Born, Newsome, 1997). Such observations make it highly plausible that all activated parts of a neural ensemble contribute to the behavior generated on the basis of these neural events. Recent quantitative estimates bolster that position (e.g., Seung, Sompolinsky, 1993). The emphasis on interaction and integration, that is at the core of the theoretical perspective outlined in Section 2, suggests that one reason activation is spread broadly within parametric representations is to enable large parts of the neural representation to interact and to contribute toward the fusion or selection of multiple sources of input.

Acknowledgments: Grant support from the Deutsche Forschungsgemeinschaft (DFG: Scho 336/4-2) and DAAD (A.B.) is gratefully acknowledged. We would like to thank Martin Giese, Axel Steinhage and Hubert Dinse for their contributions to our understanding of neural population representations.

Figure captions

1. Responses of the dynamic field model Eq. 2 to stimuli presented at seven different locations of the visual field. Input specifying one of the seven sites marked by arrows in the top panel evokes a localized activation pattern centered at that site (top). Only normalized field activity above resting level (level of activation without input) is plotted. When two of the seven inputs are applied simultaneously (arrows, bottom panel) a bimodal response pattern results (solid line on bottom). Inhibitory interaction reduces the total activation compared to the superposition of the superposition of the individual responses to the two component stimuli (dashed lines on bottom). The gradient from excitatory to inhibitory interaction with increasing distance between stimulus positions leads to repulsion of the two peaks from each other.
2. Distributions of population activation (DPAs) were constructed based on neural activity of 178 neurons recorded in primary visual cortex in response to seven localized elementary stimuli using a direct, receptive field based method in two dimensions or a optimal linear estimator method in one dimension. The width of these DPAs is shown here (receptive field based method: triangles; optimal linear estimator method: squares). Width was quantified by computing the theoretical standard deviation (SD) of the DPAs treated as probability distributions. These measures of width are shown in time slices of 10 ms for both estimators. Means over the 7 elementary stimuli are shown with error bars representing standard deviation. The observed increase is evidence for the broadening in time of the DPAs generated by phasic stimuli.
3. Activation in the linear optimization based DPA is integrated over a zone of 0.8 deg width around the left-most component stimulus. The time course of this activation level (solid line) is shown for the 6 different composite stimuli, each of which has a component at the shown left-most site and a second component

at a distance of (from top to bottom) 0.4, 0.8., 1.2, 1.6, 2.0, and 2.4 deg. For comparison, the same type of locally integrated activation variable is shown for the DPA observed in response to the left-most stimulus component presented in isolation (dashed line, same in all plots). Early excitatory interaction effects are in evidence for narrowly spaced stimuli, while the late inhibitory interaction effect persists at all stimulus separations.

4. The linear optimization based DPA (solid line) is shown for the three composite stimuli with largest interstimulus separation (1.6, 2.0, and 2.4 deg separation from left to right, positions of component stimuli indicated by vertical lines). Superpositions of the two DPAs obtained when the two component stimuli are presented in isolation are shown for comparison (dashed lines). Note the overall inhibitory interaction effect (reduced level of activation compared to superposition) and the repulsion effect (peaks repelled from each other compared to peaks of superposition).
5. The spatial repulsion effect is illustrated for the two dimensional DPAs based on receptive fields for the two largest stimulus separations (2.0 deg in (a) and 2.4 deg in (b)). The section of the visual field probed in the experiment is indicated by the outer frame. The two small squares represent the two component stimuli. Spatial contours of the DPAs are shown at an activation level corresponding to 50% of peak activation. The solid line represents the DPA of the composite stimuli, the dashed line the superposition of the two DPAs obtained when each component stimulus was presented separately. Crosses indicate the positions of the maxima of the DPAs of the composite stimuli, triangles the maxima of the superposed DPAs.
6. Optimal linear estimation is applied to the simulations shown in Figure 1. Seven Gaussians (lower cutoff and amplitude normalized) representing the seven elementary stimuli served as target DPAs. The seven field solutions ob-

tained in response to those stimuli (top panel of Figure 1) served as reference conditions. Least squares optimization determined the basis functions, $f_i(x)$, that each field site, i , contributed to the DPA. This led to good reconstruction of the shape and localization of the field in the reference conditions (bottom left panel). Keeping the basis functions, $f_i(x)$, fixed, the OLE based reconstruction also captured the inhibitory and repulsion effects predicted from the dynamic field model (original simulation repeated on top right, reconstruction on bottom right).

7. DPAs based on linear optimization representing movement direction were computed from activities of 117 neurons recorded in motor and premotor cortex of a macaque monkey. A movement to target 1 was performed following the response signal (“RS”). On the top, the preparatory signal (“PS”) provided complete prior information about the upcoming movement, target 1, whereas on the bottom, prior information specifying targets 1 and 2 was provided. The DPAs are constructed as functions over the parameter axis “direction” and are plotted as functions of time. The 6 movement directions are indicated by target numbers. Note that the DPA is a periodic function of movement direction. For more details, see text.
8. DPAs were computed from neural activity integrated over three epochs (the first 500 msec of the period between the preparatory signal and the response signal, the last 500 msec of that period, and the reaction time period between the response signal and movement onset) and three conditions of prior information (see inset).
9. The circular concentration is used to quantify the sharpness of the DPAs in two epochs (second part of the preparatory period, PP2, and reaction time period, RT). Higher values of concentration reflect sharper DPAs. All DPAs were brought to the same level of total activation by additive normalization,

such that concentration reflects shape independently of total activation. Note that all DPAs are sharper during the reaction time epoch than during the preparatory period, reflecting the specificity of these patterns of activation to the upcoming movement. The larger the range of precued targets, the broader the DPAs. This provides evidence that the shape of the DPAs reflects the range of prior information available about the upcoming movement direction.

References

- [1] S Amari. Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, 27:77–87, 1977.
- [2] C H Anderson. Basic elements of biological computational systems. *International Journal of Modern Physics C*, 5:135–137, 1994.
- [3] A Bastian, A Riehle, W Erlhagen, and G Schöner. Prior information pre-shapes the population representation of movement direction in motor cortex. *NeuroReport*, 9:315–319, 1998.
- [4] R Ben-Yishai, R L Bar-Or, and H Sompolinsky. Theory of orientation tuning in visual cortex. *Proceedings of the National Academy of Sciences USA*, 92:3844–3848, 1995.
- [5] A DAS, C D Gilbert. Distortions of visuotopic map match orientation singularities in primary visual cortex. *Nature*, 387:594–598, 1997.
- [6] R P Erickson. Parallel “population” neural coding in feature extraction. In F O Schmitt and F G Worden, editors, *The neurosciences — Third study program*, pages 155–169, 1974.
- [7] W Erlhagen and G Schöner. Dynamic field theory of movement preparation. *submitted*.
- [8] N I Fisher. *Statistical analysis of circular data*. Cambridge University Press, 1993.
- [9] A P Georgopoulos. Motor cortex and cognitive processing. In M S Gazzaniga, editor, *The cognitive neurosciences*, pages In M S Gazzaniga, editor, *The cognitive neurosciences*, pages 507–517, Cambridge, MA, 1995. MIT Press – A Bradford Book.

- [10] A P Georgopoulos, J F Kalaska, M D Crutcher, R Caminiti, and J T Massey. The representation of movement direction in the motor cortex: Single cell and population studies. In G M Edelman, W E Gall, and W M Cowan, editors, *Dynamic aspects of Neocortical Function*, pages 501–524. John Wiley, New York, 1984.
- [11] A P Georgopoulos, J T Lurito, M Petrides, A B Schwartz, and J T Massey. Mental rotation of the neural population vector. *Science*, 243:1627–1630, 1989. reprinted in Kelner, K L, Koshland, D E Jr. (eds.): *Molecules to Models*, AAAS, 1989.
- [12] A P Georgopoulos, A B Schwartz, and R E Kettner. Neural population coding of movement direction. *Science*, 233:1416–1419, 1986.
- [13] C C A M Gielen, R Glasius, and A Komoda. Interpretation of neuronal activity in neural networks. *Journal of Neurocomputing*, in press.
- [14] C C A M Gielen, G H F M Hesselmanns, and P I M Johannesma. Sensory interpretation of neural activity patterns. *Mathematical Biosciences*, 88:14–35, 1988.
- [15] A Grinvald, E E Lieke, R D Frostig, R Hildesheim. Cortical point-spread function and long -range Lateral interactions revealed by real-time optical imaging of macaque monkey primary visual cortex. *Journal of Neuroscience*, 14:2545–2568.
- [16] J M Groh, R T Born, and W T Newsome. How is a sensory map read out? effects of microstimulation in visual area MT on saccades and smooth pursuit eye movements. *Journal of Neuroscience*, 17(11):4312–4330, 1997.
- [17] S Grossberg. Biological competition: Decision rules, pattern formation, and oscillations. *Proceedings of the National Academy of Sciences (USA)*, 77:2338–2342, 1980.

- [18] D Jancke, A C Akhavan, W Erlhagen, M A Giese, A Steinhage, Schöner G, and H R Dinse. Population coding in cat visual cortex reveals nonlinear interactions as predicted by a neural field model. In C von der Malsburg, W von Seelen, J C Vorbrüggen, and B Sendhoff, editors, *Artificial neural networks – ICANN 96*, pages 641–648. Springer, Berlin, 1996.
- [19] D Jancke, W Erlhagen, H R Dinse, A C Akhavan, M Giese, A Steinhage, and G Schöner. Parametric population representation of retinal location: Neuronal interaction dynamics in cat primary visual cortex. *Journal of Neuroscience*, in press.
- [20] C. Lee, W.H. Rohrer, and D.L. Sparks. Population coding of saccadic eye movements by neurons in the superior colliculus. *Nature*, 332:357–360, 1988.
- [21] G A Orban. *Neuronal operations in the visual cortex*. Springer Verlag, Berlin, 1984.
- [22] A Pouget, K Zhang, S Deneve, and P E Latham. Statistically efficient estimation using population coding. *Neural Computation*, 10:373–401, 1998.
- [23] E Salinas and L F Abbott. Vector reconstruction from firing rates. *Journal of Computational Neuroscience*, 1:89–107, 1994.
- [24] E Salinas and L F Abbott. Transfer of coded information from sensory to motor networks. *The Journal of Neuroscience*, 15:6461–6474, 1995.
- [25] T D Sanger. Propability density estimation for the interpretation of neural population codes. *Journal of Neurophysiology*, 76:2790–2793, 1996.
- [26] G Schöner, K Kopecz, and W Erlhagen. The dynamic neural field theory of motor programming: Arm and eye movements. In P G Morasso and V Sanguineti, editors, *Self-Organization, Computational Maps and Motor Control, Psychology Series, Vol. 119*, pages 271–310. Elsevier-North Holland, 1997.

- [27] S H Scott and J F Kalaska. Changes in motor cortex activity during reaching movements with similar hand paths but different arm postures. *Journal of Neurophysiology*, 73:2563–2567, 1995.
- [28] H S Seung and H Sompolinsky. Simple models for reading neuronal population codes. *Proceedings of the National Academy of Science (U.S.A.)*, 90:10749–10753, 1993.
- [29] R G Szuborski and L A Palmer. The two-dimensional spatial structure of non-linear subunits in the RF’s of complex cells. *Vision Research*, 30:149–154, 1990.
- [30] D J Tolhurst and D J Heeger. Comparison of contrast-normalization and threshold models of the responses of simple cells in cat striate cortex. *Visual Neuroscience*, 14:193–309, 1997.
- [31] A H C van der Heijden, J Müsseler, and G Bridgeman. On the perception of position. In G Aschersleben, T Bachmann, and Müsseler J, editors, *Cognitive contributions to the perception of spatial and temporal events*. Elsevier Science B.V., in press.
- [32] H R Wilson and J D Cowan. A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik*, 13:55–80, 1973.
- [33] R S Zemel, P Dayan, and A Pouget. Probabilistic interpretation of population codes. *Neural Computation*, 10:403–430, 1998.
- [34] K Zhang. Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: A theory. *Journal of Neuroscience*, 16:2112–2126, 1996.

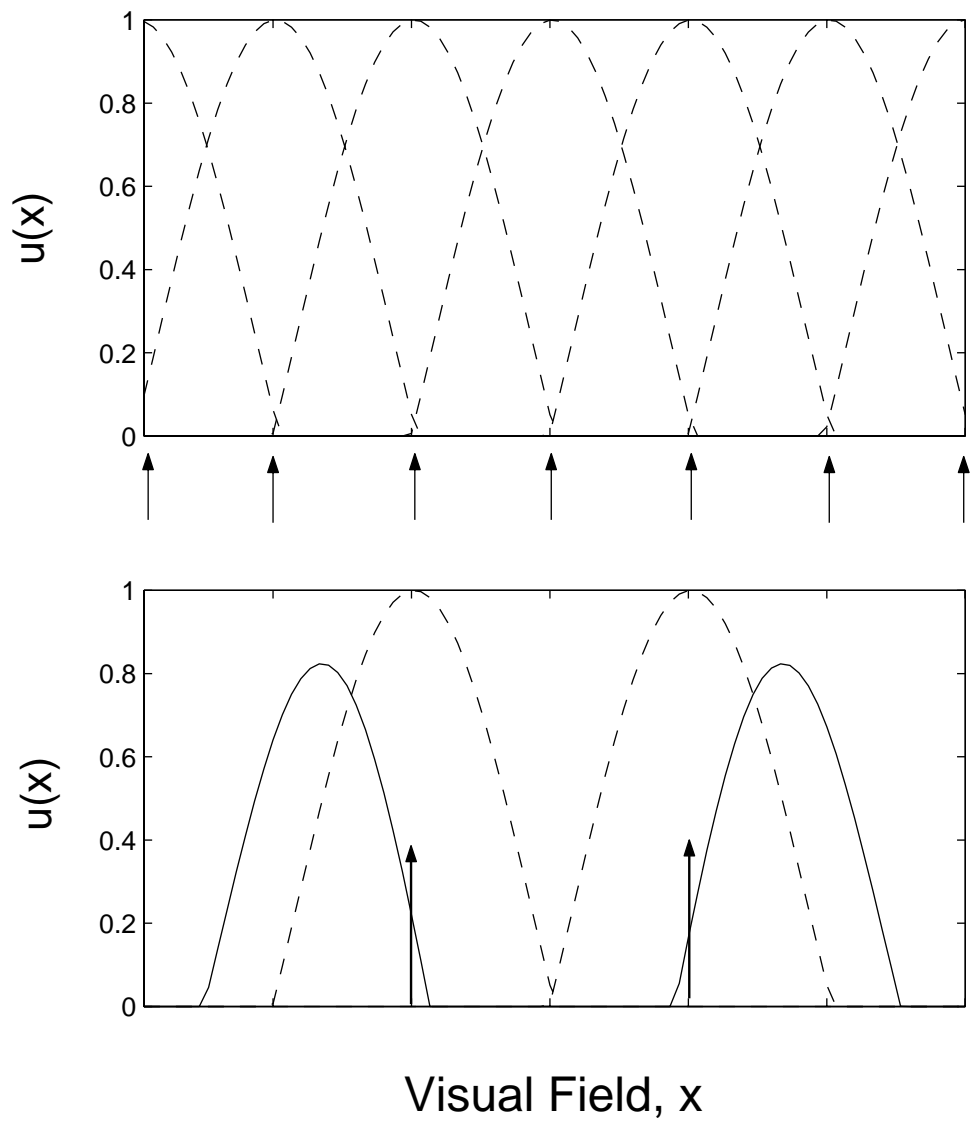


Figure 1:

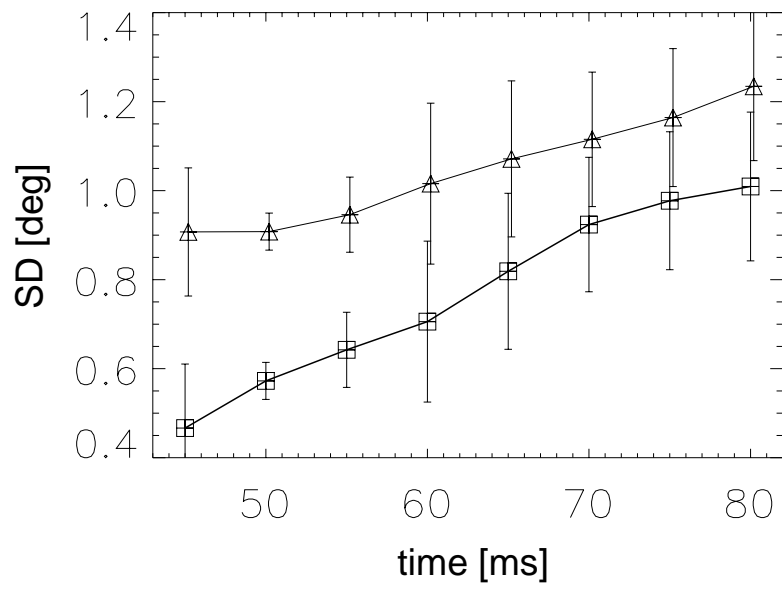


Figure 2:

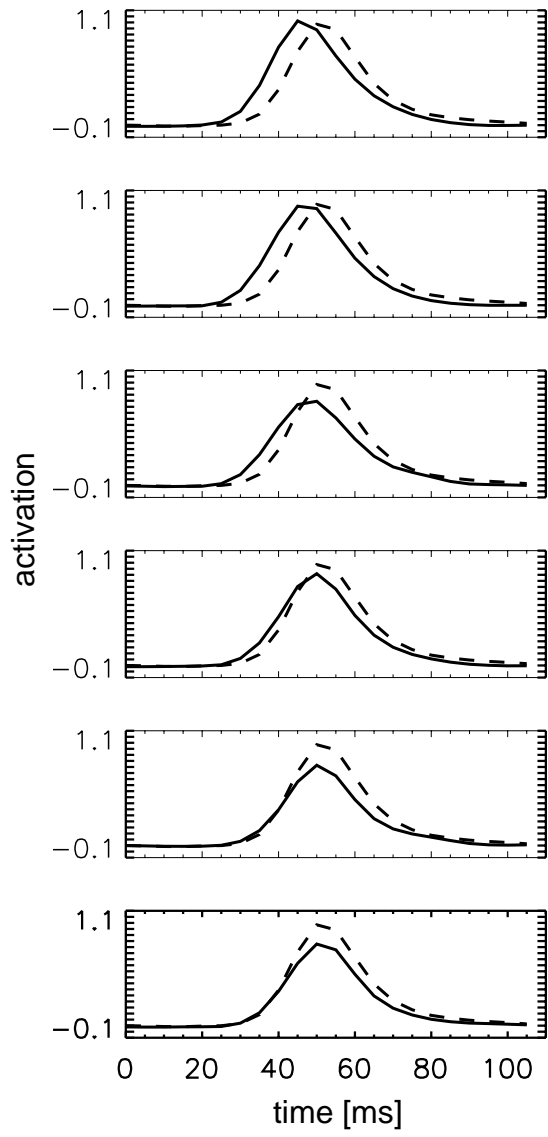


Figure 3:

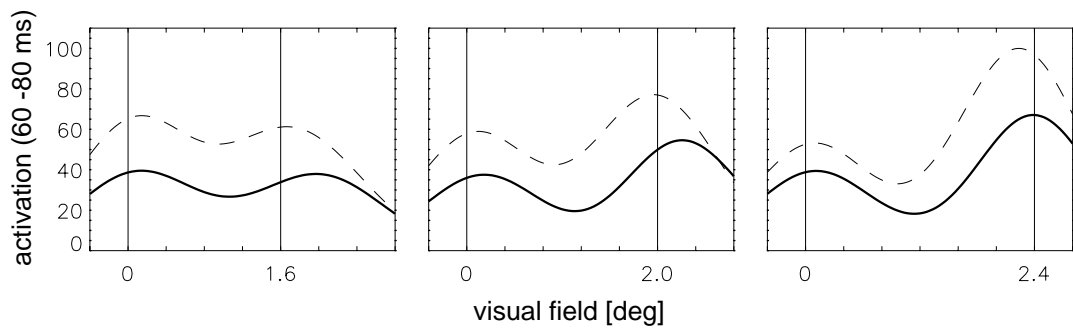
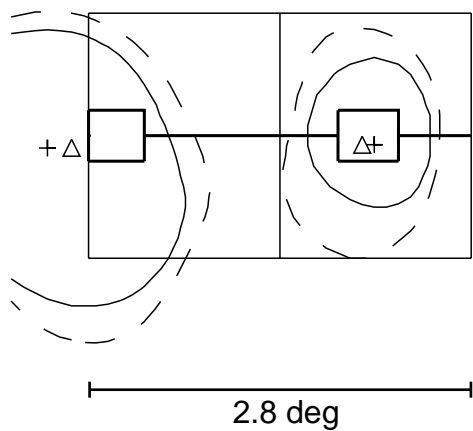


Figure 4:

a



b

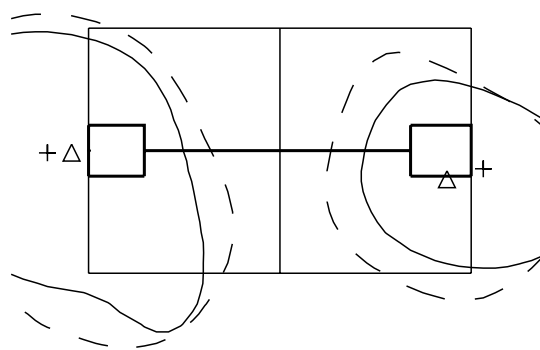


Figure 5:

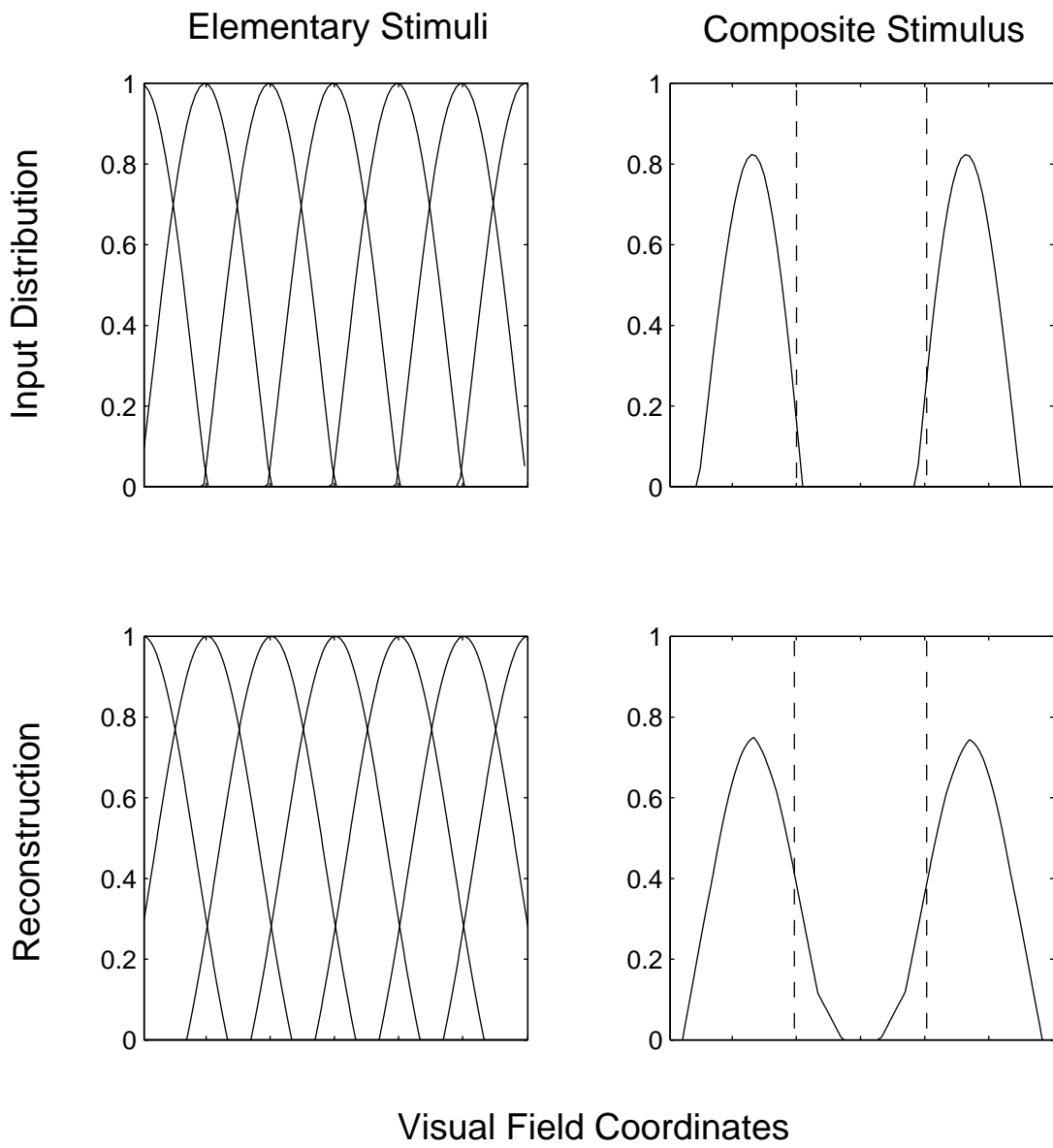
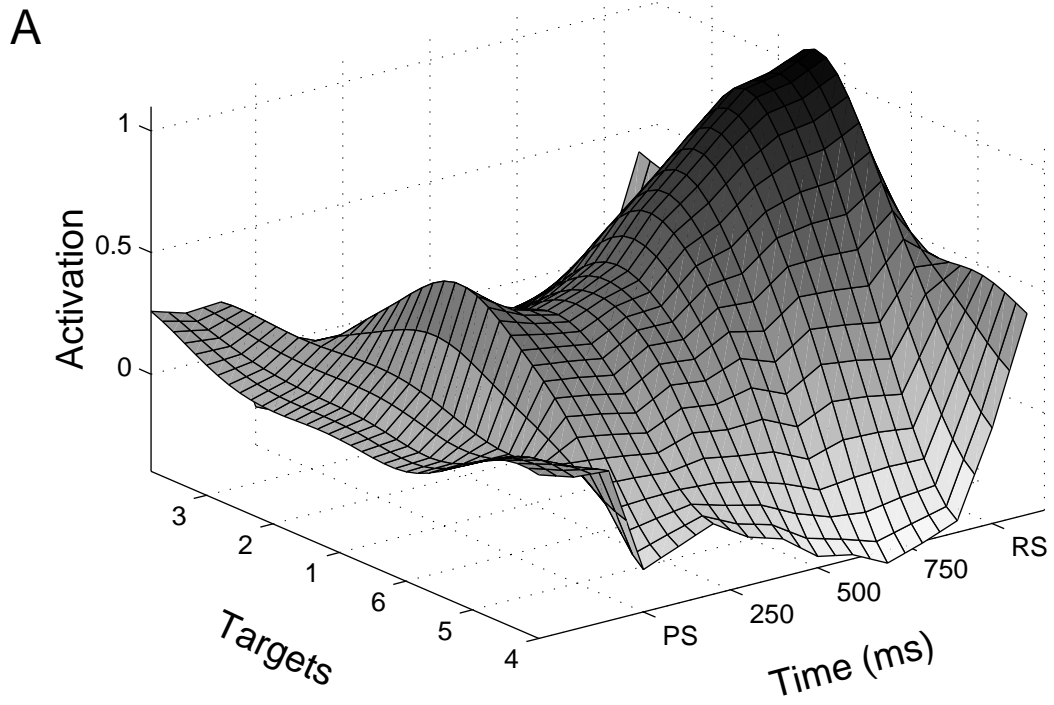


Figure 6:

Target 1
One target information (tg1)



Two target information (tg1&2)

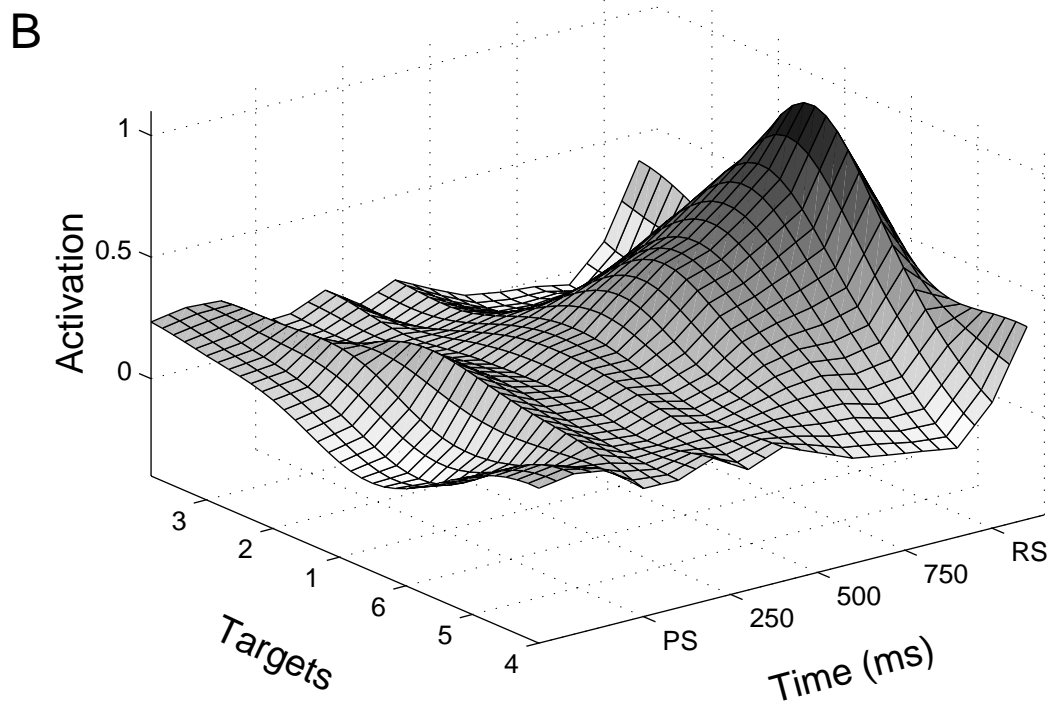


Figure 7:
36

Target 1

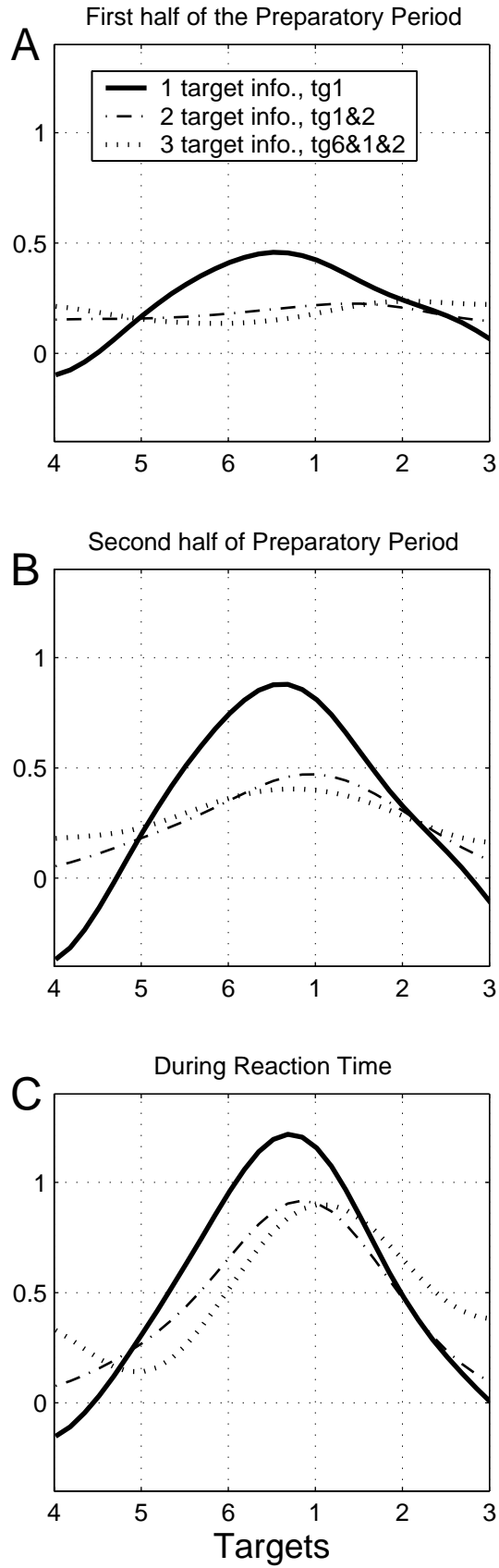


Figure 8:

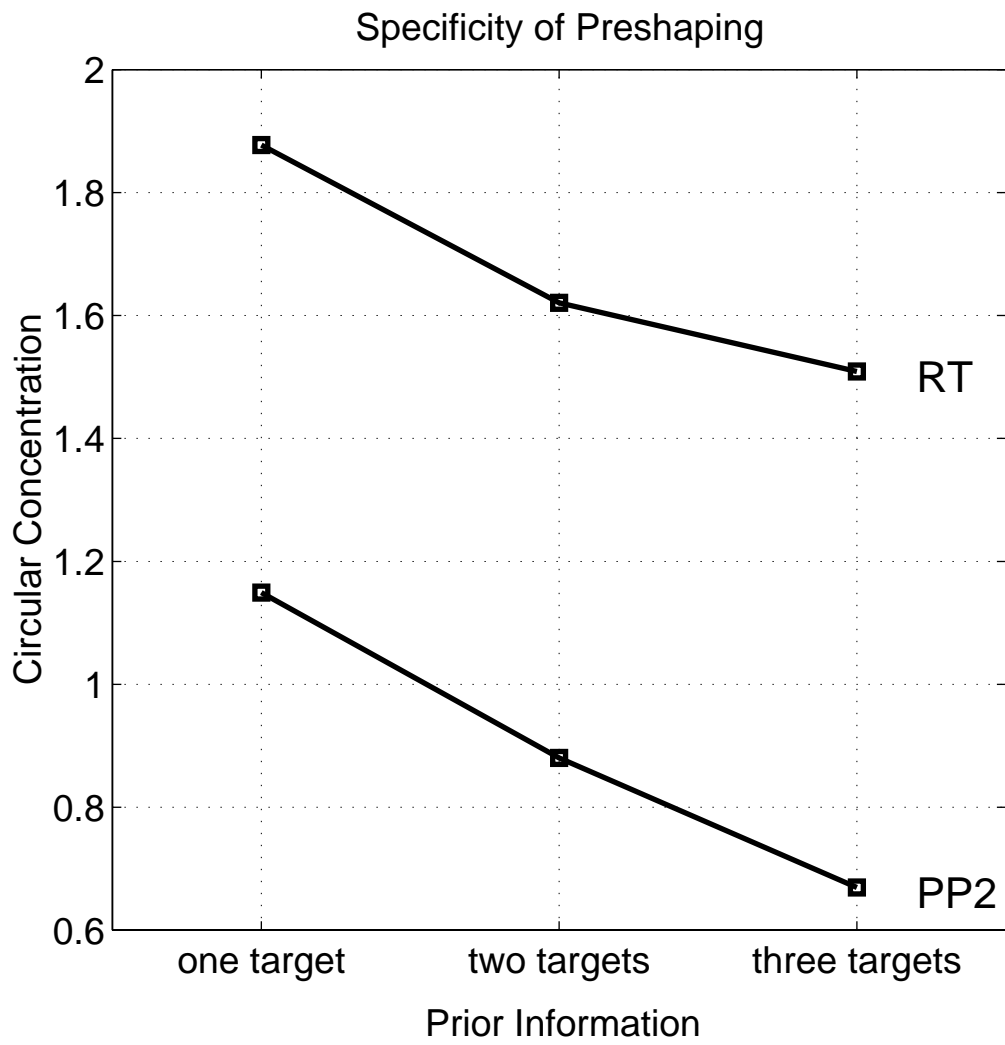


Figure 9: