Available online at www.sciencedirect.com
ScienceDirect
Journal of Physiology - Paris 101 (2007) 22–31
Journal of Physiology Paris
www.elsevier.com/locate/jphysparis

Coding processes involved in the cortical representation of complex tactile stimuli

Jean-Luc Blanc *, Jacques-Olivier Coq

UMR 6149, CNRS – Aix-Marseille Université, Centre St. Charles, Pôle 3C, case B, 13331 Marseille Cedex 03, France

Abstract

To understand how information is coded in the primary somatosensory cortex (S1) we need to decipher the relationship between neural activity and tactile stimuli. Such a relationship can be formally measured by mutual information. The present study was designed to determine how S1 neuronal populations code for the multidimensional kinetic features (i.e. random, time-varying patterns of force) of complex tactile stimuli, applied at different locations of the rat forepaw. More precisely, the stimulus localization and feature extraction were analyzed as two independent processes, using both rate coding and temporal coding strategies. To model the process of stimulus kinetic feature extraction, multidimensional stimuli were projected onto lower dimensional subspace and then clustered according to their similarity. Different combinations of stimuli clustering were applied to differentiate each stimulus identification process. Information analyses show that both processes are synergistic, this synergy is enhanced within the temporal coding framework. The stimulus localization process is faster than the stimulus feature extraction process. The latter provides more information quantity with rate coding strategy, whereas the localization process maximizes the mutual information within the temporal coding framework. Therefore, combining mutual information analysis with robust clustering of complex stimuli provides a framework to study neural coding mechanisms related to complex stimuli discrimination.

Keywords: Information theory; Neural coding; Stimulus feature extraction; Primary somatosensory cortex

1. Introduction

A major challenge in Neuroscience is to decipher how neural activity represents the physical features of objects with which animals interact. For instance, each degree of roughness scanned by whiskers corresponds both to a unique kinetic signature defined by a temporal profile of whisker velocity and to a distinct firing pattern, based on spike counts (Arabzadeh et al., 2005, 2006). Although most of the studies to date have used rate coding, several authors have emphasized the key role of spike timing in neural population coding (Borst and Theunissen, 1999). Using precise spike timing, the first spikes have been shown to transmit larger quantities of information about stimuli than the same spikes in rate coding, either in the barrel cortex (Panzeri et al., 2001) or in the cortical forepaw representation (Foffani et al., 2004).

Instead of whiskers, rats can use their forepaws to perceive object features, such as location, size, shape and texture (Bourgeon et al., 2004; Iwaniuk and Whishaw, 2000). We know that the forepaw representation in the S1 cortex is topographically organized (Coq and Xerri, 1998). This topographic organization provides a spatial frame of reference for location detection of stimuli applied on different forepaw locations. However, the large range of spatiotemporal responses in the S1 forepaw cortex (Tutunculer et al., 2006), such as found in the barrel cortex, may allow the extraction of the multidimensional kinetic features of objects or complex tactile stimuli.

Neural coding refers to how the central nervous system represents sensory information as patterns of action potentials emitted by neuronal populations. The neural coding problem is often formulated in terms of quantitizing a joint

* Corresponding author.
E-mail address: jlblanc@up.univ-mrs.fr (J.-L. Blanc).

0928-4257/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jphysparis.2007.10.004

space $(R;S)$ (Mumey et al., 2004; Slonim et al., 2006; Nadal, 2002) where S represents the input sensory stimuli and R the set of possible neural activity patterns. Both of these spaces are high-dimensional and complex. We consider the sensory system robust and adaptive, in that it must represent similar stimuli in similar ways. Thus, individual input stimuli are not important for understanding neural function, but rather classes of input stimuli and their correspondence are the key to decipher the neural representation of complex stimuli. Following this idea, this study refers to a model in which neurons are selective for a small number of stimulus dimensions out of a high-dimensional stimulus space, and within this subspace similar sensory signals are clustered. The idea that Shannon's Information Theory (Shannon, 1948) is relevant for studying neural coding goes back to Attaneve (1954) and has received considerable attention these last few years (Bialek et al., 1991; Atick, 1992; Borst and Theunissen, 1999). In this paper first of all, we report generality on information theory in the neural coding context. Secondly, we present another information quantity: the "multi-information", useful to study neural representations of complex stimuli.

In this paper the simultaneous activity of S1 neuronal populations was recorded to explore the neural coding of location detection and kinetic feature extraction of complex stimuli, based on either spike count or spike timing. Location detection is related to the different sites of forepaw stimulation, while feature extraction refers to a compression process of the high-dimensional kinetic values (i.e. random, time-varying patterns of force) of complex tactile stimuli, as encountered in the natural environment. Are stimulus localization and feature extraction independent processes? What is the time course of these processes after stimulus onset, depending on rate or temporal coding? We used mutual information (MI) to measure the stimulus-response relationship, and different stimuli clustering strategies to separate each neural process. To our knowledge, this study is the first attempt to compare these two neural processes of stimulus identification and to use "multi-information" in the neural context.

2. Methods

2.1. Information theoretic framework

2.1.1. Information carried by neuronal population response

Mutual information is a rigorous criterion to quantify how much information the neural responses convey about a sensory stimuli set (Bialek et al., 1991). We consider a time window T, associated with a sensory stimulus s chosen with a probability $p(s)$ from a stimulus set $S = \{\mathbf{s_1},\mathbf{s_2},\ldots,\mathbf{s_m}\}$, during which the activity of C neurons is recorded. The neuronal population response is denoted by the random variable $R = \{\mathbf{r_1},\mathbf{r_2},\ldots,\mathbf{r_n}\}$, where each component of the vector $\mathbf{r_n} = [r_n^1, r_n^2, \ldots, r_n^C]$ is the response of one neuron of the population within the time window T. Each neuronal response can be differently described depending on the coding framework. In a spike count code, the response is the number of spikes within the time window T. In a spike timing code, the response is a sequence of spike firing times. The MI between R and S is defined as the difference between the Shannon response entropy and the noise entropy (Shannon, 1948).

$$I(R;S) = H(R) - H(R|S) \tag{1}$$

where the entropy is

$$H(R) = -\sum_{\mathbf{r}\in R} p(\mathbf{r}) \log p(\mathbf{r})$$

and the noise entropy

$$H(R|S) = -\sum_{s\in S} p(s) \sum_{\mathbf{r}\in R} p(\mathbf{r}|s) \log p(\mathbf{r}|s)$$

The mutual information can be written as

$$I(R;S) = \sum_{s\in S} p(s) \sum_{\mathbf{r}\in R} p(\mathbf{r}|s) \log \frac{p(\mathbf{r}|s)}{p(\mathbf{r})} \tag{2}$$

$p(\mathbf{r}|s)$ is the probability of simultaneously observing a particular response $\mathbf{r}$ conditional to the stimulus s, and $p(\mathbf{r}) = \sum_{s\in S} p(s)p(\mathbf{r}|s)$ is its average across all stimuli.

2.1.2. Mutual information for three random variables

In this section we address inequalities involved in Shannon's information measures, for three discrete random variables. A region in $\mathbb{R}^{2^n-1}$, denoted by Γ^*, is identified to be the origin of all information inequalities involving n random variables in the sense that all such inequalities are partial characterizations of Γ^*. Further theoretical precisions can be found elsewhere (Yeung, 1997).

In this study we address the issue of dependence between two processes: stimulus feature extraction and stimulus localization. Within this framework, we introduce two other variables K,L describing the stimulus set S (i.e. $K = f(S)$ and $L = g(S)$). K is the random variable which describes the stimulus kinetic features extraction $K = \{k_1,k_2,\ldots,k_K\}$ and L is the random variable which describes the different stimuli locations $L = \{l_1,l_2,\ldots,l_L\}$. Using the chain-rule of information (Cover and Thomas, 1991) we can develop the MI

$$\begin{aligned} I(R;K,L) &= I(R;K) + I(R;L|K), \\ I(R;K,L) &= I(R;L) + I(R;K|L) \end{aligned} \tag{3}$$

where $I(R;K,L)$ is the MI between the neural response R and the intersection of both stimuli sets K and L (Fig. 1). This quantity can be rewritten using the mutual information definition equation (1)

$$I(R;K,L) = \sum_{k\in K} \sum_{l\in L} p(k,l) \sum_{\mathbf{r}\in R} p(\mathbf{r}|k,l) \log \frac{p(\mathbf{r}|k,l)}{p(\mathbf{r})} \tag{4}$$

where $p(\mathbf{r}) = \sum_{k\in K} p(k) \sum_{l\in L} p(l) p(\mathbf{r}|k,l)$

$I(R;K)$ is the MI between the neural response and the stimulus, considering only the kinetic features, so that the stimuli are clustered according to their time-varying intensity. $I(R;L)$ is the MI between the neural response and the stimulus, considering only its location. $I(R;L|K)$ is the MI between the neural response and the stimulus location knowing its kinetic features. $I(R;K|L)$ is the MI between the response and the stimulus kinetic features knowing its location.

The non-negativity property of the MI (Cover and Thomas, 1991), gives the following expressions

$$\begin{aligned} I(R;K,L) &\geqslant I(R;K), \\ I(R;K,L) &\geqslant I(R;L) \end{aligned} \tag{5}$$

From these inequalities we can write a lower bound of the MI between the neural response and both the stimulus kinetic feature extraction and stimulus localization

$$I(R;K,L) \geqslant \frac{1}{2}[I(R;K) + I(R;L)] \tag{6}$$

Suppose the random variables $R \to K \to L$ form a Markov chain, then the data-processing inequality (Cover and Thomas, 1991) gives

$$\begin{aligned} I(R;K|L) &\leqslant I(R;K) \\ I(R;L|K) &\leqslant I(R;L) \end{aligned} \tag{7}$$

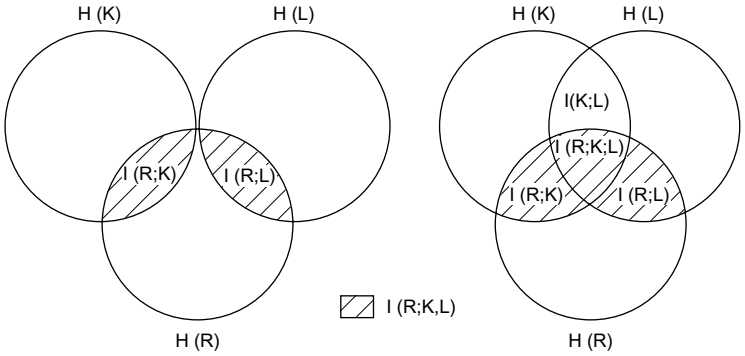

Fig. 1. Venn diagrams that explain the introduction of variables k and l describing each stimulus s. Left panel. The random variables K and L are independent. Right panel. The variables K and L are dependent or synergistic. $I(R;K,L)$ is the mutual information between the neural responses (variable R) and the intersection of sets K and L with the set R.The "multi-information" $I(R;K;L)$ corresponds to the intersection between the sets R, K and L.

Note that these information inequalities are symmetric. We can now write an upper bound of the MI using the definition (3) and (7)

$$I(R;K,L) \leqslant I(R;K) + I(R;L) \tag{8}$$

With equality only if the random variables K and L are independent. Thus, under this regime the mutual information is framed by

$$\frac{1}{2}[I(R;K) + I(R;L)] \leqslant I(R;K,L) \leqslant I(R;K) + I(R;L) \tag{9}$$

Note that if R,K,L do not form a Markov chain, then it is also possible that

$$\begin{aligned} I(R;K|L) &\geqslant I(R;K), \\ I(R;L|K) &\geqslant I(R;L) \end{aligned} \tag{10}$$

Based on this assumption, for this special case we can write a lower bound of the mutual information

$$I(R;K,L) \geqslant I(R;K) + I(R;L) \tag{11}$$

Using the Venn diagrams in Fig. 1 we defined the information quantity for three random variables as

$$\begin{aligned} I(R;K;L) &= I(R;K) - I(R;K|L), \\ I(R;K;L) &= I(R;L) - I(R;L|K) \end{aligned} \tag{12}$$

We preferred to call this measure "multi-information" rather than mutual information because in contrast to mutual information, which is always positive, multi-information can be either positive or negative. This is possible since the effect of holding one of the variables may increase or decrease the dependency between the others. As a trivial case, consider the situation in the trivariate product where variables K and L are independent when R is not known, but become dependent given R, in this case, $I(R;K;L)$ is clearly negative. Han (1980) has shown that multi-information need not be always positive, by expanding it in terms of parameters of probability up to the second order.

In this study we define a "synergy threshold" as

$$I(R;K;L) \leqslant 0 \tag{13}$$

With $I(R;K;L) = 0$, if R, K and L are mutually independent.

The larger the negativity of the quantity $I(R;K;L)$ the stronger the synergy between the localization and kinetic feature extraction processes. We then explore this hypothesis estimating the different information quantities from neuronal responses to complex tactile stimuli.

2.2. Experimental data acquisition

Multi-unit neuronal activity was recorded in the S1 cortex and complex tactile stimuli were applied to the forepaw in four anesthetized rats. All experiments were carried out in accordance with National Institutes of Health Guidelines for the Care and Use of Laboratory Animals. The details of the surgical preparation and multi-unit recordings are described elsewhere (Coq and Xerri, 1998). Briefly, adult Long-Evans rats were anesthetized with pentobarbital sodium with an initial dose of 50 mg kg^{-1} (i.p.). Supplementary doses (5 mg kg^{-1} i.p.) were given as needed to keep the rats at an areflexive level of anesthesia throughout the experiments by monitoring the heart rate, spontaneous whisker movements, eye-blinking and paw-withdrawal reflexes. The core temperature was continuously monitored by a rectal thermistor probe and was maintained at around 38 °C by a heating pad. A craniotomy (about 16 mm^2) was done with bregma as the initial point of reference (ant. 3.0 mm; post. 1.0 mm; lat. 3–6 mm) to expose the somatosensory cortex. The dura was incised and resected. The exposed somatosensory cortex surface was covered with warm silicone fluid (30,000 cs) to prevent drying and oedema. At the end of the experiment, the animal received a lethal injection of pentobarbital sodium (150 mg kg^{-1} i.p.).

Multi-unit activity was recorded with parylene-coated tungsten electrodes (1 MΩ at 100 Hz) in layer III-IV of the S1 cortex at a depth of 600–700 μm. A single electrode was moved perpendicular to the cortical surface at different locations to map the S1 forepaw representation of interest: from palmar pad 1–digit 5 (about 1.6 mm in length) in the rostrocaudal direction. A digitized image of the cortex was used to place the microelectrode penetrations which were identified relative to the cortical vasculature. The multi-unit signal was preamplified, filtered (bandwidth 0.5–5 kHz) and displayed on an oscilloscope. Under our recording conditions, the amplitude of the background noise usually ranged from 10 to 20 μV, with a signal-to-noise ratio of 5–10.

After the S1 area of interest was located and roughly mapped using single electrodes, the activity of neuronal populations was then simultaneously recorded in layers III–IV using one row-arrays of eight electrodes (MicroProbe, MD) and a multichannel acquisition system at a frequency of 20 kHz (Plexon, TX). The array was made of eight platinium/irridium electrodes (2 MΩ at 100 Hz) spaced by 200 μm from each other. The electrode array was inserted along the rostrocaudal axis of the S1 forepaw representation to sample neurons whose receptive fields were located from digits 2–5 and from palmar pads 1–3 (Fig. 2), according to the single electrode mapping described above. The activity signal recorded with the electrode array was processed and monitored as well as the single electrode activity and its quality was equivalent to what was described above. We generally recorded 2–5 units simultaneously for each electrode of the array. Two neurons per electrode with well-separated spikes were sorted using Offline Sorter (Plexon), so that we obtained 16 neurons per animal. In this paper, MI estimations were based on eight sets of eight neurons: one neuron out of the two recorded with each electrode of the array, and two sets of eight neurons pseudo-randomly chosen from the 16 neurons for each rat. To calculate MI, the spike trains of each neuron in a 150 ms time window for each trial, were represented in 4 ms bins, with ones for each spike and zeros elsewhere.

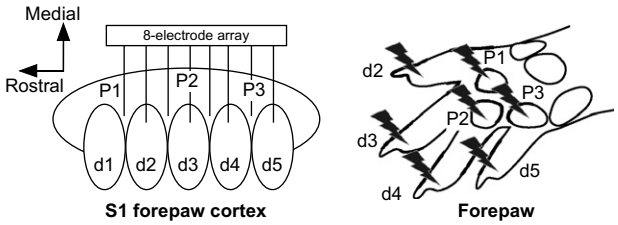

Fig. 2. The activity of S1 neuronal populations was recorded simultaneously in response to single complex tactile stimuli applied to the rat forepaw. Left panel. Schematic drawing of the topographic organization of the S1 cortical representation of the glabrous forepaw skin surfaces on which has been superimposed the implantation location of the 8-electrode array. Right panel. Schematic drawing of the rat forepaw with the locations (black symbols) of single stimuli applied on digit tips 2–5 (d2–d5) and palmar pads 1–3 (P1–P3). The recorded neurons had their receptive field distributed along the 7-forepaw locations of stimulation.

Simultaneous neuronal responses were recorded while single complex stimuli (or trials) were successively applied on each of seven different locations (from digit tip 2–5 and from palmar pad 1–3) of the rat forepaw (Fig. 2). These stimulus locations corresponded to the recorded neurons' receptive fields. The complex stimuli were applied using a hand-held electronic Von-Frey device, designed in our laboratory to apply single, punctuate stimuli through a probe (0.8 mm in diameter) onto the forepaw skin surfaces. Usually, Von-Frey hairs are used to determine stimulus detection thresholds on the skin. The force envelope randomly applied (mN) by hand on the skin was converted by the electronic Von-Frey into a potential variation (1 mN corresponded to 10 mV) which was recorded at 1 kHz with the Plexon multichannel acquisition system. Stimuli and neuronal responses were synchronized using a TTL signal emitted by the Von-Frey at the onset of the skin contact. The force pattern, shape and duration (ranging from 130 to 250 ms) of the complex stimuli varied for each trial (Fig. 3, left panels). The pseudo-random variability of our stimuli was assumed to reproduce the ethological variability of stimulation in the natural environment. Neuronal activity was recorded in a 150 ms time window which started from the trial onset. Stimuli shorter than 150 ms were discarded from analyses. We elicited about 1300 trials on each of the seven forepaw locations and we finally obtained a stimuli set S of $n = 9100$ trials in each animal.

2.3. Clustering of complex tactile stimuli

To address the issue of the relative independence between the processes of location detection and kinetic feature extraction of complex stimuli, two sets of variables L and K were introduced to describe the set S of stimuli. The sets L and K were respectively attributed to the processes of stimulus localization and feature extraction. The seven clusters of stimuli from the variable L corresponded to the seven locations of the forepaw stimulation (Fig. 4, right upper panel). These stimulations were applied from the digit tip 2–5 and from the palmar pad 1–3 (Fig. 2, right panel).

We assume that the natural environment produces at each instant of time a new stimulus with some probability. This complex stimulus results from the combination of multidimensional features. In our case, the kinetic features of complex tactile stimuli correspond to random, time-varying patterns of force applied onto the rat forepaw skin. As a robust system the central nervous system is able to extract the kinetic features of sensory inputs and then to make clusters of stimuli with similar features.

The stimulus kinetic feature extraction process was modelled by first projecting each stimulus from the whole stimuli set S (i.e. matrix of dimension $n * m$) onto a lower dimensional d subspace ($d < m$) (Fig. 3). Each complex stimulus was initially described by the acquisition system as a vector with m samples $s = [s_1,\ldots,s_m]$. We decomposed each stimulus on an orthogonal basis, finding the eigenvalues and eigenvectors of the sample covariance matrix $Cov = S^tS$. Using a principal component analysis-like method (Lebart et al., 2002), we obtained a shorter stimulus description $Y = As$ where A was the basis formed with the d most significant eigenvectors of Cov. To achieve the stimulus compression (i.e. the kinetic feature extraction), we then performed a general clustering step using a K-Means algorithm (Fig. 3). The K-Means algorithm is an iterative method, used to solve the vector quantization (VQ) problem (Gersho and Gray, 1991). In VQ, a codebook C that includes k code words is used to represent a wide family of stimuli-signals $Y = y^n_{i_{i=1}}$ $(n \gg K)$ by a nearest neighbor assignment. This leads to an efficient compression or description of those signals, as clusters in $\mathbb{R}^n$ surrounding the chosen code words. The VQ problem is often formulated as an optimization problem

$$\min_{C,X} \|Y - CX\|_2^2 \text{ subject to } \forall i,\ \mathbf{x}_i = \mathbf{e}_k \quad \text{for some } k \tag{14}$$

The K-Means algorithm is used for designing the optimal codebook for VQ. In each iteration there are two stages, one for sparse coding that essentially evaluates X by mapping each stimulus/signal to its closest code word (under l^2-norm distance) in C, where $\mathbf{e}_k$ is a vector from the trivial basis, with all zero entries except a one in kth position. The second stage is for updating the codebook, changing sequentially each column c_k in order to better represent the signals mapped to it. At the end of this procedure we obtained k stable clusters of trials.

As the set K summarizes the time-varying components of the complex stimuli, the set L corresponds to the stimuli or trials according to their location when applied on the forepaw skin (Fig. 4). Information theory allows us to consider not only the response variance, but also the conditional probability distributions. To describe the conditional stimuli set $(K|L)$, we used the same procedure as for the set K and then we partitioned each cluster into the seven forepaw locations to obtain $7*k$ new clusters. The other conditional stimuli set $(L|K)$ was obtained by an inverse procedure, in which we performed the clustering for each forepaw location, initializing the K-Means using the same k centers describing the variable K (Fig. 4).

The number of centers k was chosen optimally to obtain homogenous clusters containing similar amounts of stimuli. The clustering procedure was replicated with a new value of k at each time and we retained the solution with the lowest within-cluster sums of points to centroid distances.

2.4. Mutual information estimation methods

In this section we present the different methods we used to estimate the MI with both coding strategies. Each neuronal response was recorded from $C = 8$ neurons within a time window of $T = 150$ ms long. To estimate the information from neural responses, we used a varying time window 0–T ms after the onset of tactile stimuli. We divided the spike train of each cell into bins of size $\Delta t = 4$ ms (Foffani et al., 2004), with each time bin containing 0 or 1 spike. Each possible spike train of each cell hence became a 'binary word' of length $N = T/\Delta t$ (i.e. N is varying 0–38 bins).

2.4.1. Rate coding

We used the amount of information conveyed by spike count responses of neurons. Each response was the sum of spikes in the 0–N bins time window for each neuron. The response vector $\mathbf{r}$ was 8-dimensional, in which each component was the firing rate of the neuron. In this case, the number of possible neuronal responses varied from 1 to NC = 304. We computed the different mutual information probability distributions using a large number of trials ($n = 9100$, see Section 2.2) so the MI with rate coding was weakly biased. However, we used a jacknife procedure (Shao and Dongsheng, 1995) to correct the small biased error due to the different trial numbers used to compute the probability associated with the different clustering strategies (see Section 2.3).

2.4.2. Temporal coding

We also used the amount of information conveyed by the spike timing response of neurons. With this coding strategy the neuronal response vector $\mathbf{r}^c$ for each neuron c, was decomposed as $\mathbf{r}^c = [r^c(1),r^c(2),\ldots,r^c(N)]$,

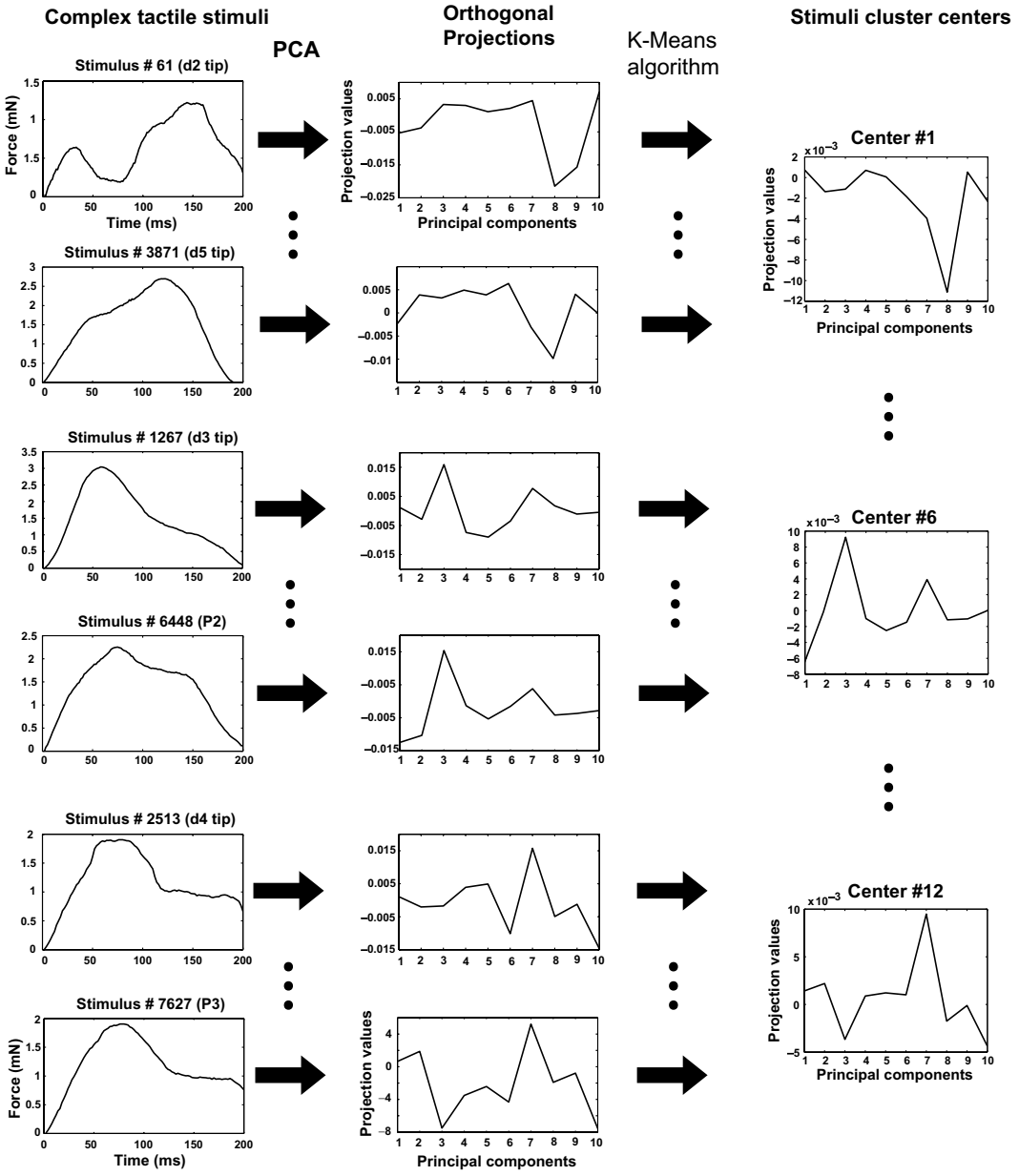

Fig. 3. Schematic view of the stimulus clustering corresponding to the kinetic feature extraction process. The left panels show the force envelope of several complex tactile stimuli, along with their location on the forepaw. The middle panels correspond to the orthogonal projections of the complex stimuli after a principal component analysis (PCA) to reduce the high-dimensionality of these stimuli. The right panels illustrate the centroid plots of some low-dimension stimulus average after the clustering procedure with a *K*-Means algorithm (see text).

where $r^c(t)$ was the response (i.e. 0 or 1 spike) in the time bin t. To compute the MI from the full responses was considered too complex because of the high-dimensionality of the space and of all the possible patterns of responses. Such a high-dimensionality leads to large undersampled probability distributions and to large biases (Panzeri and Treves, 1996; Paninski, 2003).

Numerous approaches have been suggested for the MI estimation. One of them consists of expanding the MI in terms of moments of the probability distribution, until the second order (Panzeri and Schultz, 2001). This approximation has the advantage of clearly separating the different information sources as a sum of components.

$$I(R;S) = I_{\text{lin}} + I_{\text{sig-sim}} + I_{\text{cor-ind}} + I_{\text{cor-dep}} \tag{15}$$

The components of this approximation have different magnitudes and sampling properties, and are discussed elsewhere (Pola et al., 2003). Here we briefly resume the definition of each components.

I_{lin} the ''linear'' component, is the information conveyed by spikes emitted in different time bins and cells independently. $I_{\text{sig-sim}}$ the ''signal similarity'' component quantifies the redundancy arising from similarity across stimuli of the mean response in each time bin. $I_{\text{cor-ind}}$ the ''stimulus independent correlation'' component is the information associated with correlations in neural responses not modulated by the stimulus. $I_{\text{cor-dep}}$ the ''stimulus-dependent correlation'' component is associated with stimulus modulation of correlation.

In our paper, a lower bound approximation of the spike timing information was used. This lower bound (Pola et al., 2005) removes the contribution related to the stimulus-dependent correlations in neural responses from the previous information expansion (15).

$$\begin{aligned} I_B &= I_{\text{lin}} + I_{\text{sig-sim}} + I_{\text{cor-ind}} \\ I_B &= -\sum_{\mathbf{r}\in R} p(\mathbf{r}) log_2 p_{\text{ind}}(\mathbf{r}) - \sum_{s\in S} p(s) \sum_{\mathbf{r}\in R} p_{\text{ind}}(\mathbf{r}|s) log_2 p_{\text{ind}}(\mathbf{r}|s) \end{aligned} \tag{16}$$

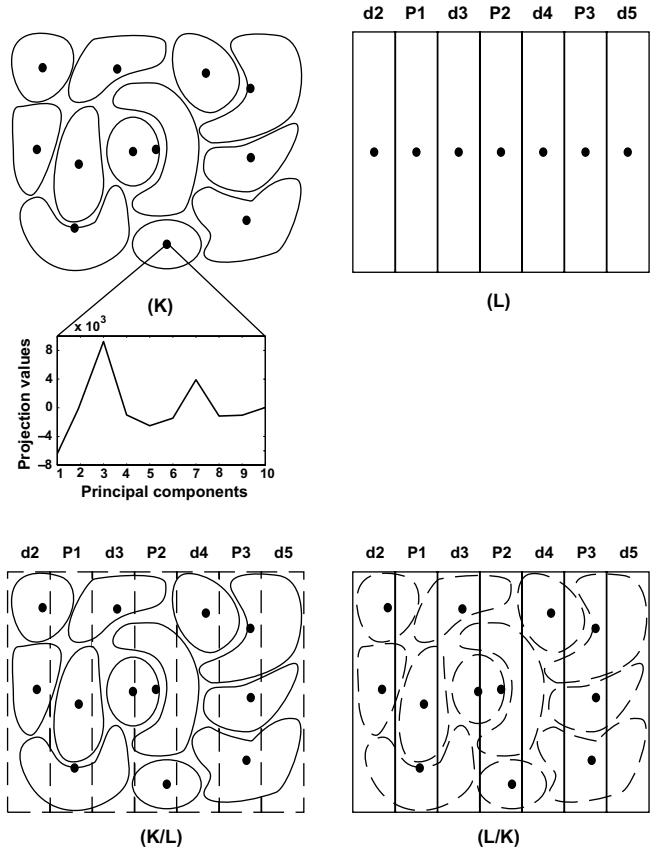

Fig. 4. Schematic view of the clustering of complex tactile stimuli. The set K corresponds to the kinetic feature extraction process while the set L corresponds to the stimulus localization process. To model stimulus feature extraction the stimuli set was clustered using K-Means to obtain 12 stable and homogeneous clusters of stimuli (K). Illustrated is the centroid plot of one stimuli cluster, which corresponds to the low-dimension stimulus average. The set L corresponds to the seven forepaw locations of stimulation. To describe the conditional stimuli set (K/L) we used the same procedure described before and then we partitioned each cluster according to the seven forepaw locations to obtain 84 new clusters. The other conditional stimuli set (L/K) was obtained by an inverse procedure in which we performed the clustering for each forepaw location, initializing the K-Means using the same 12 centers describing the variable K.

where $p_{\text{ind}}(\mathbf{r}) = \prod_{c=1}^{C}\prod_{t=1}^{N} p(r^c(t))$ follows an independent probability model as the product of $p(r^c(t))$, the marginal probability of responses of individual neuron in each time bin. While estimating $p(\mathbf{r})$ requires an evaluation of $2^{NC} - 1$ parameters, estimating $p(\mathbf{r})$ needs only NC parameters. This approximation has the advantage of being weakly biased according to the amount of trials we used in our paper (see Section 2.2).

3. Results

3.1. Clustering of complex tactile stimuli

Using the clustering procedure described in Section 2.3 we obtained $k = 12$ stable and homogenous clusters of tactile stimuli (set K). We partitioned the $n = 9100$ trials in each set K and L of stimuli. We chose the same number of trials (1300 trials in each of the seven clusters of L and 758 trials in each K cluster of stimuli) to obtain equiprobable clusters of stimuli, and thus $P(K = k) = 1/12$ and $P(L = l) = 1/7$. We computed the mutual information between the two stimuli variables $I(K;L) = 0.00480$ bits and test the "quasi-independence" of the variables K and L ($I_{\text{bootstrap}} = 0.00497$ bits; $P < 0.05$) using a bootstrap procedure (Shao and Dongsheng, 1995). In fact, the true independence corresponds to $I(K;L) = 0$.

3.2. Neuronal and information analyses

Estimation of the mutual information (MI) was based on eight sets of eight neurons recorded simultaneously. Typical responses of a representative set of eight neurons to stimulation on pad one are illustrated in the post-stimulus time histograms (PSTHs) shown in Fig. 5. The three most responsive neurons (neurons #1, #2 and #3) to pad one stimulation showed a sharp initial increase in firing rate 4–8 ms after stimulus onset, rising to a peak a few milliseconds later. There was a small increase in firing rate around the stimulus end. Neuron #3 was the most responsive to pad 1 stimulation, while neurons #4–#8 were almost unresponsive. In each set of recorded cells, the neuron #1 was located in the most rostral part of the S1 forepaw representation and generally its receptive field was located on either digit 2 or pad 1, while the neuron #8 was located toward the caudal part of the S1 map and responded mainly to the cutaneous stimulation of either digit 5 or pad 3 (see Section 2.2 and Fig. 2). The lower panel of Fig. 5 illustrates the "tuning curve" of each neuron that comprise the representative set, i.e. the average firing rate of each neuron according to its topographic location within the S1 forepaw map in response to each stimulation location.

Fig. 6 illustrates the different MI contributions between neuronal responses and both sets of stimuli processes (K and L), based on either spike timing or spike count. These MI estimations are shown for a representative set of eight neurons (Fig. 6, upper panels) and averaged on eight sets of eight neurons (Fig. 6, lower panels). It is worth noting that the MI curves correspond to the information quantity in bits divided by the average number of spikes in the growing post-stimulus time window (from 0 to T ms) (Arabzadeh et al., 2004). In addition, we cannot rule out that the MI estimates based on spike timing were slightly underestimated relative to the true MI quantities since we computed a lower bound approximation of the MI (see Section 2.4).

Using temporal coding, the MI between the neural response and the stimulus kinetic feature extraction, i.e. $I(R;K)$, first peaked at 4–8 ms post-stimulus and then gradually increased over time (Fig. 6, left panels). In contrast, the MI between the neural response and the stimulus location detection, i.e. $I(R;L)$, peaked at 4–8 ms after the stimulus onset and then gradually decreased. Both curves tend asymptotically toward the same value (0.25 bits/spike in spike timing). The behavior of both information curves was found to be similar with rate coding (Fig. 6, right panels); however, the deviation between $I(R;L)$ and $I(R;K)$ was

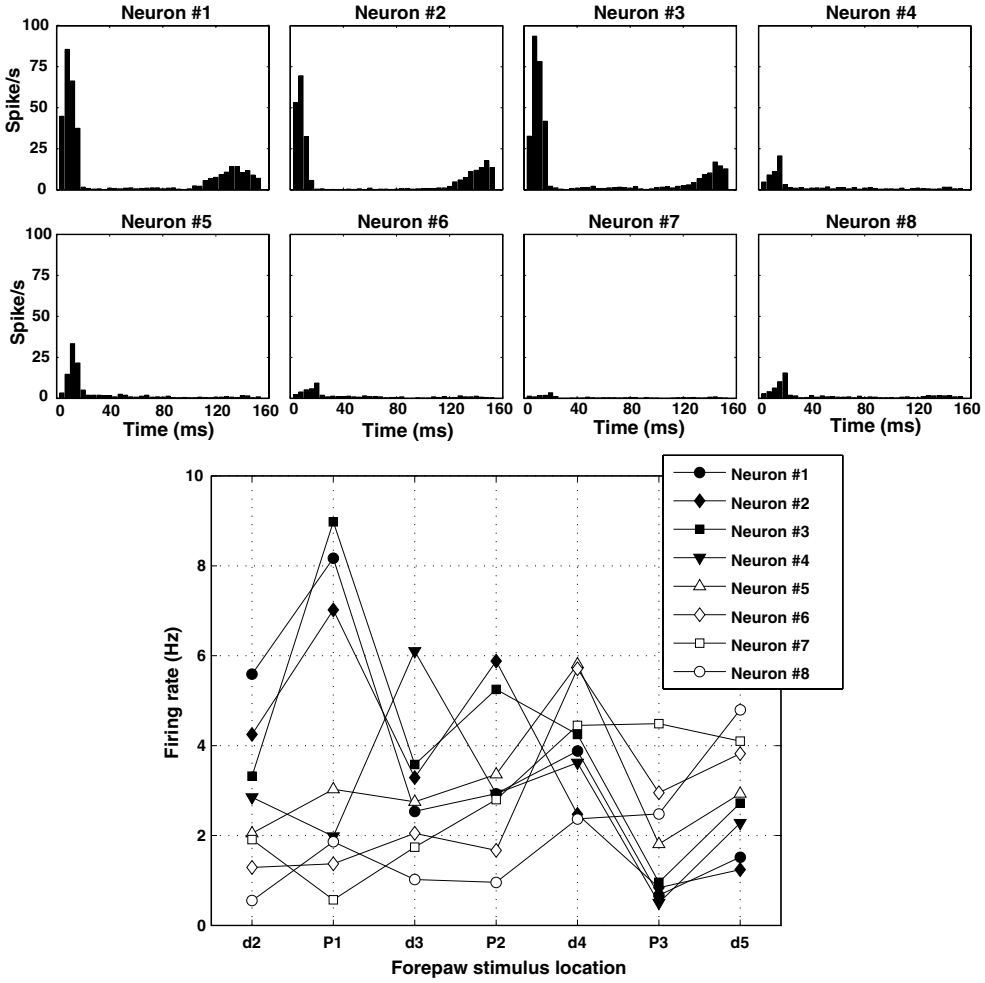

Fig. 5. Examples of post-stimulus time histograms (PSTHs) and of firing rates of a representative set of eight neurons. In the upper panels is illustrated the neuronal activity of each neuron recorded simultaneously in response to tactile stimulation of pad 1 (P1). The lower panel shows the average firing rate of each neuron to the stimulation of each of the 7 forepaw locations (from digit tip 2–5 and from pad 1–3).

greater in spike timing than in spike count. In fact, the difference between both MIs was 0.74 ± 0.09 bits/spike at 4–8 ms and 0.64 ± 0.03 bits/spike on average during the 20 first ms post-stimulus in spike timing. This difference was smaller at 4–8 ms (0.47 ± 0.08 bits/spike; t-test $= 2.15$; $P < 0.02$) and on average during the 20 first ms (0.38 ± 0.02 bits/spike; t-test $= 4.85$; $P < 0.003$) when using spike count rather than spike timing (see above). Therefore, more information about the stimulus localization rather than about the feature extraction was available to the neuronal population in the beginning of the stimulus presentation.

In addition, the amount of information $I(R;L)$ at 4–8 ms post-stimulus (0.84 ± 0.25 bits/spike) was not significantly larger (t-test $= 1.63$; $P < 0.1$) for the temporal coding than for the rate coding (0.64 ± 0.20 bits/spike) strategies. However, $I(R;L)$ in average during the 20 first ms post-stimulus was significantly larger (t-test $= 3.4$; $P < 0.01$) for the temporal coding than for the rate coding strategy (Fig. 6). In addition, the $I(R;K)$ peak at 4–8 ms was significantly larger (t-test $= 3.4$; $P < 0.01$) for the rate coding (0.16 ± 0.02 bits/spike) than for the temporal coding (0.09 ± 0.05 bits/spike) strategy (Fig. 6). Again, $I(R;K)$ on average during the 20 first ms post-stimulus was larger (t-test $= 5.5$; $P < 0.003$) with rate coding (0.14 ± 0.03 bits/spike) than with temporal coding (0.06 ± 0.03 bits/spike). Thus, the stimulus localization seems to be a spike timing-dependent process, while $I(R;K)$, the information relative to the stimulus feature extraction does not depend on spike timing in the beginning of the stimulus. The conditional MI curves $I(R;L|K)$ and $I(R;K|L)$ are indicated in the Fig. 6 and correspond to the prediction of the Eq. (3). $I(R;K,L)$ corresponds to the global stimulus identification process (i.e. both localization and feature extraction at the same time). The peak of $I(R;K,L)$ at 4–8 ms post-stimulus was larger (t-test $= 2.92$; $P < 0.03$) for the temporal coding (1.41 ± 0.22 bits/spike) than for the rate coding (1.06 ± 0.05 bits/spike) strategies.

From a theoretical point of view, the negativity of the multi-information (defined in Section 2.1.2, Eq. (13)) indicates that the variables R,K,L do not form a Markov chain, because the variables K and L are considered as

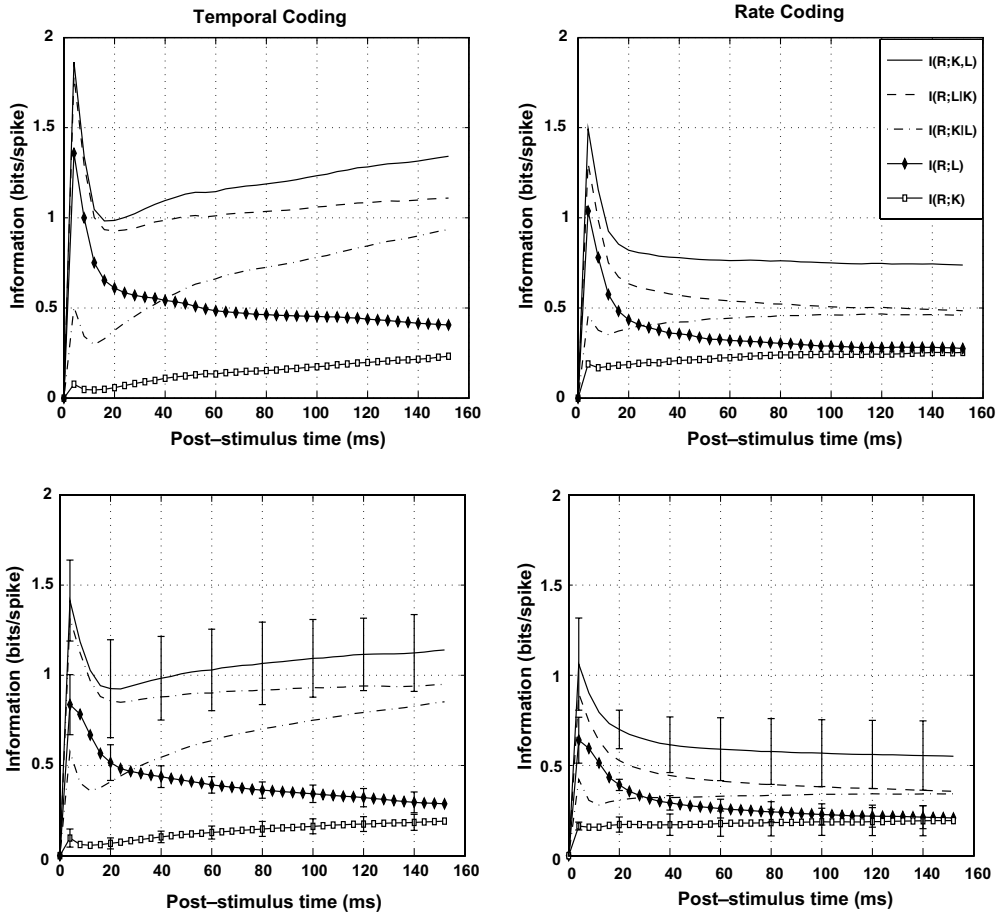

Fig. 6. Time course of the mutual information (MI) from the stimulus onset. The MIs are estimated for a representative set of eight neurons (upper panels) and for all sets of neurons (lower panels). Neural responses are based on either spike timing (left panels) or spike count (right panels). Neurons were recorded simultaneously in the S1 forepaw cortex. MIs curves shown in the lower panels are averaged over eight experiments and the error bars represent the SEM.

independent (see Section 3.1). More importantly, these variables (K and L) become dependent when we take the third variable R into account. From a neuronal point of view, we defined this "change of dependency" as the synergy between the identification processes. Fig. 7 shows the values of the multi-information for a representative set of eight neurons (left panel) and for the eight sets of eight neurons (right panel) when considering temporal and rate coding strategies. The dashed line ($y = 0$) represents the "synergy threshold". The larger the negativity of the multi-information values the stronger the synergy between the two processes. The multi-information quantity was negative during the stimulus presentation whatever the coding strategy used and peaked at 4–8 ms post-stimulus. The quantity of multi-information was lower (i.e. the synergy was greater) for the temporal coding (-0.48 ± 0.15 bits/spike at 4–8 ms and -0.29 ± 0.13 bits/spike on average during the 20 first ms post-stimulus) than for the rate coding (-0.26 ± 0.05 bits/spike at 4–8 ms; t-test $= 7.88$; $P < 0.0005$ and -0.13 ± 0.06 bits/spike on average during the 20 first ms post-stimulus; t-test $= 4.71$; $P < 0.003$). Interestingly, the multi-information curve displayed a plateau after 40 ms post-stimulus for the rate coding, whereas it kept decreasing (i.e. increasing the synergy) over time when using the temporal coding strategy.

4. Discussion

Our results show that the two stimulus identification processes are not performed in parallel by neuronal populations, so the stimulus localization and kinetic feature extraction are synergistic processes. From an information theory point of view, this synergy is described as a special case where two sets apparently independent (namely the stimulus location and the stimulus feature extraction) become anti-correlated given another set (the neural response). The multi-information quantity is more negative when using temporal coding rather than rate coding strategy. Thus, both identification processes can be considered as more synergistic if stimulus decoding by neural populations is based on spike timing. This synergy between neural processes allows us to consider the joint encoding of two different degrees of freedom of the complex stimulus (localization and kinetic features). Localization process has been found to be optimal in barrel cortex with spike timing (Panzeri et al., 2001). Our results show that the joint

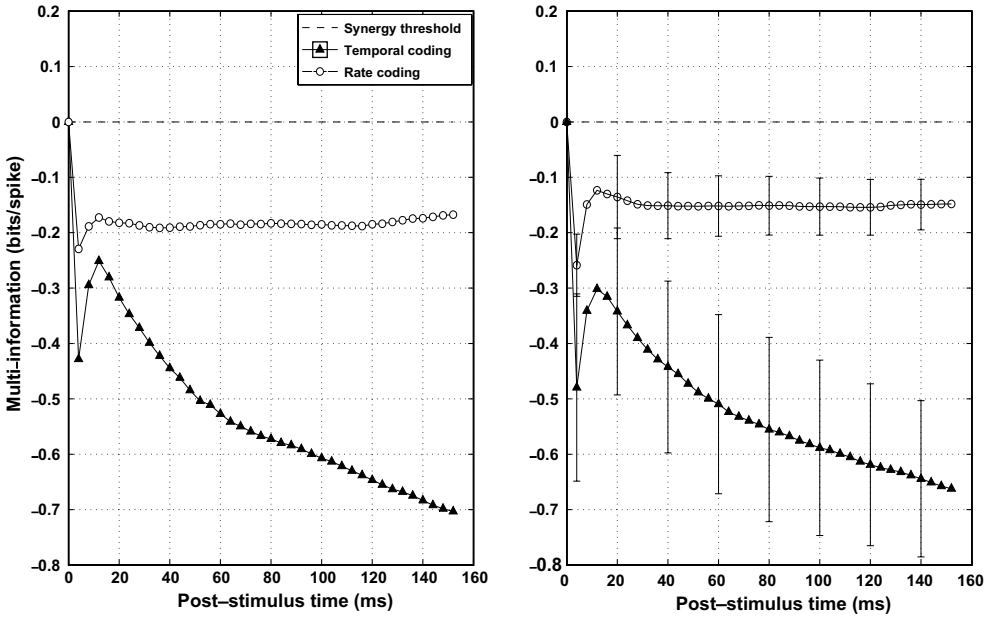

Fig. 7. Time course of the multi-information $I(R;K;L)$ from the stimulus onset using temporal and rate coding strategies. The multi-information quantities are estimated for a representative set of eight neurons (left panel) and for all sets of neurons (right panel). The synergy threshold is represented as the dashed line of zero value.

encoding of stimulus localization and feature extraction is optimal with spike timing in the rat forepaw cortex.

From a neural population point of view, we have studied how a neuronal population topographically distributed in the S1 forepaw cortex represents the combination of multidimensional stimulus features, and processes tactile information. One possible explanation is that a small number of neurons represents each combination of specific feature values. This is called a sparse code (Barlow, 1972) and it requires neurons that are highly selective in their responses. The sparse representation problem can be viewed as a generalization of the vector quantization objective described in Section 2.3 (Tropp, 2004). The advantage of this sparse code is that at any given point in time, only a small number of neurons is active. However, there is a serious disadvantage. In fact, to represent all possible combinations of many stimulus features, a very large number of neurons is necessary, growing exponentially with the number of dimensions, this is the "curse of dimensionality" (Bellman, 1961).

In our study, S1 neuronal populations perform two different processes to perceive objects and tactile stimuli. The kinetic feature extraction process consists in projecting each stimulus into a new subspace specific to the feature of interest, and then to classify the stimulus within this subspace. The second process, the stimulus localization, is related to the topographic organization of the neurons recorded in the rat S1 forepaw cortex. The interest of the present paper is that it demonstrates the synergy between two fundamental sensory identification processes. Information analyses show that the stimulus localization depends on spike timing and is faster than the feature extraction, which depends on spike count. In addition, all the MI quantities, except the MI related to the feature extraction, are greater in spike timing than in spike count. We can speculate that the localization process, which is faster, depends on feedforward connections from the periphery to the S1 cortex (Diamond et al., 2003), whereas the feature extraction, which requires a much longer time to reach the maximum information value, could depend on cortico-cortical connections, such as suggested in the visual cortex (Roelfsema, 2006). Both neural processes become more and more synergistic during the stimulus presentation in spike timing, but not in spike count. With spike timing, the joint encoding of these processes provides more information to the central nervous system for discriminability of complex stimuli.

Acknowledgements

The authors are grateful to Loic Bonnier for technical help in workstation management, Jean-Pierre Roman for design and technical assistance on the electronic Von-Frey, Laurent Pezard for helpful remarks and suggestions on the manuscript, and Stefano Panzeri for stimulating discussions on information bias.

References

Arabzadeh, E., Panzeri, S., Diamond, M.E., 2004. Whisker vibration information carried by rat barrel cortex neurons. Journal of Neuroscience 24 (26), 6011–6020.

Arabzadeh, E., Zorzin, E., Diamond, M.E., 2005. Neuronal encoding of texture in the whisker sensory pathway. PLoS Biology 3 (1), 155–165.

Arabzadeh, E., Panzeri, S., Diamond, M.E., 2006. Deciphering the spike train of a sensory neuron: counts and temporal patterns in the rat whisker pathway. Journal of Neuroscience 26 (36), 9216–9220.

Atick, J.J., 1992. Could information theory provide an ecological theory of sensory processing. Network 3, 213–251.

Attaneve, F., 1954. Informational aspects of visual perception. Psychological Review 61, 183–193.

Barlow, H., 1972. Single units and sensation: a neuron doctrine for perceptual psychology? Perception 1, 371–394.

Bellman, R., 1961. Adaptive Control Processes. Princeton University Press.

Bialek, W., Rieke, F., de Ruyter, van Steveninck, R., Warland, D., 1991. Reading the neural code. Science 252, 1854–1857.

Borst, A., Theunissen, F.E., 1999. Information theory and neural coding. Nature Neuroscience 2 (11), 947–950.

Bourgeon, S., Xerri, C., Coq, J.O., 2004. Abilities in tactile discrimination of textures in adult rats exposed to enriched or impoverished environments. Behavioral Brain Research 153 (1), 217–231.

Coq, J.O., Xerri, C., 1998. Environmental enrichment alters organizational features of the forepaw representation in the primary somatosensory cortex of adult rats. Experimental Brain Research 121 (2), 191–204.

Cover, T., Thomas, J., 1991. Elements of Information Theory. Wiley, New York.

Diamond, M.E., Petersen, R.S., Harris, J.A., Panzeri, S., 2003. Investigations into the organization of information in sensory cortex. Journal of Physiology (Paris) 97, 529–536.

Foffani, G., Tutucunler, B., Moxon, K.A., 2004. Role of spike timing in the forelimb somatosensory cortex of the rat. Journal of Neuroscience 24 (33), 7266–7271.

Gersho, A., Gray, R., 1991. Vector Quantization and Signal Compression. Kluwer Academic Publishers, USA.

Han, T.S., 1980. Multiple mutual informations and multiple interactions in frequency data. Information and Control 46 (1), 26–45.

Iwaniuk, N., Whishaw, I.Q., 2000. On the origin of skilled forelimb movements. Trends in Neuroscience 23 (8), 372–376.

Lebart, L., Morineau, M., Piron, M., 2002. Multidimensional exploratory statistics. Dunod.

Mumey, B., Sarkar, A., Gedeon, T., Dimitrov, A., Miller, J., 2004. Finding neural codes using random projections. Neurocomputing 25 (19), 58–60.

Nadal, J.-P., 2002. Information theoretic approach to neural coding and parameter estimation: a perspective. In: Probabilistic Models of the Brain: Perception and Neural Function. MIT Press.

Paninski, L., 2003. Estimation of entropy and mutual information. Neural Computation 15, 1191–1253.

Panzeri, S., Schultz, S., 2001. A unified approach to the study of temporal, correlational and rate coding. Neural Computation 13 (6), 1311–1349.

Panzeri, S., Treves, A., 1996. Analytical estimates of limited sampling biases in different information measures. Network 7 (2), 87–107.

Panzeri, S., Petersen, R.S., Schultz, S.R., Lebedev, M., Diamond, M.E., 2001. The role of spike timing of stimulus location in rat somatosensory cortex. Neuron 29 (3), 769–777.

Pola, G., Thiele, A., Hoffmann, K.P., Panzeri, S., 2003. An exact method to quantify the information transmitted by different mechanisms of correlational coding. Network 14 (1), 35–60.

Pola, G., Petersen, R., Thiele, A., Young, M., Panzeri, S., 2005. Data-robust tight lower bounds to the information carried by spike times of a neuronal population. Neural Computation 17 (9), 1962–2005.

Roelfsema, P.R., 2006. Cortical algorithms for perceptual grouping. Annual Review of Neuroscience 29, 203–227.

Shannon, C., 1948. A mathematical theory of communication. Bell System Technical Journal (27), 379–423.

Shao, J., Dongsheng, T., 1995. The Jacknife and Bootstrap. Springer Series in Statistics. Springer-Verlag.

Slonim, N., Friedman, N., Tishby, N., 2006. Multivariate information bottleneck. Neural Computation 8 (18), 1739–1789.

Tropp, J., 2004. Greed is good: algorithmic result for sparse approximation. IEEE Transactions on Information Theory 10 (50), 2231–2242.

Tutunculer, B., Foffani, G., Himes, B.T., Moxon, K.A., 2006. Structure of the excitatory receptive fields of infragranular forelimb neurons in the rat primary somatosensory cortex responding to touch. Cerebral Cortex 16 (6), 791–810.

Yeung, R., 1997. A framework for linear information inequalities. IEEE Transactions on Information Theory 43 (6), 1924–1934.