Available online at www.sciencedirect.com
ScienceDirect
Neural Networks 20 (2007) 194–209
Neural Networks
www.elsevier.com/locate/neunet

# A machine learning approach to the analysis of time–frequency maps, and its application to neural dynamics

François B. Vialatte[a,b,1], Claire Martin[b], Rémi Dubois[a], Joëlle Haddad[a], Brigitte Quenet[a], Rémi Gervais[b], Gérard Dreyfus[a,*]

[a] *École Supérieure de Physique et de Chimie Industrielles de la Ville de Paris (ESPCI-Paristech), Laboratoire d'Électronique (UMR 7084 CNRS), 10 rue Vauquelin, 75005 Paris, France*

[b] *Institut des Sciences Cognitives (ISC), UMR 5015 CNRS Université Claude Bernard Lyon I 69675 Bron and IFR 19 Neurosciences, Lyon, France*

Received 5 April 2005; accepted 7 September 2006

**Abstract**

The statistical analysis of experimentally recorded brain activity patterns may require comparisons between large sets of complex signals in order to find meaningful similarities and differences between signals with large variability. High-level representations such as time–frequency maps convey a wealth of useful information, but they involve a large number of parameters that make statistical investigations of many signals difficult at present. In this paper, we describe a method that performs drastic reduction in the complexity of time–frequency representations through a modelling of the maps by elementary functions. The method is validated on artificial signals and subsequently applied to electrophysiological brain signals (local field potential) recorded from the olfactory bulb of rats while they are trained to recognize odours. From hundreds of experimental recordings, reproducible time–frequency events are detected, and relevant features are extracted, which allow further information processing, such as automatic classification.


*Keywords:* Machine learning; Time–frequency; Wavelet; Neural network; Local field potential; Bump; Olfaction; Rat; Odour recognition; Electro-encephalography

## 1. Introduction and motivation

The purpose of this paper is to provide a new method for comparing automatically complex signals, related to the investigation of brain activity patterns, which are usually analyzed 'visually' in the time–frequency domain. The motivation is the extraction of differences and similarities between experimental signals with large variability. In such situations, the main problem is that of finding an appropriate data representation, which must be parsimonious enough for processing with present-day algorithms and computers, yet retain the information that is relevant for fast and reliable comparison and classification.

The automatic analysis of time-varying signals has been spurred to a large extent by industrial applications such as automatic speech recognition. In that context, numerous data representations (see for instance Bogert, Healey, and Tukey (1963)) and classification techniques (such as Hidden Markov Models, see for instance Rabiner (1989)) have been designed. However, those representations and methods are not especially useful for discovering structure in time–frequency patterns, which are ubiquitous in the investigation of brain dynamics; they are of great interest to neurobiologists, who have been actively investigating the possible functional significance, for cognition, of specific types of dynamics of neural assemblies. Those activities can be recorded either from the surface of the brain (electroencephalogram, hereinafter abbreviated to EEG) or from deep implanted electrodes (local field potentials, hereinafter abbreviated to LFPs). Typical challenging problems can be found in the Brain–Computer Interface (BCI) competition (Blankertz et al., 2004), in which electrophysiological signals must be classified using statistical methods that range from elementary, such as nearest-neighbour classifiers (Cincotti et al., 2002) and linear separators, to elaborate ones (neural networks, support vector machines).

* Corresponding author. Tel.: +33 1 40 79 45 41; fax: +33 1 47 07 13 93.
*E-mail address:* Gerard.Dreyfus@espci.fr (G. Dreyfus).
[1] Present address: Riken Brain Science Institute, Laboratory for Advanced Brain Signal Processing, 2-1, Hirosawa, Wako Saitama, 351-0198, Japan.

0893-6080/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2006.09.013

**Symbols**

| | |
|---|---|
| $a$ | Amplitude of a bump function |
| $b_r$ | Centroid of group $G_r$ |
| $B$ | Number of bumps |
| $c_{ft}$ | Wavelet time–frequency coefficient at frequency $f$ and time $t$ |
| $\delta_f$ | Estimated standard deviation of $c_{ft}$ for each time $t$ |
| $\delta_x$ | Normalized distance in time between two bumps |
| $\delta_t$ | Distance in time between two bumps |
| $\delta_y$ | Normalized distance in frequency between two bumps |
| $\delta_f$ | Distance in frequency between two bumps |
| $D_r$ | Distance of a bump to its neighbours |
| $d$ | Distance between two bumps |
| $F$ | Ratio of intensity modelled by a bump to the total intensity of the map |
| $\varphi_b(f,t)$ | Bump function |
| $G_r$ | Group of bumps |
| $\{G_z\}$ | Set of groups |
| $G_z$ | The most compact group relatively to $D_r$ |
| $H$ | Frequency extension of a time–frequency window |
| $l_f$ | Half-length of the frequency axis of a half ellipsoid function |
| $l_t$ | Half-length of the time axis of a half ellipsoid function |
| $\lambda$ | Losses during normalization |
| $L$ | Time extension of a time–frequency window |
| $\mu_f$ | Frequency centre of a bump function |
| $\mu_t$ | Time centre of a bump function |
| $N$ | Number of time–frequency maps |
| $\Omega$ | Time–frequency map |
| $P$ | Number of periods defining an oscillation |
| $\rho$ | Modelling residual |
| $R$ | Invariance rate of a group of bumps |
| $\sigma_t$ | Standard deviation of the Gaussian envelope of a Morlet wavelet |
| $\sigma_f$ | Standard deviation of the Gaussian frequency spectrum of a Morlet wavelet |
| $s$ | Scale step |
| $S$ | Sum of intensities of the pixels of a window |
| $s_f$ | Standard deviation along the vertical (frequency) axis of a Gaussian function |
| $s_t$ | Standard deviation along the horizontal (time) axis of a Gaussian function |
| $\tau$ | Time step |
| $\Theta$ | Threshold of neighbourhood distance |
| $w$ | Wavelet function |
| $W$ | Time–frequency window |
| $W_{\max}$ | Time–frequency window containing the maximal amount of oscillatory activity |
| $z_{ft}$ | Normalized time–frequency coefficient at frequency $f$ and time $t$ |

In EEG data analysis, source separation is performed by Independent Component Analysis (Makeig, Bell, Jung, & Sejnowski, 1996), and distances between signals are computed separately from the frequency components and from the time components (Anemuller, Sejnowski, & Makeig, 2003). However, to the best of our knowledge, no available method allows a full description of complex signals in both time and frequency domains in conjunction with pattern extraction, leading to statistical comparison between sets of signals. In neurobiology, this might be particularly useful for identifying specific dynamics of a given cognitive task, as will be exemplified below.

The automatic analysis of large sets of time–frequency maps involves difficult problems related (i) to the large number of 'pixels' in the maps, and (ii) to the inter-map variability due to time jitter and other sources of noise. This makes automatic map comparisons very complex and time consuming, even with powerful computers. Moreover, it is desirable to use, as a starting point, the time–frequency representations that neurobiologists are familiar with, in order to take advantage of the expertise gained by the visual analysis of time–frequency maps. Image segmentation techniques, using the EM algorithm for instance (Kapur, Grimson, Wells, & Kikinis, 1996) are not readily applicable to such analyses. We describe a method that models a time–frequency map as a sum of appropriate elementary functions; that method is a 2-dimensional generalization of the Gaussian mesa function modelling technique described for one-dimensional signals (electrocardiogram analysis) in Dubois (2004) and Dubois, Quenet, Faisandier, and Dreyfus (2006). The method will be applied in this paper to the analysis of electrophysiological signals, but it could be used as well for any other problem that requires a simple description of a time–frequency map.

Brain rhythms are traditionally described by different oscillatory regimes including theta (4–10 Hz), beta (15–40 Hz) and gamma (40–90 Hz) rhythms. The functional importance of oscillatory activities for brain computation is still a central open problem. Experimentally, their possible role in perception can be assessed by finding specific and reproducible patterns of oscillatory activities in response to a large set of stimuli having the same behavioural significance. Interestingly, at least some specific stimulus-induced oscillatory regimes have been observed in mammals mainly in response to olfactory (Freeman & Schneider, 1982) and visual stimuli in the cat and monkey (Singer, 1993), and in human (Tallon-Baudry, Bertrand, Delpuech, & Pernier, 1997; Tallon-Baudry, Bertrand, Peronnet, & Pernier, 1998). In those experiments, induced oscillatory activity has been found mainly in the gamma band (around 40 Hz), and interpreted as a neural correlate of the transient formation of neural assemblies related to information processing. Responses were investigated using Fast Fourier transforms, and, more recently, wavelet transforms. The latter technique is particularly powerful for studying transient phenomena without any prior knowledge of frequency bands of interest.

Even though repeated presentation of the same stimulus generates similar neural responses, it is well known that the exact time course of the response exhibits jitter, i.e. varies

from one trial to another. Although averaging is efficient in showing events that are perfectly time-locked to the stimulus onset (evoked-activity), it blurs out events (induced-activity) that occur at variable intervals following stimulus onset (Tallon-Baudry et al., 1997, 1998). That difficulty can be circumvented by selecting a time–frequency window, and performing an ANOVA test on the amplitude of the maximal activity (Gruber & Müller, 2005), or mean activity (Herrmann, Mecklinger, & Pfeifer, 1999) within the window; however, that requires assumptions that are not necessarily true, and, in addition, a fairly accurate prior knowledge of the time–frequency area of interest (see Section 3.2.2). In addition, the measurement of the maximal activity does not take into account peaks of lower amplitude, which can also convey meaningful information.

The purpose of the present paper is thus to describe a new method for signal analysis and modelling, which addresses efficiently the problem of detecting automatically reproducible time–frequency events in spite of inter-trial variability. The method relies on the analysis of Morlet wavelets coupled with modelling the time–frequency maps by specific functions called "bumps", which provide very parsimonious representations. Furthermore, it provides a quantitative index of the reproducibility of events, termed 'invariance rate'.

The method is designed to be general-purpose. It was set up in the specific context of the analysis of LFP recordings of the olfactory bulb of rats that were trained to solve a Go/No Go task based on olfactory cues (see Section 3.3); however, it has also been applied successfully to the analysis of electro-encephalographic data for the early detection of Alzheimer's disease (Vialatte, Cichocki, Dreyfus, Musha, & Gervais, 2005a, 2005b).

## 2. Presentation of the method

The bump[2] modelling of a time–frequency map aims at representing the map with a limited number of elementary functions. The purpose is to reduce the huge quantity of parameters (tens to hundreds of thousands) that describe a time–frequency map to a sum of parametric functions (a few functions with some tens of parameters). A parsimonious representation is then obtained, which is relevant for further analysis. The method is somewhat similar in spirit to the matching-pursuit type algorithms (Mallat & Zhang, 1993), applied to the modelling of the time–frequency maps instead of being applied directly to the signal; as mentioned above, the rationale for modelling the map is the fact that experts of the analysis of oscillatory signals in LFP's are familiar with the time–frequency maps.

In the present section, the time–frequency map of interest is first defined, and the preprocessing operations applied to the signal are described; the method of bump modelling is subsequently described. Finally, the resulting map representation is exploited for finding events that are invariant across experiments. A procedure for assessing quantitatively the invariance of bumps and their mutual distances will be described.

[2] In the present paper, two-dimensional local functions are loosely referred to as 'bumps', although those functions do not comply with the mathematical definition of a bump function, i.e. an infinitely differentiable function that is constant outside a domain. Typically, the function $f(x) = \exp\left(-\frac{1}{(1-x^2)}\right)$ if $-1 < x < 1$, $f(x) = 0$ otherwise, is a *bona fide* bump function.

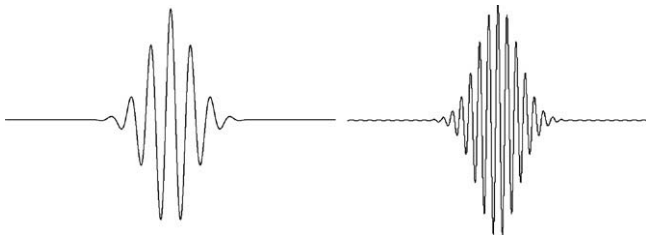

Fig. 1. Real part of $w(t)$ (relation (1)), with $2\pi\sigma_t f = 7$ (left) and $2\pi\sigma_t f = 14$ (right).

### 2.1. Time–frequency maps

The continuous wavelet transform (Mallat, 1989; Poularikas, 1996) of a function $g(\tau)$ is defined as:

$$W_g(s,t) = \int_{-\infty}^{+\infty} g(\tau) w^*\left(\frac{\tau - t}{s}\right) \mathrm{d}\tau$$

where $w^*(\tau)$ is the complex conjugate of the "mother wavelet" function $w(\tau)$, and $t$ and $s$ are the translation and scale factors respectively. In practice, the signal $g(\tau)$ is sampled at a known frequency. The scale factor acts on the central frequency $f$ of the Fourier spectrum of the wavelet: $f$ will be loosely referred to as the 'frequency' of the wavelet.

A time–frequency map $\Omega$ is a convenient 3-dimensional representation of the result of the wavelet transformation of a sampled signal. The $x$-axis is $\tau$, which takes on discrete values that are multiples of the sampling period; therefore, the horizontal axis is essentially time. The $y$-axis is the frequency axis: the discrete values of the scales are chosen in such a way that the frequency step, or log-frequency step, along that axis, has a pre-defined value. The magnitude $c_{ft}$ of the wavelet transform of the signal for a given value of the frequency $f$ and of the translation factor $t$ is plotted along the $z$-axis.

There is a wide variety of wavelets. In the present study, complex Morlet wavelets were investigated. Complex Morlet wavelets are appropriate for time–frequency analysis of electroencephalographic signals (Caplan, Madsen, Raghavachari, & Kahana, 2001; Düzel et al., 2003; Tallon-Baudry, Bertrand, Delpuech, & Pernier, 1996). Complex Morlet wavelets (Kronland-Martinet, Morlet, & Grossmann, 1987) of Gaussian shape in time (standard deviation $\sigma_t$) are defined as:

$$w(t) = A.\exp(-t^2/2\sigma_t^2).\exp(2\mathrm{i}\pi f t) \quad (1)$$

where $\sigma_t$ and $f$ are appropriately chosen parameters; they cannot be chosen independently, since the product $\sigma_t f$ determines the number of periods that are present in the wavelet. In the present investigation, the wavelet family defined by $2\pi\sigma_t f = 7$ was used (see Fig. 1) as described in (Tallon-Baudry et al., 1996).

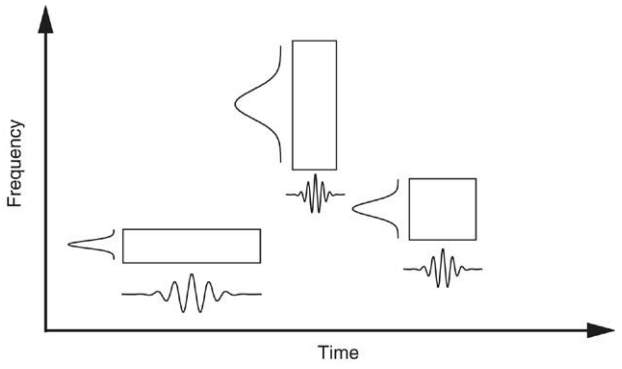


Fig. 2. The time and frequency extensions of wavelets in a time–frequency map (adapted from Mallat (1989)).

The Fourier spectrum of a Morlet wavelet is Gaussian with standard deviation $\sigma_f$, with:

$$\sigma_f = \frac{1}{2\pi\sigma_t}. \tag{2}$$

Thus, in the present investigation:

$$\sigma_f = \frac{1}{2\pi\sigma_t} = \frac{f}{7}. \tag{3}$$

Therefore, the higher the central frequency of the wavelet, the wider its frequency spectrum (the poorer its frequency resolution), as shown on Fig. 2.

In the experiments described in the present paper, the time step is one sampling period (0.5 ms), and the frequency step is 1 Hz. A time–frequency map of a synthetic signal is shown on Fig. 3. Its information content is quite large: in the above conditions, 1 s of signal gives rise to $86 \times 2000 = 172{,}000$ values of $c_{ft}$. Comparing directly hundreds of such maps is thus essentially impossible.

### 2.2. Preprocessing of time–frequency maps

#### 2.2.1. Undersampling of the time (translation) axis of the map

LFP signals were sampled at 2 KHz, which provides an unnecessarily high resolution for our analysis between 15 and 100 Hz. Time–frequency maps were undersampled along the time axis by a factor of 10. That reduces the number of parameters to 17,200 coefficients for 1 s of signal.

#### 2.2.2. Normalization

The second preprocessing step is the normalization of the coefficients of the time–frequency map (also termed '$z$-score transformation'): this is mandatory because time–frequency maps of LFPs have much higher energy for low frequencies than for high frequencies. Since the bump modelling algorithm detects events of highest amplitudes first (as will be shown in Section 2.3.3), non-normalised signals would favour low frequencies, which are not necessarily more relevant to brain computation than higher frequencies.

Frequencies within the map are thus balanced by defining a normalised and centred variable $z_{ft}$ as:

$$z_{ft} = \frac{c_{ft} - m_f}{\delta_f} \tag{4}$$

where $m_f$ and $\delta_f$ are the estimated mean and standard deviation of $c_{ft}$ for frequency $f$ during a time window where the animal behaves freely. $z_{ft}$ may be negative, for coefficients smaller than $m_f$. That normalization strategy is sometimes used for noise elimination in EEG time–frequency maps: as an example, in Kahana, Sekuler, Caplan, Kirschen, and Madsen (1999) coefficients $c_{ft}$ whose normalized values $z_{ft}$ are smaller than 1 are deleted. In our case, it is desirable to keep as much information as possible before the detection of the invariant parts of the maps, as described in Section 2.4. In order to have only positive normalized values $z_{ft}$, the latter must be shifted in the positive direction. By adding 2 units to $z_{ft}$ (i.e. shifting

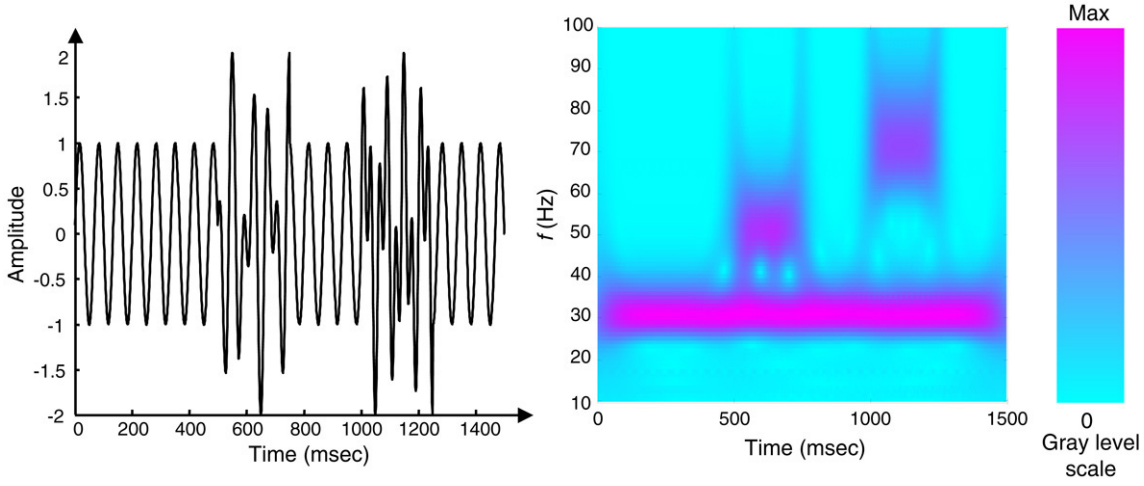


Fig. 3. An example of wavelet transform. Left, an artificial signal made of three components: A low frequency (30 Hz) oscillation, and two oscillatory bursts around (50 Hz; 500–700 ms) and (70 Hz; 1000–1200 ms). Right, the time–frequency map of this signal obtained with the complex Morlet wavelet. A color version of the figure is posted at: http://www.neurones.espci.fr/NNT_1005_color_figures/Figures.html.

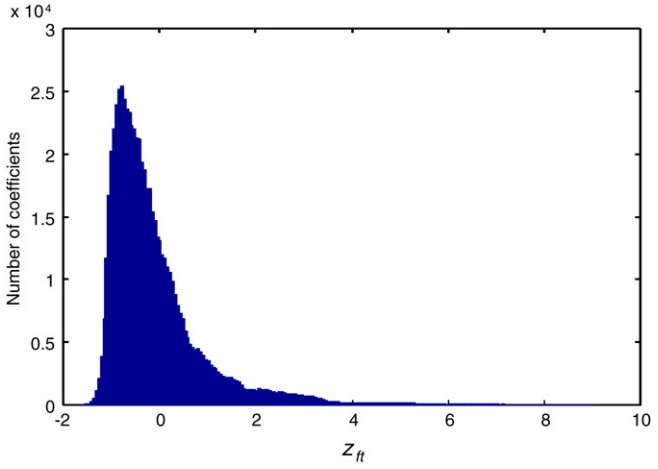


Fig. 4. Distribution of the normalized wavelet coefficients $z_{ft}$ (relation (4)) for a typical map, before shifting them in the positive direction: all values of $z_{ft}$ are larger than $-2$.

$z_{ft}$ up by two standard deviations) and setting negative $z_{ft}$'s to zero, losses λ (as defined by relation (5)) are negligible (on the order of $10^{-5}$).

$$\lambda = \sum_{f=f_{\min}}^{f_{\max}} \sum_{t=t_{\min}}^{t_{\max}} \frac{(z_{ft}+2) - \mathrm{abs}(z_{ft}+2)}{2}. \quad (5)$$

Fig. 4 shows a typical histogram of $z_{ft}$: since no value of $z_{ft}$ is smaller than $-2$, the distribution can be trimmed below $-2$ without any information loss.

Note that this normalization procedure cannot be used for the detection of event-related desynchronization; in the present investigation, the time localization of desynchronized events was not of interest. If such is not the case, no offset is applied, and the positive and negative parts of the resulting map are modelled separately.

#### 2.2.3. Dealing with noise

Various sources of noise or disturbances may be present:

- noise arising from the signal recording technique,
- artefacts arising from the time–frequency transformation.

The first kind of noise is seldom present in LFP signals investigated in this paper, and can be eliminated using an amplitude threshold or through visual check.

Artefacts arising from the transform must be taken into account. Essentially, wavelet transforms exhibit boundary effects at the ends of the signal; therefore, the part of the map that is analysed must be enclosed within a boundary zone where artefacts may appear, as described in Section 2.3.1.

### 2.3. Bump modelling algorithm

In the present section, the generation of a parsimonious representation of the map by bump modelling is described. A typical result of that procedure is shown on Fig. 5, displaying (top) a time–frequency map of a signal recorded in an 'expert' rat, and (bottom) the result of the bump modelling of the original map: that model is described by 35 parameters (5 parameters per bump), while the original map is described by 9000 values of the wavelet coefficients $c_{ft}$ (90 frequency steps, 100 time steps).

The purpose of the algorithm is to approximate a time–frequency map as a sum of known elementary parameterized functions $\varphi_b$ called bumps. The parameters of the bumps are adapted in order to minimize the modelling error. For each signal, the algorithm features the following steps:

- (i) Define the boundaries of the map in order to avoid finite-size effects.
- (ii) Window the map in order to define the zones to be modelled.
- (iii) Find the zone that contains the maximum amount of energy.
- (iv) Adapt a bump to the selected zone, and withdraw it from the original map.
- (v) Stop if the amount of information modelled by the bumps reaches a threshold, else iterate to (iii).
- (vi) Prune the model for optimal information contents.

#### 2.3.1. Boundaries

Due to the abrupt changes occurring at the beginning and at the end of the recordings, spurious frequencies are generated by the wavelet transform. Therefore, the section of recorded signal to be transformed started 750 ms before the part of the signal that was of interest and stopped 750 ms after the end of the signal of interest. After wavelet transformation, the two 750 ms parts of the signal that were likely to exhibit artefacts were removed from the map.

#### 2.3.2. Resolution and windowing

As mentioned in Section 2.1, the time extension of wavelets is frequency-dependent: for high frequencies, wavelets have a small time extension (high time resolution), but their frequency spectrum is large (low frequency resolution), whereas the inverse occurs at low frequencies. Time–frequency patterns to be modelled are thus characterized by:

- a frequency-dependent time resolution as defined by relation (2): $\sigma_t = \frac{7}{2\pi f}$,
- a time duration, corresponding to the duration of a biologically relevant oscillation of the local field potential (LFP): such an oscillation is defined as a periodic LFP signal that lasts at least $P$ time periods, with $3 \leq P \leq 4$. In this paper, $P = 4$ time periods will be used. In all the following, the term 'oscillation' will be used to refer to a biologically relevant oscillation.

The height and width of the bumps, expressed in terms of numbers of 'pixels' of the map (i.e. numbers of coefficients of the wavelet expansion of the signal, after preprocessing as described in Section 2.2), depend on $\sigma_t$ and $P$. Since time–frequency patterns organized in oscillations lasting 3–4 periods are searched for, sliding windows defined as follows will be used:

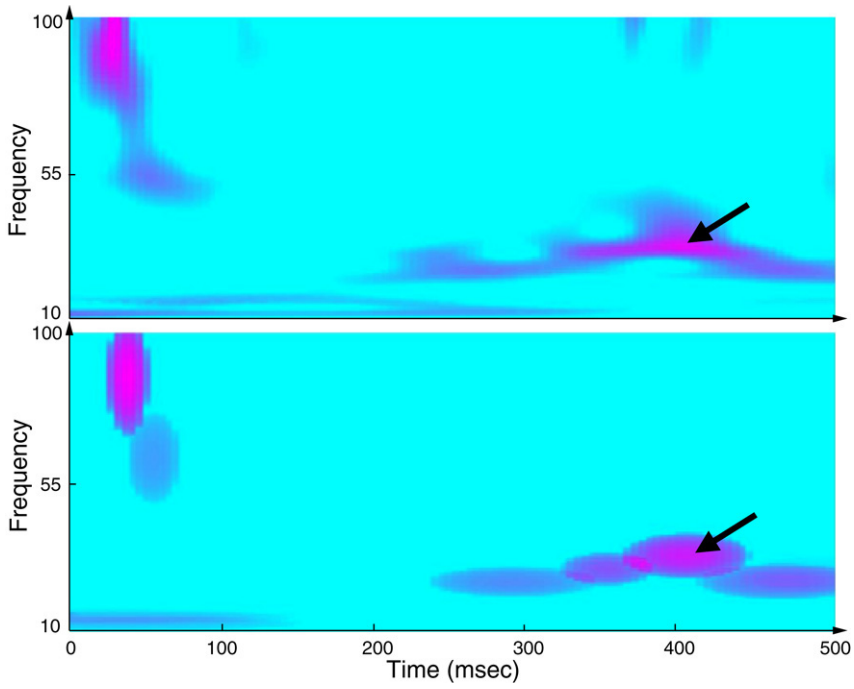


Fig. 5. Example of bump modelling. Top: Normalized time–frequency map of a LFP signal; bottom: Bump modelling (using 'half ellipsoid' functions, defined in Section 3.1) of the map: The arrow points at the bump that has the largest amplitude. A colour version of the figure is posted at: http://www.neurones.espci.fr/NNT_1005_color_figures/Figures.html.

- The time extension $L$ of a window centred at frequency $f$ is equal to the time duration of an oscillation:

$$L = \frac{P}{f}. \tag{6}$$

  Therefore, the ratio of the time extension of the window to the time resolution of a wavelet at that frequency is $\frac{P/f}{\sigma_t} = \frac{8\pi}{7}$;
- The frequency extension $H$ of that window is such that its ratio to the frequency resolution of the wavelet is also equal to $8\pi/7$ : $\frac{H}{\sigma_f} = \frac{8\pi}{7}$, hence

$$H = \frac{2\pi}{49} Pf. \tag{7}$$

Those values are computed for each point of the time–frequency map (within the boundaries defined in Section 2.3.1). Thus, for each point of the map, a time–frequency window centred at that point is obtained (Fig. 6).

Since windows that are centred on points that lie at a boundary of the map extend beyond the map, provision is made for time and frequency margins, as shown on Fig. 6.

### 2.3.3. Search for the zone containing the maximal amount of oscillatory activity

The normalized intensities of the pixels contained in a window describe the amount of oscillatory activity within the window. The modelling algorithm searches for the window $W_{\max}$ containing the maximal amount of that activity: for each window $W$, the sum $S$ of the intensities of the 'pixels' $S = \sum_{t,f\in W} z_{ft}$ that it contains is computed; the summations

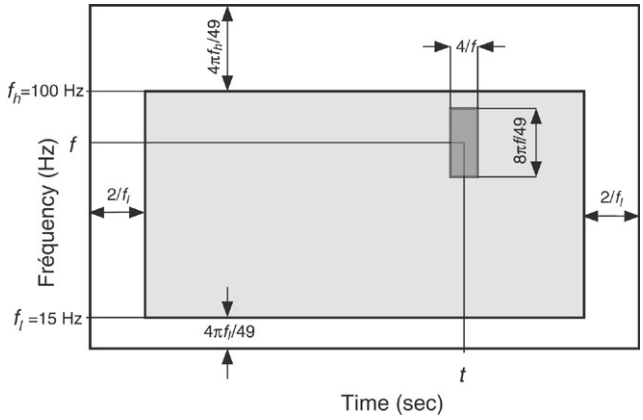


Fig. 6. Definition of the sliding windows, and of time and frequency margins, for $P = 4$. The grey area is the map under investigation. $f_l$ and $f_h$ are the lowest and highest frequencies of interest. The dark grey rectangle is a sliding window.

run on all points within window $W$. The window $W_{\max}$ with maximal $S$ is selected.

### 2.3.4. Bump adaptation

Within the selected window, a bump function $\varphi_b(f,t)$ is adapted, starting with a bump extending over the whole area of the window. The definition of the bump functions will be given in Section 3.1. As an illustration, consider a 'half-ellipsoid' bump function (Fig. 7), defined by

$$\varphi_b(f,t) = a\sqrt{1-v} \quad \text{for } 0 \le v \le 1$$
$$\varphi_b(f,t) = 0 \quad \text{for } v > 1$$

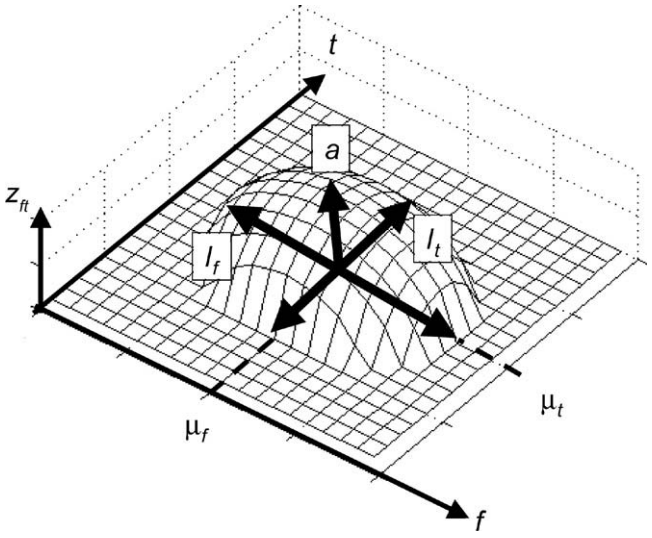


Fig. 7. Half ellipsoid bump, with its five parameters $(a, l_f, l_t, \mu_f, \mu_t)$.

where $v = (e_f^2 + e_t^2)$ with $e_f = (f - \mu_f)/l_f$ and $e_t = (t - \mu_t)/l_t$. $\mu_f$ and $\mu_t$ are the coordinates of the centre of the ellipsoid, $l_f$ and $l_t$ are the half-lengths of the principal axes along the frequency and time axes respectively, and $a$ is the amplitude of the function. Thus, the bump function has five parameters, subject to the following constraints:

- $\mu_f > 0$, $\mu_t > 0$ and such that the centre of the bump lies within the window,
- $0 < l_t < L$, $0 < l_f < H$, where $L$ and $H$ are the height and width of the window, as defined by relations (6) and (7).
- $a > 0$.

During the adaptation phase, the parameters of the function are optimized under the above constraints with a second order gradient descent using the Broyden–Fletcher–Goldfarb–Shanno (BFGS) method (see for instance Press, Flannery, Teukolsky, and Vetterling (2002)). The cost function to be optimized is the modelling error of the bump, defined by the usual sum of squared errors:

$$C = \frac{1}{2} \sum_{t,f \in W} (z_{ft} - \varphi_b(f,t))^2,$$

where the summation runs on all pixels within the window $W$ under consideration.

For instance, if $\varphi_b$ is a half ellipsoid function, $C$ becomes:

$$C = \frac{1}{2} \sum_{f,t \in W} (z_{ft} - \varphi_b(f,t; a, l_f, l_t, \mu_f, \mu_t))^2,$$

where the summation runs on all parameters within the window $W$ under consideration, and $a, l_f, l_t, \mu_f, \mu_t$ are the parameters of the function.

Thus, the time–frequency patterns modelled by the bump are restricted to one 'biologically relevant oscillation' with a duration of four periods or less. Longer oscillations will be modelled by two bumps or more, for reasons that will be explained below. The amplitude $a$ is constrained to have positive values. In the following, the value of the cost function of a bump after adaptation is termed the *modelling error* of the bump.

If, after adaptation, a bump happens to extend outside its window (e.g. because the centre of the adapted bump is away from the centre of the window), the window itself is shifted in order to follow the bump (within the limits of the map boundaries). When the bump is finally adapted, it is subtracted from the time–frequency map, and the process is iterated with the next bump.

#### 2.3.5. Termination criterion and model pruning

As usual in signal modelling, a tradeoff must be performed between accuracy and relevance (also termed 'bias-variance dilemma'): if the number of bumps in the model is too low, the latter will not be accurate; if it is too large, the noise will be modelled, hence irrelevant information will be built into the model. Two approaches may be taken: (i) design a model with the largest number of bumps compatible with a reasonable computation time, and then prune the model; or (ii) stop the modelling procedure when additional bumps contain an amount of information that is deemed negligible. In the present investigation, the former strategy was used. To that effect, the fraction of the total intensity of the map modelled by a given bump is computed:

$$F = \sum_t \sum_f \varphi_B(f,t) \Big/ \sum_t \sum_f z_{ft},$$

where the summations run on all points of the map.

In the examples shown in Section 3, the stopping criterion was the following: modelling was terminated after computing three bumps with $F < 5 \times 10^{-3}$; that led to a typical number of biologically relevant oscillations per map on the order of 40. The pruning strategy that is used after the termination criterion defines an acceptance threshold $F^*$ for $F$: all bumps that model a fraction $F$ of the map that is smaller than $F^*$ are discarded; the definition of the threshold involves the *invariance rate* of the groups of bumps present in the model, as defined in Section 2.4. The pruning procedure is described in Appendix B.

### 2.4. Analysis of the invariance of the bumps

As mentioned above, one of the purposes of the present investigation is the automatic extraction of spatio-temporal patterns that are common to different time–frequency maps, e.g. maps derived from LFP recordings of expert rats as compared to beginner rats. Therefore, after modelling the time–frequency maps of interest, it is important to find 'invariant' bumps, i.e. bumps that are consistently found in similar maps (irrespective of their amplitudes). To that effect, bumps are clustered into groups of radius $\Theta$ in time–frequency space. The *invariance rate* of a group is defined as the proportion of modelled signals that have a bump within that group (or, equivalently, the probability of occurrence of a bump, in the time–frequency maps under investigation, within a distance $\Theta$ of the centroid of the group). Therefore, the choice of the parameter $\Theta$ depends on the expected amount

of jitter in the maps: a large value of $\Theta$ will generate large groups, hence will be appropriate if jitter is large, but the groups may encompass irrelevant events; conversely, if jitter is small, small values of $\Theta$ are more appropriate. The details of the procedure, and illustrations of the influence of $\Theta$, are described in Appendix A.

## 3. Results

### 3.1. Choice of the bump functions

Different kinds of five, six and seven-parameter functions were tested for bump modelling. The most parsimonious functions that were tested are the Gaussian function and the half-ellipsoid function, which have five parameters.

- Gaussian function:

$$\varphi_b(f,t) = \frac{a}{2\pi s_f s_t} \exp\left[-\frac{(f-\mu_f)^2}{2s_f^2}\right] \exp\left[-\frac{(t-\mu_t)^2}{2s_t^2}\right]$$

  where $\mu_f$ and $\mu_t$ are the coordinates of the centre of the bump, $s_f$ and $s_t$ are the standard deviations along the vertical and horizontal axes respectively, and $a$ is the amplitude of the function.
- Half ellipsoid function (shown on Fig. 7):

$$\varphi_b(f,t) = a\sqrt{1-v} \quad \text{for } 0 \le v \le 1$$
$$\varphi_b(f,t) = 0 \quad \text{for } v > 1$$

  where $v = (e_f^2 + e_t^2)$ with $e_f = (f-\mu_f)/l_f$ and $e_t = (t-\mu_t)/l_t$. $\mu_f$ and $\mu_t$ are the coordinates of the centre of the ellipsoid, $l_f$ and $l_t$ are the half lengths of the principal axes, and $a$ is the amplitude of the function.

The quality of the modelling of time–frequency maps, in terms of accuracy and number of bumps required, was assessed on real signals. Functions with more than five parameters provided more accurate modelling, with fewer bumps and a smaller modelling error, than the Gaussian and half-ellipsoid functions. However, the latter are more parsimonious for the application at hand. Half ellipsoids were chosen because they led to a smaller modelling error than Gaussians: the average modelling error, and the residuals defined by:

$$\rho = \frac{\sum_t \sum_f z_{ft} - \sum_b \sum_t \sum_f \varphi_B^b(f,t)}{\sum_t \sum_f z_{ft}},$$

where $\varphi_B^b(f,t)$ denotes the $b$-th bump function involved in the model, are smaller for half ellipsoids (Fig. 8). Therefore, the half ellipsoid function is appropriate for the signals under investigation.

### 3.2. Validation of the method on artificial signals

In order to validate the method, two types of signals (type $A$ and type $B$) were generated, containing similar oscillations, at the same time–frequency locations, but with different probabilities of occurrence. Those signals contain

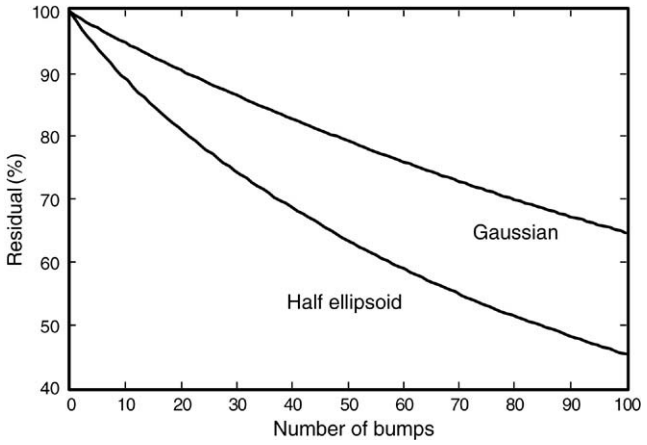


Fig. 8. Modelling residual $\rho$ for real signals modelled with Gaussians (top curve) or half ellipsoids (bottom curve) as a function of the number of bumps (results of the modelling of 100 maps of signals recorded from the rat olfactory bulb during 1.5 s).

three oscillations, denoted by $a$, $b$, and $c$, centred at points $a$ (55 Hz; 1.5 s), $b$ (80 Hz; 1.15 s) and $c$ (30 Hz; 0.85 s) respectively. Those oscillations last 3.5 periods, and their amplitude $U$ can take one of three values (0, 1, or 4) with the following probabilities:

- Type $A$ signals (Fig. 9):

  $\Pr(U=1|a) = 1$
  $\Pr(U=4|b) = 0.4$ and $\Pr(U=0|b) = 0.6$
  $\Pr(U=4|c) = 0.4$ and $\Pr(U=0|c) = 0.6$.

  In addition, jitter was generated by shifting the $a$ components in time, with respect to point $a$, by random shifts $s$, uniformly distributed in [−50 ms, +50 ms]; $b$ and $c$ components were not shifted.
- Type $B$ signals (Fig. 10):

  $\Pr(U=1|b) = 1$
  $\Pr(U=4|a) = 0.4$ and $\Pr(U=0|a) = 0.6$
  $\Pr(U=4|c) = 0.4$ and $\Pr(U=0|c) = 0.6$.

  In addition, jitter was generated by shifting the $b$ components in time, with respect to point $b$, by random shifts $s$, uniformly distributed in [−50 ms, +50 ms]; $a$ and $c$ components were not shifted.

A low-amplitude Gaussian white noise (mean 0, standard deviation 0.5) was added to the signals. 100 type $A$ signals and 100 type $B$ signals, 2.5 s in duration, were generated.

Therefore, the bump modelling method is expected to find out automatically that $a$ components are invariant in type $A$ signals, and that $b$ components are invariant in type $B$ signals. On that basis, an automatic classifier discriminating type $A$ signals from type $B$ signals should be designed. Note that the problem is made difficult by the fact that the amplitude $U$ of the discriminant features is small ($U = 1$) with respect to that of the irrelevant components ($U = 4$).

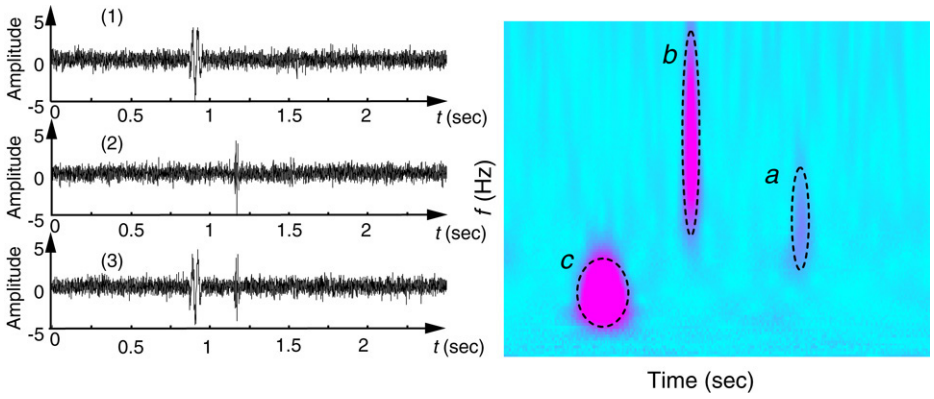


Fig. 9. Type *A* signals. Left: Examples of artificial signals containing oscillations located at points *a*, *b* and *c*. From top to bottom (1) *a* and *c*; (2) *b* and *a*; (3) *a*, *b* and *c* oscillations. Right: time–frequency map of the third signal, containing the three oscillations. Oscillation *a* is close to the background noise, but is nevertheless present in all signals. The time–frequency map on the right is to be compared with the maps of Figs. 10 and 11; note that signal (1) (*a* and *c* components only) cannot be discriminated visually from signal (1) of Fig. 10 (*b* and *c* components only). A colour version of the figure is posted at: http://www.neurones.espci.fr/NNT_1005_color_figures/Figures.html.

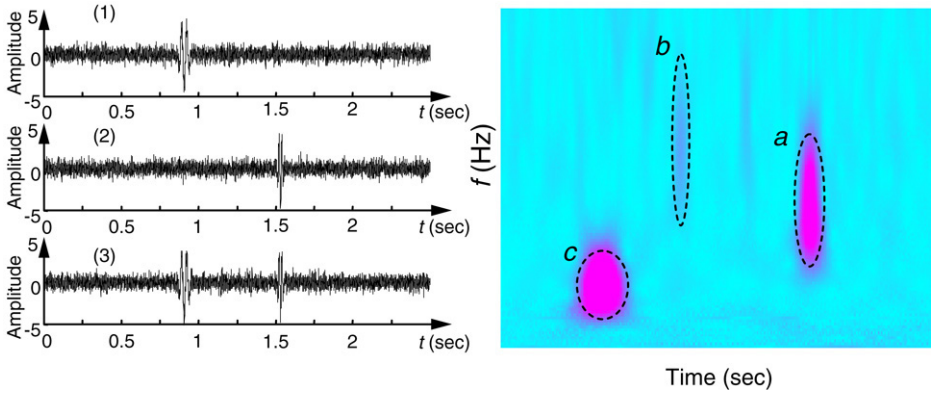


Fig. 10. Type *B* signals. Left: examples of artificial signals containing oscillations located at points *a*, *b* and *c*. From top to bottom (1) *b* and *c*; (2) *a* and *b*; (3) *a*, *b* and *c* oscillations. Right: Time–frequency map of the third signal, containing the three oscillations. Oscillation *b* is close to the background noise, but is nevertheless present in each signal. The time–frequency map on the right is to be compared with the map of Fig. 9; note that signal (1) (*b* and *c* components only) cannot be discriminated visually from signal (1) of Fig. 9 (*a* and *c* components only). A color version of the figure is posted at: http://www.neurones.espci.fr/NNT_1005_color_figures/Figures.html.

### 3.2.1. Invariance rates

Fig. 11 shows that averaging over maps does not provide any relevant information: the *c* components appear prominently, although they are irrelevant for our purpose.

Table 1 shows, for each oscillation and each signal type, the invariance rate computed for the group that is closest, in the time–frequency map, to the oscillation of interest. High invariance rates are obtained for the relevant oscillations, i.e. oscillation *a* for type *A* signals, and oscillation *b* for type *B* signals. More specifically, 47% of invariant bumps of type *A* signals model oscillation *a*, 23% of invariant bumps of type *A* signals model oscillation *b*, and 30% of invariant bumps of type *A* signals model oscillation *c*. If no noise and no jitter were present, the percentages would be 56%, 22% and 22% respectively.

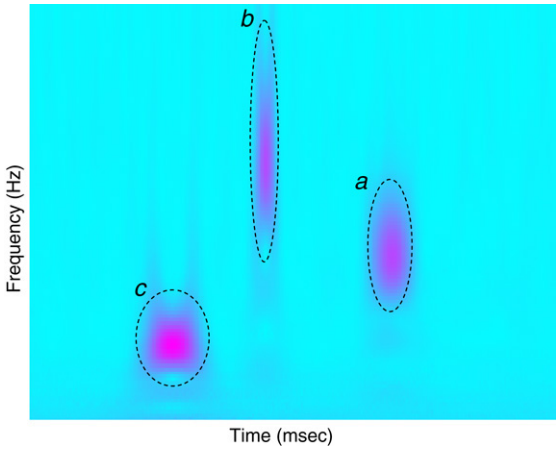


Fig. 11. Normalized time–frequency map of type *A* signals, averaged over the 100 generated examples. Oscillations *c* appear prominently, although they are less invariant than *a* oscillations. A colour version of the figure is posted at: http://www.neurones.espci.fr/NNT_1005_color_figures/Figures.html.

### 3.2.2. Relevance of ANOVA statistical tests

As mentioned in Section 1, analysis of variance is frequently used in the processing of time–frequency maps, in order to

Table 1
Detection of invariant areas

| Invariance rates | Type-$A$ signals | Type-$B$ signals |
|---|---|---|
| Oscillation $a$ | **0.91**, centroid 56 Hz/1.43 s | 0.52, centroid 39 Hz/1.45 s |
| Oscillation $b$ | 0.44, centroid 106 Hz/1.08 s | **0.82**, centroid 86 Hz/1.07 s |
| Oscillation $c$ | 0.58, centroid 26 Hz/0.88 s | 0.44, centroid 31 Hz/0.90 s |

Invariance rates of the groups that are closest to the location of the oscillations in the time–frequency map, with $\Theta = 5$, for $A$ and $B$ type signals. Boldface figures indicate the most invariant group detected.

determine whether a significant activity is present in some time–frequency domain. The most popular algorithm to that effect is ANOVA. It is based on the assumption of Gaussian distributions, an assumption whose validity is seldom checked, and which is frequently invalid. Typically, applying ANOVA to the artificial signals described in the present section gives extremely poor results. Alternatively, non-parametric tests such as the Wilcoxon rank sum test may provide satisfactory results, provided the windows are chosen from prior knowledge, and located very accurately in order to avoid boundary effects; when performed on the above signals with arbitrary windows (e.g. a regular grid), that test, although applicable in principle, provides unsatisfactory results. However, choosing windows from prior knowledge obviously generates 'clustering fallacies' (see for instance Gilovich, Vallone, and Tversky (1985)). Therefore, ANOVA statistical tests are not an appropriate approach for the automatic determination of areas of interest and feature extraction; they may be useful for posterior confirmation, if necessary, provided the underlying gaussian assumptions are true.

### 3.2.3. Automatic signal discrimination

Type $A$ and type $B$ signals cannot be discriminated easily, since the oscillations are at the same time–frequency locations in both signals, differing only by their probabilities of occurrence. Moreover, background noise is present, together with jitter. Therefore, automatic classification can be expected to be a difficult task. In the present section, it is shown that bump modelling extracts the appropriate features for classification.

A window of extension (30 Hz × 150 ms) was defined around each position $a$, $b$ and $c$. The most natural feature, in the context of the present problem, is the number of bumps per window. However, the time shift of the bumps with respect to the centre of the windows is also a quantity of interest, since components $a$ are shifted in type $A$ signals while component $b$ is shifted in type $B$ signals. Therefore, the classifier is fed with two numbers per window:

- the number of bumps within the window,
- the normalized distance $e$ between the centre of the window and the centre of the bump that is closest to the centre of the window: $e = (b_t - w_t)/(L/2)$, where $b_t$ is the time location of the centre of the bump that is closest to the centre of the window, $w_t$ is the time location of the centre of the window, and $L$ is the time extension of the window. If there is no bump within the zone, then $e = 1$.

In the usual empirical risk minimization framework (Vapnik, 1995), neural network classifiers of increasing complexity were trained, starting from linear classifiers, i.e. multilayer Perceptrons with zero hidden unit (see for instance Dreyfus (2005)). The loss function was the squared modelling error. In view of the low complexity of the model and the large number of examples, no regularization term was added to the cost function.[3] Model selection was performed by leave-one-out cross-validation (Stone, 1974), whereby each signal is withdrawn in turn, training is performed on the $N_S - 1$ remaining signals, and the classification error of the left-out signal is computed. The procedure is iterated $N_S$ times, and the leave-one-out score, i.e. the error rate on the left-out examples, is computed as

$$\frac{1}{N_S}\sum_{k=1}^{N_S} e^{(-k)}$$

where $e^{(-k)}$ is the classification error (1 or 0) on example $k$ when the latter is withdrawn from the training set. The leave-one-out score is an unbiased estimator of the generalization error (Vapnik, 1995).

For each model complexity, the procedure was iterated ten times with different parameter initializations; Table 2 shows that, as expected, the training error decreases steadily with increasing complexity, while the optimal generalization ability is obtained with 4 hidden neurons.

As a baseline for assessing the relevance of the features derived from bump modelling, consider signal (1) shown on Fig. 9 (type $A$ signal with components $a$ and $c$ only), and signal (1) shown on Fig. 10 (type $B$ signal with components $b$ and $c$ only); they cannot be discriminated visually, since the discriminant features (oscillation $a$ for type $A$ and oscillation $b$ for type $B$) are buried in noise. Such 'difficult' signals occur with probability $P[(U = 0|b) \wedge (U = 4|c)] = P(U = 0|b)P(U = 4|c) = 0.24$; therefore, a classifier that would classify those signals randomly would have an error rate of 12%, assuming an error rate of 0 on all other signals. The classifier using the features derived from bump modelling performs clearly better than that, which shows that the method provides relevant features even in a very difficult case.

## 3.3. Results on real signals

Signals recorded from the postero-ventral part of the olfactory bulb of rats were investigated, as described in Martin, Gervais, Hugues, Messaoudi, and Ravel (2004), Ravel et al. (2003). In those experiments, freely behaving rats had to learn to discriminate between two odours, one associated with a palatable drink and the other with no reward or with an aversive drink (trained rats will be referred to as 'expert rats', other rats as 'beginners' or 'naïve rats'). By inspection of the time–frequency maps obtained by Morlet wavelet transforms of recorded signals, it was observed that odour sampling

[3] Similar results to those reported here were obtained with SVM classifiers, which have a built-in regularization term.

Table 2
Error rates for the classification of 100 type *A* signals and 100 type *B* signals, for 100 different initializations of each classifier

| Number of hidden neurons | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| % error on training set | 11.4 | 10.1 | 6.9 | 3.9 | 2.3 | 1.7 | 1.7 | 1.4 |
| % leave-one-out error | 11.5 | 9.5 | 8.5 | 8.0 | 7.0 | 7.5 | 7.5 | 8.00 |

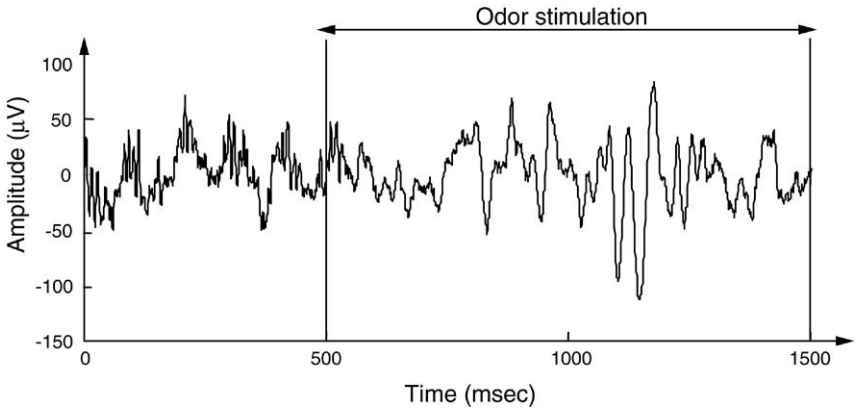


Fig. 12. Example of real signal (expert rat, odour = eugenol, stimulus onset at 500 ms).

induced marked changes in both gamma and beta oscillatory frequencies, and that beta activity could represent a correlate of olfactory recognition. The demonstration of the existence of a statistically significant correlation would be of importance in assessing the hypothesis that some oscillatory regimes would play a functional role in cognition.

A typical signal is shown on Fig. 12. Sampling started 500 ms before odour onset, and stopped 1 s after odour onset. The signals were sampled at 2 kHz, filtered between 1 and 300 Hz, and wavelet transformed between 10 and 100 Hz. 200 trials, recorded from five different rats performing the Go/No Go odour memory task described above were available. For each rat, 20 signals recorded before training, and 20 signals recorded after training, were present in the database.

In Section 3.3.1, we describe the design of a simple classifier, whose inputs are appropriate parameters of the bump model of a time–frequency map; we show that the classifier can discriminate recordings of expert rats from recordings of naïve rats, which demonstrates that the bump model extracts from the map the relevant features of the oscillatory activity related to the task.

In Section 3.3.2, we show an example of information that can be derived from the analysis of the bump models of time–frequency maps of signals recorded from expert rats.

### *3.3.1. Beginner/expert rat discrimination*

In the present section, our classification method (described in Section 3.2.3) is applied to discriminate beginners from expert rats on the basis of experimental recordings.

Each olfactory bulb signal, recorded during odour stimulation, was modelled with bumps as described above. Twelve windows, covering the whole time–frequency area of our signals, were defined: each window was 700 ms in length, and 40 Hz in height. Those windows were centred at 35, 55 and 75 Hz, and at 200, 700, 1200 and 1700 ms. The number of bumps whose time–frequency centre lies within the limits of each window was computed, resulting in twelve values that were input to the classifier. Model selection was performed by leave-one-out cross-validation, as described in Section 3.2.3.

Table 3
Error rates for the classification of 100 recordings from beginner rats and 100 recordings from expert rats *during olfactory stimulation*, as a function of the number of hidden neurons

| Number of hidden neurons | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| % training error | 22.6 | 15.1 | 12.3 | 7.5 | 3.76 |
| % leave-on-out error | 28.5 | 24.0 | 23.5 | 22.0 | 24.0 |

Table 3 shows the classification results. The most appropriate classifier complexity is three hidden neurons. To the best of our knowledge, no published results on the automatic classification of LFP signals in the olfactory bulb are available for comparison. Nevertheless, the relevance of the features extracted by bump modelling can be shown as follows: to serve as a baseline, signals recorded from the same animals, and the same electrodes, during inter-trial intervals (resting spontaneous activity), were classified. In those conditions, the smallest percentage of error rates is 40.0% in the validation set (Table 4, most appropriate classifier complexity is two hidden neurons); therefore, that classification is essentially random.

Fig. 13 shows pictorially the separation of the classes performed by the classifiers for signals recorded during odour stimulations and between odour stimulations. In those experiments, a single feature (number of bumps) was used for discrimination. Additional features (bump width, bump

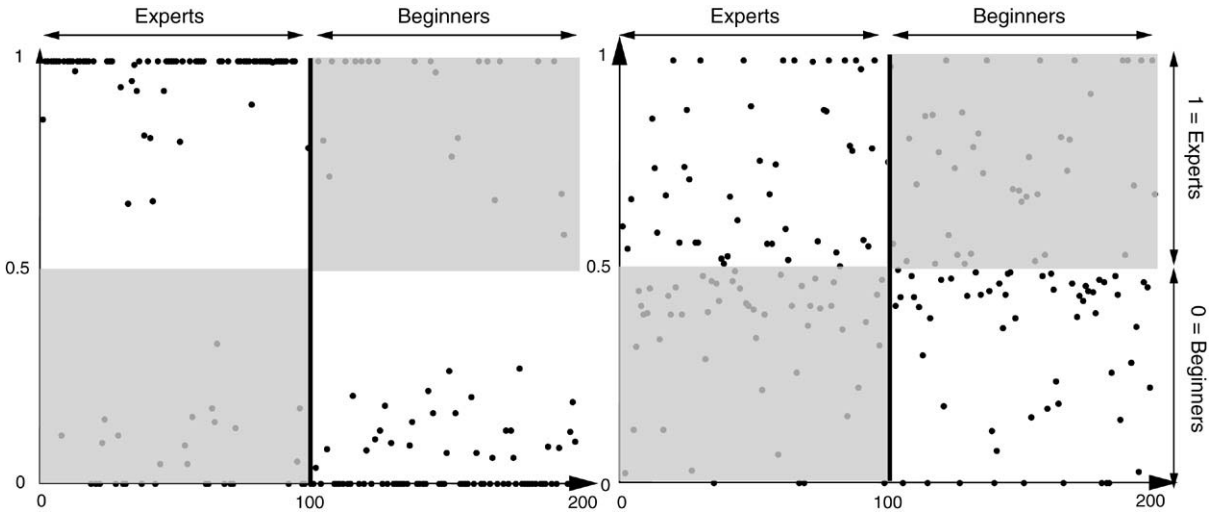


Fig. 13. Classifier outputs for 100 recordings (numbered 1–100) of expert rats and 100 recordings (numbered 101–200) of beginner rats, of any origin (five different rats, postero-ventral electrodes), of a validation set; classifier: Neural network; expert rats are labelled as 1, and beginner rats are labelled as 0. Left: Most appropriate (three hidden neurons) classifier outputs for recordings during odour stimulation; right: Most appropriate (two hidden neurons) classifier outputs for recordings during inter-trial intervals. Grey areas show erroneous classifications (bottom = false negatives, top = false positives).

Table 4
Error rates for the classification of 100 recordings from beginner rats and 100 recordings from expert rats *in the absence of olfactory stimulation* (during inter-trial intervals), as a function of the number of hidden neurons

| Number of hidden neurons | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| % training error | 43.5 | 35.6 | 29.3 | 16.9 | 9.4 |
| % leave-one-out error | 62.0 | 50.0 | 40.0 | 40.0 | 42.0 |

location, ...) for improving the error rate are currently being considered.

### 3.3.2. *Physiological and medical relevance of bump models*

A detailed discussion of the physiological relevance and implications of the results obtained on LFP signals can be found in Vialatte (2005), and will be published elsewhere. An example of time–frequency map of a signal recorded from an expert rat in response to eugenol, and of its bump model, is shown on Fig. 14.

Fig. 15 shows the invariant groups found from 182 recordings of naïve rats (left) and 166 recordings from expert rats (right), in response to eugenol. All invariant groups are found in the beta band, which is in agreement with the results described in Martin et al. (2004). For all odours on which experiments were performed, the significant differences between naïve and expert rats were found in the beta band, 0.3 to 0.6 s after the inception of the stimulus. Investigations of the responses to other odours are under way.

Detailed discussions of the results obtained by bump modelling of EEG signals have been published in Vialatte (2005) and Vialatte et al. (2005a, 2005b). As an example, Fig. 16 shows the bump model of an EEG signal recorded from a patient with mild cognitive impairment. It has been shown that the bump modelling of EEG signals is a promising candidate for population-wide early detection of Alzheimer's disease.

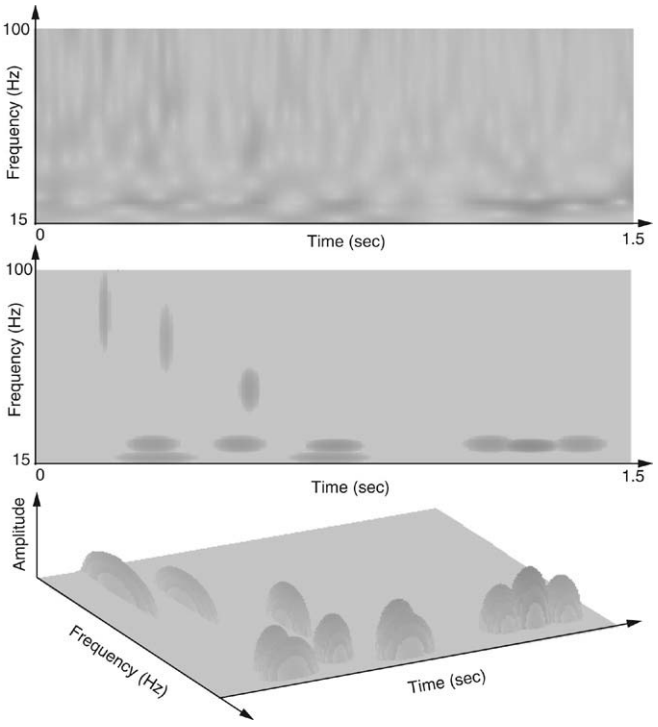


Fig. 14. Time–frequency map (top) and its bump model in two (middle) and three (bottom) dimensions. Recording of a LFP signal from an expert rat in response to eugenol.

## 4. Conclusion

The present paper describes a novel method for analyzing automatically large amounts of experimental data from electrophysiological experiments. It is based on the modelling of the time–frequency maps, obtained by wavelet transformation of the signals, which are usually analyzed 'visually'. The map is modelled as a sum of simple 'bump' functions, and it has

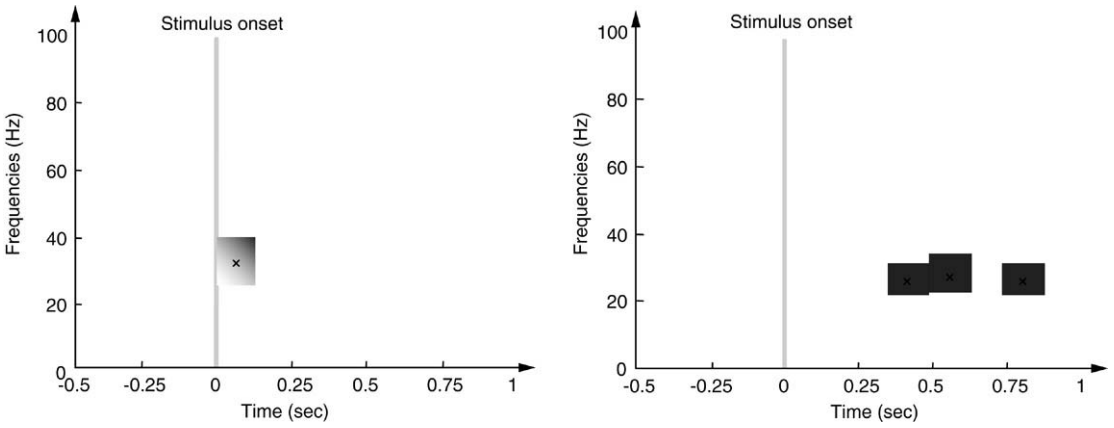


Fig. 15. Invariant groups of bumps in response to eugenol. Left: Naïve rats (182 recordings); right: Expert rats (166 recordings).

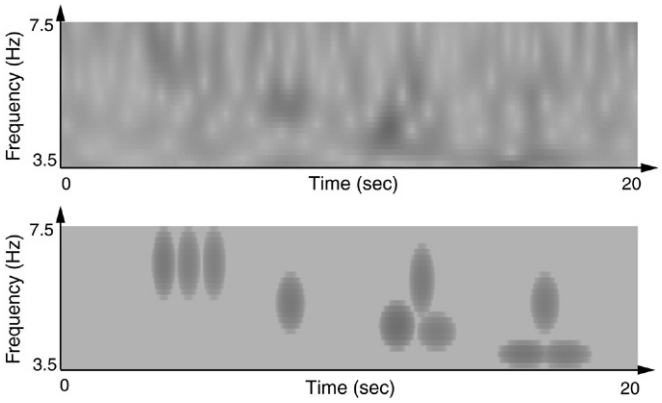


Fig. 16. Time–frequency map (top) and its bump model (bottom) for an EEG recording from a patient with mild cognitive impairment. Result obtained in collaboration with A. Cichocki, T. Rutkowski (Riken Brain Science Institute), and T. Musha (Brain Function Laboratory). A color version of the figure is posted at: http://www.neurones.espci.fr/NNT_1005_color_figures/Figures.html.

been shown that the resulting model allows the detection of time–frequency events that occur with some degree of invariance in the experiments, and are related to the activation of different neural populations.

One of the assets of the method is the fact that it allows experts to validate the result of bump modelling: the starting point is the time–frequency maps with which they are familiar, and the accuracy and relevance of the resulting bump model can be checked visually and estimated numerically. In addition, it provides a quantitative estimate of the reproducibility of the time–frequency events detected by the model.

The method was first validated on synthetic signals, and it was shown that automatic discrimination, based on features extracted from the bump modelling, is feasible and efficient. The method was also applied to real signals recorded from the olfactory bulb during Go/No-go experiments. The relevance of the bump model was demonstrated by the design of an automatic classifier, whose inputs are features derived from the bump model, which allowed the automatic discrimination of beginner rats from naïve rats, based solely on their LFP recordings during the tasks. A detailed discussion of the biological relevance of the bump models derived from the recordings can be found in Vialatte (2005), and will be published in a forthcoming article.

The feasibility and efficiency of the automatic analysis and classification of large numbers of time–frequency maps have also been demonstrated on EEG data for early Alzheimer's disease detection, as described in Vialatte et al. (2005a, 2005b). However, there is still room for improvements, especially by a more thorough selection of the features that are used for automatic classification; that topic is still under investigation.

## Acknowledgments

The authors would like to thank J. Kahana and J.B. Caplan, W.J. Freeman, O. Bertrand, J.P. Lachaux and S. Roux for their help and discussions. This work was supported by the Centre de Microélectronique de Paris-Île de France. François Vialatte and Claire Martin were supported by M.E.N.E.S.R grants.

## Appendix A. Procedure for the generation of groups and the computation of their invariance rates

### A.1. Distance

The above approach requires the definition of a distance on a time–frequency map. Since the variables on the $x$- and $y$-axes have different physical meanings, dimensionless variables must be defined. The time resolution is a natural reference for the distance between bumps along the time axis, and the frequency resolution for the distance along the frequency axis: the dimensionless distance $\delta x$ between two points lying at the same frequency, and at a distance $\delta t$ equal to the time resolution $\sigma_t$, should be equal to the dimensionless distance $\delta y$ between two points lying at the same time, and at a distance $\delta f = (f_1 - f_2)$ equal to the frequency resolution $\sigma_{\bar{f}}$ at the mean frequency $\bar{f}$: if $\delta t = \sigma_t$ and $\delta f = \sigma_{\bar{f}}$, then $\delta x = \delta y$.

Moreover, it is convenient to define the dimensionless distance in time as the ratio of the distance in time to the period $T$ at the frequency of interest:

$$\delta x = \delta t / T = f \delta t.$$

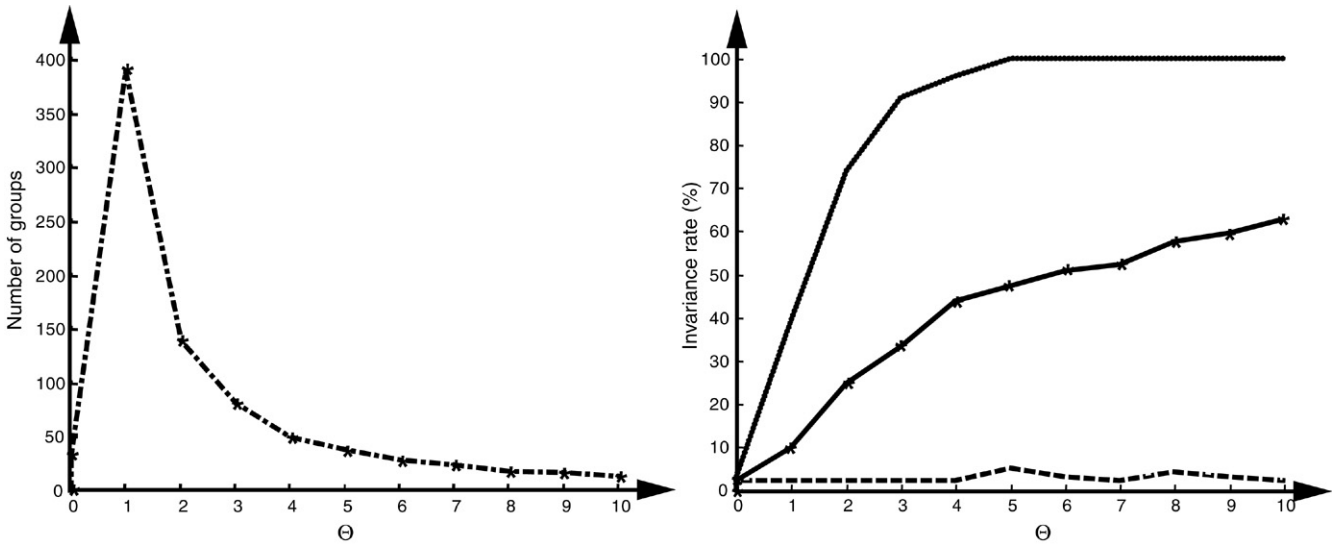


Fig. 17. Effect of $\Theta$ on the generation of groups, for 100 real signals recorded from the rat olfactory bulb (Ravel et al., 2003) of five expert rats, sampled during 500 ms prior to and 1 s after the stimulation. Left: Number of invariant groups vs. $\Theta$, Right: Maximal, average and minimal (dotted, solid and dashed lines respectively) invariance rate $R$ of the generated groups vs. $\Theta$. $\Theta$ is varied from 0 to 10 in increments of one; additionally, results for $\Theta = 0.01$ are shown. If the threshold is small (e.g. $\Theta = 0.01$), the number of groups is small; as the threshold increases, the tolerance to jitter increases, so that a larger number of groups are generated, with higher invariance rates; when the threshold becomes large (for $\Theta > 1$), groups coalesce so that the number of groups decreases. For $\Theta > 5$, the most invariant groups reach the maximum rate $R = 100\%$.

As a consequence, one has for two bumps centred at frequencies $f_1$ and $f_2$:

$$\delta y = \frac{\sigma_t}{\sigma_f}\frac{f_1+f_2}{2}\delta f = \frac{49}{\pi}\frac{\delta f}{f_1+f_2}$$

(using relations (2) and (3)) with $\delta f = (f_1 - f_2)$.

Finally, the distance between two bumps $b_1$ $(f_1, t_1)$ and $b_2$ $(f_2, t_2)$ is equal to:

$$d(b_1, b_2) = \sqrt{\delta x^2 + \delta y^2}.$$

### A.2. Detection of invariant areas

The $B$ bumps arising from the modelling of $N$ maps within a neighbourhood of size $\Theta$ are clustered iteratively by the following algorithm:

- Initialization
  - Compute the matrix of distances between the bumps of all possible pairs of bumps belonging to two different maps.
- Cluster formation
  - For each bump $b_r (r = 1, \ldots B)$, find its nearest neighbour in each map (except the map $b_r$ belongs to); denote by $b_{r,c}$ the nearest neighbour of bump $b_r$ in map $c$ ($c = 1$ to $N - 1$).
  - Among those bumps, consider the bumps $\{b_{r,k}\}$ ($k = 1 \ldots K_r$, $K_r \leq N-1$) such that $d(b_r, b_{r,k}) < \Theta$, where $\Theta$ is a radius that is chosen *a priori* from considerations that will be described below. Those bumps form a group $G_r$, containing $K_r$ bumps $b_{r,k}$ that are located at a distance of $b_r$ smaller than $\Theta$; denote by $D_r$ the quantity $D_r = \sum_{k=1}^{K_r} d(b_r, b_{r,k})$. $b_r$ will be referred to as the *centroid* of group $G_r$.
- Selection and bump elimination
  - Among all groups $G_r$ $(r = 1 \ldots B)$, find the group(s) $\{G_{r0}\}$ that contain(s) the largest number of bumps $K_{r0}$: $K_{r0} = \max_r(K_r)$.
  - Within the group(s) $\{G_{r0}\}$, select the most compact group $G_z$, i.e. the group such that $D_z = \min_{r0}(D_{r0})$, with invariance rate $R = K_z/N$: thus, the *invariance rate* of a group is the proportion of modelled signals that have at least one bump within that group.
  - Search for the bumps $b_{r,n}$ neighbouring $b_r$: find all bumps on all signals such that $d(b_r, b_{r,n}) < \Theta$. Withdraw the bumps $b_{r,n}$ from the set of bumps.
  - Iterate the formation of new groups $\{G_r\}$ until the largest group contains less than 2 bumps.

The groups $\{G_z\}$ thus obtained are the invariant areas of interest.

The maximal distance $\Theta$ has an influence on the number of groups $\{G_z\}$, and on the area of those groups. The smaller $\Theta$, the closer the bumps must be in order to be considered as neighbours. Therefore, a small value of $\Theta$ will select small zones that have a high density of invariant bumps, whereas a large value of $\Theta$ will generate larger clusters encompassing a larger number of bumps (Fig. 17). If one is interested in highly invariant events, and willing to accept a risk of missing some of them, a small value of $\Theta$ must be chosen; conversely, if a large amount of jitter is expected, then a larger value of $\Theta$ must be chosen, but irrelevant events might be captured. Therefore, a case-dependent tradeoff must be performed. In the absence of prior information, the value of the threshold that gives rise to the maximum number of groups $G_z$ should be chosen, and the groups with low invariance rates should be discarded.

The results reported in Section 3 were obtained with $\Theta = 5$, which is a satisfactory tradeoff between tolerance to jitter (just

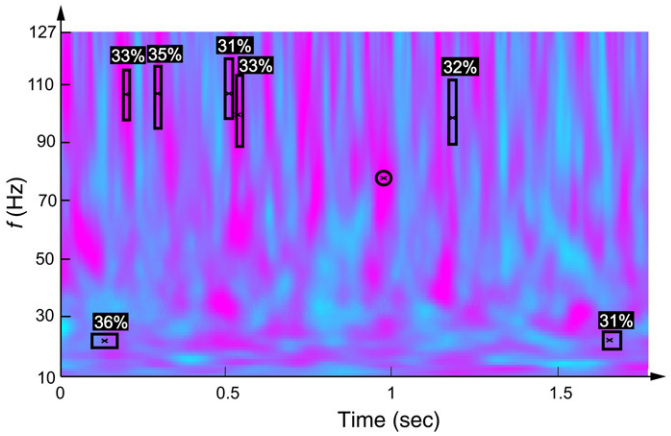


Fig. 18. Time–frequency map obtained by averaging 100 time–frequency maps of real signals recorded from the rat olfactory bulb (Ravel et al., 2003) of five beginner rats, sampled during 500 ms prior to and 1 s after the stimulation. The rectangles are the locations of the groups with invariance rate larger than 30%, obtained with $\Theta = 1$. The figures are the invariance rates; crosses indicate the centroids of the groups. The area of maximum amplitude of the averaged map is shown as a circle. The examples illustrate the fact that averaging the maps does not give reliable information on the invariant areas of the maps. A color version of the figure is posted at: http://www.neurones.espci.fr/NNT_1005_color_figures/Figures.html.

over the duration of an oscillation) and the number of bumps present in each group.

Fig. 18 shows the invariance analysis of real signals: the locations of invariant groups are superimposed on a time–frequency map obtained by averaging 100 such maps, from real signals. Clearly, the analysis provides information about invariant groups that cannot be provided by map averaging, because of the variability of the signals.

## Appendix B. Pruning of the bump models

The present appendix describes the procedure used for pruning a model made of a large number of bumps, in order to discard bumps that model noise. As described in Section 2.3.5, we denote by $F$ the fraction of the energy of the map that is described by a bump. Since there may be invariant bumps with small energies, a selection criterion must take into account the invariance rate of the bump, in addition to the energy modelled by that bump.

Consider a bump model built with a stopping criterion $F_s$, i.e. a model containing bumps that model a fraction of the map that is larger than $F_s$. Assume that the pruning method consists in discarding all bumps whose invariance rate is smaller than a threshold $T(F_s)$; an appropriate choice of $T(F_s)$ will be discussed below. We denote by $I(F_s)$ the number of 'informative' groups, i.e. the groups whose invariance rate is larger than $T(F_s)$, and by $N(F_s)$ the number of 'noise' groups, whose invariance rate is below the threshold. Therefore, we seek the value $F^*$ of $F_s$ for which the ratio $I(F_s)/N(F_s)$ is maximum. To that end, $F_s$ is varied between two values $F_{\min}$ and $F_{\max}$, where $F_{\max}$ is the fraction of the map that is modeled by the first bump built into the model, and $F_{\min}$ is the fraction modelled by the bumps when the amplitude of 10% of them is smaller than the maximum amplitude of the wavelet transform of white Gaussian noise.

The value of the threshold $T(F_s)$ is found as follows: having built a model with termination criterion $F_s$, additional maps are generated by randomly shuffling the bumps, and groups are built as described in Appendix A. The threshold is defined as the value of $T(F_s)$ whose cumulative distribution function is larger than 99%.

## References

Anemuller, J., Sejnowski, T. J., & Makeig, S. (2003). Complex independent component analysis of frequency-domain electroencephalographic data. *Neural Networks*, *16*, 1311–1323.

Blankertz, B., Müller, K. R., Curio, G., Vaughan, T. M., Schalk, G., Wolpaw, J. R., et al. (2004). The BCI Competition 2003: Progress and perspectives in detection and discrimination of EEG single trials. *IEEE Transaction on Biomedical Engineering*, *51*, 1044–1051.

Bogert, B. P., Healey, M. J. R., & Tukey, J. R. (1963). The quefrency analysis of time series for echoes: Cepstrum, pseudo-autocovariance, cross-cepstrum and saphe cracking. In M. Rosenblatt (Ed.), *Proceedings of the symposium on time series analysis* (pp. 209–243). New York: John Wiley and Sons.

Caplan, J. B., Madsen, J. R., Raghavachari, S., & Kahana, M. J. (2001). Distinct patterns of brain oscillations underlie two basic parameters of human maze learning. *Journal of Neurophysiology*, *86*, 368–380.

Cincotti, F., Mattia, D., Babiloni, C., Carducci, F., Bianchi, L., Millán, J., et al. (2002). Classification of EEG mental patterns by using two scalp electrodes and mahalanobis distance-based classifiers. *Methods of Information in Medicine*, *41*, 337–341.

Dreyfus, G. (2005). *Neural networks, methodology and applications*. Springer.

Dubois, R. (2004). Application de nouvelles méthodes d'apprentissage à la détection précoce d'anomalies en électrocardiographie. Thèse de Doctorat de l'Université Pierre et Marie Curie, Paris. Available from http://pastel.paristech.org/bib/archive/00000571/.

Dubois, R., Quenet, B., Faisandier, Y., & Dreyfus, G. (2006). Building meaningful representations for nonlinear modelling of 1D- and 2D-signals: Applications to biomedical signals. *Neurocomputing*, *69*, 2180–2192.

Düzel, E., Habib, R., Schott, B., Schoenfeld, A., Lobaugh, N., McIntosh, A. R., et al. (2003). A multivariate, spatiotemporal analysis of electromagnetic time–frequency data of recognition memory. *Neuroimage*, *18*, 185–197.

Freeman, W. J., & Schneider, W. (1982). Changes in spatial patterns of rabbit olfactory EEG with conditioning to odours. *Psychophysiology*, *19*, 44–56.

Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology*, *17*, 295–314.

Gruber, T., & Müller, M. (2005). Oscillatory brain activity dissociates between associative stimulus content in a repetition priming task in the human EEG. *Cerbral Cortex*, *15*, 109–1116.

Herrmann, C. S., Mecklinger, A., & Pfeifer, E. (1999). Gamma responses and ERPs in a visual classification task. *Clinical Neurophysiology*, *110*, 363–642.

Kahana, M. J., Sekuler, R., Caplan, J. B., Kirschen, M., & Madsen, J. R. (1999). Human theta oscillations exhibit task dependence during virtual maze navigation. *Nature*, *399*, 781–784.

Kapur, T., Grimson, W. E., Wells, W. M., & Kikinis, R. (1996). Segmentation of brain tissue from magnetic resonance images. *Medical Image Analysis*, *1*, 109–207.

Kronland-Martinet, R., Morlet, J., & Grossmann, A. (1987). The wavelet transform. In C. H. Chen (Ed.), *In expert systems and pattern analysis* (pp. 97–126). World Scientific.

Makeig, S., Bell, A. J., Jung, T. -P., & Sejnowski, T. J. (1996). Independent component analysis of electroencephalographic data. *Advances in Neural Information Processing Systems*, *8*, 145–151.

Mallat, S. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *11*, 674–693.

Mallat, S., & Zhang, Z. (1993). Matching pursuit with time–frequency dictionaries. *IEEE Transactions on Signal Processing*, *41*, 3397–3414.

Martin, C., Gervais, R., Hugues, E., Messaoudi, B., & Ravel, N. (2004). Learning modulation of odor-induced oscillatory responses in the rat olfactory bulb: A correlate of odor recognition? *Journal of Neuroscience*, *24*, 389–397.

Poularikas, A. D. (1996). *The transforms and applications handbook*. CRC Press.

Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (2002). *Numerical recipes in C: The art of scientific computing* (pp. 425–430). New York: Cambridge Univ. Press.

Rabiner, L. R. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, *77*, 257–286.

Ravel, N., Chabaud, P., Martin, C., Gaveau, V., Hugues, E., Tallon-Baudry, C., et al. (2003). Olfactory learning modifies the expression of odour-induced oscillatory responses in the gamma (60–90 Hz) and beta (15–40 Hz) bands in the rat olfactory bulb. *European Journal of Neuroscience*, *17*(2), 350–358.

Singer, W. (1993). Synchronization of cortical activity and its putative role in information processing and learning. *Annual Reviews of Physiology*, *55*, 349–374.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions (with discussion). *Journal of the Royal Statistical Society B*, *36*, 111–147.

Tallon-Baudry, C., Bertrand, O., Delpuech, C., & Pernier, J. (1996). Stimulus specificity of phase-locked and non-phase-locked 40 Hz visual responses in human. *Journal of Neuroscience*, *16*, 4240–4249.

Tallon-Baudry, C., Bertrand, O., Delpuech, C., & Pernier, J. (1997). Oscillatory gamma-band (30–70 Hz) activity induced by a visual search task in humans. *Journal of Neuroscience*, *17*, 722–734.

Tallon-Baudry, C., Bertrand, O., Peronnet, F., & Pernier, J. (1998). Induced gamma-band activity during the delay of a visual short-term memory task in humans. *Journal of Neuroscience*, *18*, 4244–4254.

Vapnik, V. (1995). *The nature of statistical learning theory*. New York: Springer.

Vialatte, F. (2005). Modélisation en bosses pour l'analyse des motifs oscillatoires reproductibles dans l'activité de populations neuronales: Applications à l'apprentissage olfactif chez l'animal et à la détection précoce de la maladie d'Alzheimer. *Thèse de l'Université Pierre et Marie Curie, Paris.* Available from http://pastel.paristech.org/bib/archive/00001508/.

Vialatte, F., Cichocki, A., Dreyfus, G., Musha, T., & Gervais, R. (2005a). Early detection of Alzheimer's disease by blind source separation, time frequency representation, and bump modeling of EEG Signals. In *Lecture Notes in Computer Science*: *Vol. 3696* (pp. 683–692). Springer.

Vialatte, F., Cichocki, A., Dreyfus, G., Musha, T., & Gervais, R. (2005b). Blind source separation and sparse bump modelling of time–frequency representation of EEG signals: New tools for early detection of Alzheimer's disease. In *Machine learning for signal processing (MLSP 2005)*.