

A Glimpse into Your Vision

Sara L. Gonzalez Andino,^{1,2*} Rolando Grave de Peralta,^{1,2} Asaid Khateb,^{3,4,5}
Alan J. Pegna,^{3,4,5} Gregor Thut,⁵ and Theodor Landis²

¹Electrical Neuroimaging Group, Neurology Department, Geneva University Hospitals,
Geneva, Switzerland

²Neurology Department, Geneva University Hospitals, Geneva, Switzerland

³Laboratory of Experimental Neuropsychology, Department of Neurology, Geneva University
Hospitals, Geneva, Switzerland

⁴Neuropsychology Unit, Department of Neurology, Geneva University Hospitals, Geneva, Switzerland

⁵Human Brain Mapping Laboratory, Department of Neurology and Department of Fundamental
Neurosciences, University of Geneva, Geneva, Switzerland

Abstract: Invasive recordings of local field potentials (LFPs) have been used to “read the mind” of monkeys in real time. Here we investigated whether noninvasive field potentials estimated from the scalp-recorded electroencephalogram (EEG) using the ELECTRA source localization algorithm could provide real-time decoding of mental states in healthy humans. By means of pattern recognition techniques on 500-ms EEG epochs, we were able to discriminate accurately from single trials which of four categories of visual stimuli the subjects were viewing. Our results show that it is possible to reproduce the decoding accuracy previously obtained in animals with invasive recordings. A comparison between the decoding results and the subjects’ behavioral performance indicates that oscillatory activity (OA), elicited in specific brain regions) codes better for the visual stimulus category presented than the subjects’ actual response, i.e., is insensitive to voluntary or involuntary errors. The identification of brain regions participating in the decoding process allowed us to construct 3D-functional images of the task-related OA. These images revealed the activation of brain regions known for their involvement in the processing of this type of visual stimuli. Electrical neuroimaging therefore appears to have the potential to establish what the brain is processing while the stimuli are being seen or categorized, i.e., concurrently with sensory-perceptual processes. *Hum Brain Mapp* 28:614–624, 2007. © 2006 Wiley-Liss, Inc.

Key words: EEG; LFP; Inverse solutions; oscillations; pattern recognition; decoding; ELECTRA; brain rhythms

Contract grant sponsor: Swiss National Science Foundation; Contract grant numbers: 3200B0-105867, 3151A0-102271/1, 3151A0-102271/1, 320000-109928; Contract grant sponsors: IM2.MI on Brain Machines Interfaces, the European IST Program FET Project FP6-003758 (MAIA).

*Correspondence to: Sara L. Gonzalez Andino, Neurology Department, Geneva University Hospital, Rue 24 Micheli du Crest, 1211 Geneva 14, Geneva, Switzerland.

E-mail: Sara.GonzalezAndino@hcuge.ch

Received for publication 1 September 2005; Accepted 5 April 2006

DOI: 10.1002/hbm.20302

Published online 28 November 2006 in Wiley InterScience (www.interscience.wiley.com).

© 2006 Wiley-Liss, Inc.

INTRODUCTION

Until recently, “mind reading” was considered to belong to the realm of science fiction. Although still a controversial issue, the ability to decode a conscious subject’s cognitive state is on the brink of turning into reality [Mehring et al., 2003; Pesaran et al., 2002; Wessberg et al., 2000]. A recent functional MRI (fMRI) study has shown that it is possible to determine with a high level of accuracy which of eight stimulus orientations a subject is seeing [Kamitani and Tong, 2005]. However, due to the properties of the hemodynamic response on which the fMRI technique relies [Buckner et al., 1996; Menon and Kim, 1999], decoding of stimulus

orientation required periods of analysis extending beyond 16 s. Since during this period a normal brain can effectively process a multitude of visual stimuli, alternative neuroimaging approaches to the question of “mind reading” in healthy subjects have to be explored. Ideally, such approaches should possess a high temporal resolution and be able to trace neural activity in real time, while remaining noninvasive.

The most extensively explored technique for assessing mental chronometry is electroencephalography (EEG), which reflects millisecond-by-millisecond activity of neural populations. However, EEG signals measured on the scalp are severely limited in terms of spatial resolution, as they result from the spatial integration of the activity of large groups of neurons producing local field potentials (LFPs). Animal studies using intracerebral invasive recordings show that neural processes are coded within the temporal structure of LFPs in the form of oscillatory activity (OA). The information contained in OA has been shown to be efficient in predicting animal behavior [Mehring et al., 2003; Pesaran et al., 2002] or cognitive states [Gervasoni et al., 2004; Pesaran et al., 2002]. It is therefore reasonable to assume that similar information could be extracted from human EEG traces, providing that we can estimate the temporal structure of field potentials based on scalp recordings.

We have developed a method allowing a noninvasive, although coarse, estimation of field potentials from scalp-recorded EEG data [Grave de Peralta Menendez et al., 2000]. This method, termed ELECTRA (see below), is based on the neurophysiological properties of EEG generators and the fields they produce [Grave de Peralta Menendez et al., 2004]. Simulation studies [Grave de Peralta Menendez et al., 2000] and experimental findings [Gonzalez Andino et al., 2001, 2005; Grave de Peralta Menendez et al., 2004; Thut et al., 2000] indicate that this method can produce trustworthy estimates of the temporal structure of LFP. Since OA depends on the temporal structure of LFP, we hypothesized that accurate real-time decoding of mental states might be possible based on this method. In this investigation, the accuracy of OA estimates was assessed using a visual recognition task in which healthy subjects were asked to identify words (W), nonwords (NW), images (I), or nonimages (NI), while EEG was simultaneously recorded [Khateb et al., 2002]. We developed a procedure to select the OA that best discriminated between the categories of visual stimuli and then used this information to decode, on a separate set of trials, the stimuli that were being presented to the subjects. Given the correlation that has been observed between the BOLD responses and the LFP oscillations in animals [Logothetis et al., 2001; Niessing et al., 2005], we predicted that areas yielding the highest discrimination in terms of OA between the different types of stimuli would correspond to brain networks activated in similar tasks using fMRI paradigms.

MATERIALS AND METHODS

The dataset presented here was randomly selected from a larger one that has been described in detail elsewhere [Khateb et al., 2002].

Participants

Ten native French-speaking subjects (mean age 22 ± 3 years, five female) were included in this study. They were all right-handed with normal or corrected-to-normal vision and had given written informed consent as recommended by the ethical committee of the Geneva University Hospitals. They were all medication-free and had no history of neurological diseases at the time of the experiment.

Experimental Paradigm

Four categories of stimuli were presented. They consisted of either verbal or pictorial items. The verbal items were either concrete imaginable high-frequency French words (W; $n = 40$; e.g., “train”) or phonologically plausible nonwords (NW; $n = 40$; e.g., “prande”). In the pictorial items, the stimuli selected from the Snodgrass and Vanderwart [1980] set were either recognizable black-and-white drawings (representing living and nonliving items), referred to as the image condition (I, $n = 40$), or a scrambled unrecognizable version of the same drawings referred to as the nonimage condition (NI, $n = 40$, see Fig. 1). Altogether, these stimuli provided a set of 160 pseudorandomized trials.

Experiments were carried out in an isolated, electrically shielded room. The stimuli were presented using a Power Macintosh computer (17" screen, refresh rate 67 Hz) running MacProbe (v. 1.6.69, Aristrometrics, Woodland Hills CA). Following the presentation of a fixation cross for 500 ms, the stimulus was presented for 150 ms at the center of the screen. Manual responses were collected for 1,300 ms following stimulus onset leading to ~ 2 -s interstimulus interval. Subjects had to decide as quickly and as accurately as possible whether the stimulus represented a meaningful word/image or not. Responses were given by pressing a key with the (right hand) index finger for words/images and with the middle finger for nonwords/nonimages. Before the experimental block, subjects underwent a training session in order to ensure maximum comprehension of the task demands.

The EEG was recorded (using a 64-channel system, hardware: M&I, Prague, Czech Republic; software: Neuroscience Technology Research, Prague, Czech Republic) from 47 equidistant electrodes placed manually over the scalp according to the extended 10/20 system [Khateb et al., 2000]. The signals, recorded against Cz site at 500 Hz, were recomputed off-line against the average reference. Filter settings were between 0.15–250 Hz and the impedances were kept around 5 kOhm.

Analysis Window and Data Preprocessing

Stimulus categorization necessarily operates on a single trial basis, with subjects responding to the items one by one, as they appear. Thus, to mimic this procedure using an automatic system, the whole analysis must be based on single trials. This differs from conventional EEG analysis, where off-line data processing typically involves the rejection of epochs

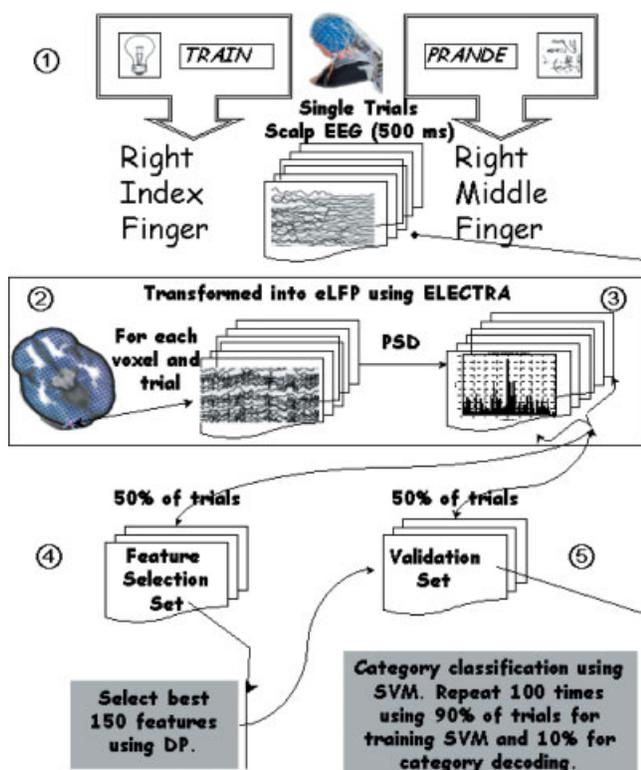


Figure 1.

Block diagram of the analysis procedure employed in this study. The whole procedure can be roughly divided into the five basic steps enumerated in the diagram: 1) data recording and selection of epoch length; 2) local field potential (LFP) estimation for all voxels and epochs (trials); 3) power spectral density estimation for each estimated LFP; 4) feature selections using the first half of the trials (learning set); and 5) categorization of the cognitive state over the second half of the trials (validation set).

containing artifacts, interpolation of bad electrodes, baseline correction, filtering, and finally averaging. Here, the analysis procedure was applied directly to the epochs of EEG data as they had been collected, without any off-line preprocessing.

Analysis Procedure

Here we use the term cognitive state to refer to the ongoing processing of a given stimulus in this task. Classification of a cognitive state into one of two (or more) classes involved several steps schematically shown in the block chart represented in Figure 1. The whole procedure can be roughly divided into the following five basic steps explained below: 1) data recording and selection of epoch length; 2) estimation of LFP using ELECTRA (eLFP) for all voxels and epochs (trials); 3) power spectral density estimation for each eLFP; 4) feature selections using the first half of the trials (learning set); and 5) categorization of the cognitive state over the second half of the trials (validation set).

Data recording and selection of analysis epoch length

Data recording procedures were described in the previous section. The length of the window of analysis included the period of presentation/categorization of the stimulus but minimized the influence of motor responses. Based on the shortest mean reaction time over subjects (W: 520 ± 96 ms, NW: 690 ± 120 ms, I: 529 ± 88 ms, NI: 564 ± 60 ms), this window was set between 0 and 500 ms after stimulus onset. This selection ensured that all subjects had completed visual object recognition within the defined period. In addition, the selection of a fixed analysis window for all subjects considerably facilitated the comparison between subjects and the averaging of results, since it led to identical spectral resolution and frequencies for all subjects and conditions. Still, some of the manual responses could have been included in this window, particularly in the case of rapid response times. To rule out the potential confound of classification based on motor responses, we designed additional specific analyses.

Transforming scalp-recorded EEG data into intracranial estimates of LFP

A technique widely used in signal processing to improve automatic categorization consists in the projection of features belonging to a low-dimensional space (the EEG in our case) into a high-dimensional space so as to improve class separability. While many different high-dimensional mathematical spaces can be selected, we employed here the space formed by the noninvasive intracranial estimates of LFP. It can be seen intuitively that cognitive states stand a better chance of being separated within this space than in the space of the data originally recorded. This is obvious, as the EEG data represents the overlapping, noisy spatiotemporal activity arising from very diverse brain regions; i.e., a single scalp electrode picks up and mixes the ongoing activity of different brain areas. Consequently, temporal or spectral features are merged on the same recording. For example, an electrode placed on the frontal midline picks up the combined activity of different motor areas (primary motor cortex, supplementary motor areas, anterior cingulate cortex, and motor cingulate areas) known to have different functional roles. Transforming EEG into eLFPs aims to “deconvolve” or “unmix” the scalp signals, attributing to each brain area its own temporal activity and leading to cleaner temporal/spectral estimates. This approach differs from usual mathematical transformations (e.g., ICA, PCA) since biophysical a priori information is incorporated that is not actually contained in the EEG measurements. The first a priori information included in the eLFP estimates concerns the fact that EEG measurements are a convolution of the intracranial LFPs at each brain voxel with a mixing matrix (the lead field). The latter expresses how electric fields propagate from the brain to the scalp. Further a priori information is incorporated through the choice of one specific inverse solution (see below). This information is neither contained in the original data, nor considered in pure mathematically based approaches.

Scalp EEG was transformed into eLFP using a distributed linear inverse solution termed ELECTRA [Grave de Peralta Menendez et al., 2000, 2004]. ELECTRA selects a unique solution to the bioelectromagnetic inverse problem based on physical laws governing propagation of field potentials in biological media. Such restrictions lead to a formulation of the inverse problem in which the unknowns are the electrical potentials within the whole brain, rather than the current density vector. In simpler terms, ELECTRA allows an estimation of the 3D distribution of electrical potentials (field potentials) within the whole brain as if they were recorded with intracranial electrodes. This is an important difference between ELECTRA and other existing inverse solutions that estimate the intracranial current density. Since the current density is the first spatial derivative of the electrical potential, there is little sense in comparing current density-based estimates with potentials recorded intracranially. We use the term eLFP to denote the retrieved magnitude because ELECTRA estimates field potentials using a local autoregressive model [Grave de Peralta Menendez et al., 2004]. Autoregression aims to filter out the contribution of distant activity. However, the LFP estimates obtained with ELECTRA should not be confused with the LFPs recorded with implanted microelectrodes in animals, since they differ considerably in terms of spatial localization. Direct LFP recordings in animals integrate neural activity in the submillimeter range, while ELECTRA estimates possess a spatial resolution that is at best comparable to that of intracranial recordings in epileptic patients.

Here we computed the ELECTRA inverse solution in a solution space composed of 4,024 nodes (referred to as voxels) homogeneously distributed within the inner compartment of a realistic head model (Montreal Neurological Institute average brain). The voxels were restricted to the gray matter and formed an isotropic grid of 6 mm resolution.

Power spectral density (PSD) estimation for each ELFP

The decoding of cognitive states can be attempted in the time or the spectral domain. Here we chose to perform the decoding in the spectral domain for the following reasons: 1) the temporal structure of actual field potentials (although not the amplitude) seems to be correctly estimated by a linear inverse method based on sensitive constraints [Grave de Peralta Menendez et al., 2004]; and 2) spectral estimates depend on the temporal structure of the LFPs and have proven to be efficient in decoding cognitive states in animals [Mehring et al., 2003; Pesaran et al., 2002].

In brief, for each individual subject the power spectral density (PSD) was computed for all brain voxels and single trials during the selected window using a multitaper method with seven Slepian data tapers. The multitaper method proposed by Thomson [1982] provides a trade-off between minimizing the variance of the estimate and maximizing the spectro-temporal resolution. The application of tapers to the data allows an estimation of power that is robust against bias. This is particularly important for time series with a

large dynamic range. Hence, for the $T = 500$ ms windows used here, a bandwidth parameter of $W = 8$ Hz and a variance reduction by a factor of $1/7$ was attained by using seven Slepian data tapers. Each 500-ms time series was multiplied by each of the tapers and the Fourier components were then computed via FFT. The PSD was computed by taking the square of the modulus of these complex numbers corresponding to frequencies from 0 (DC) to 100 Hz.

Feature selections using the first half of the trials (learning set)

Two cognitive states can be differentiated in the spectral domain if their power spectral densities do not completely overlap over trials for at least some voxels at a given frequency. Thus, a first step before trying to classify the cognitive states is to identify which voxels and frequencies (features) provide the maximal discriminative power between categories. The use of large feature sets leads to intractable computational problems and degrades the performance of the classifier if the features are redundant. Note that in our case the data size under scrutiny for each subject is about 65,027,840 real numbers, i.e., the product number of brain voxels \times number of considered frequencies \times number of trials \times number of categories.

In order to select the most discriminative features, the whole set of trials was divided into two parts: 1) A feature selection set composed of the first half of trials used to detect the most discriminative features; and 2) A validation set formed by the second half of the trials on which we inferred the subject's cognitive state using features extracted from the selection set.

Many approaches to feature selection have been developed [Pal and Mitra, 2004]. These approaches can be roughly divided into two types: filter and wrapper. Filter methods select the best features over the training set, independently of the classification algorithm or its error criteria. The idea of filter methods is to rank the features according to some measure of their capability to separate the classes that are being considered. A first ranking of the features can be obtained from the probability values obtained after comparing the classes statistically. However, statistical tests rely on central tendency measures and thus have a propensity to assign highly significant values to features that considerably overlap between classes. Thus, to select voxels and frequencies producing significant differences between cognitive states in the feature selection set, we employed here the discriminative power (DP) introduced in Gonzalez et al. [2006]. The discriminative power is a measure graded from 0 to 100, with zero representing complete overlap between classes (no discrimination is possible) and 100 representing perfect separation. It quantifies the percentage of trials that can be readily attributed to one of the classes based on a single feature, i.e., one frequency at a single voxel. It thus estimates the minimum correct classification rate that could be obtained using a single feature. As shown below, the use of multiple (independent) features yields substantial increases in classifi-

cation rates. As with the relief method [Kira and Rendell, 1992] or the Fisher criterion [Bishop, 1995], the DP discards possible interactions (redundancies) with other variables/features. However, in contrast to these methods the DP yields absolute bounded values that can be used to compare different features. Nonetheless, the DP measure has some limitations that are shared with other statistical methods based on the distribution of extreme values, namely: 1) sensitivity to outliers, and 2) possible failure for unbounded support distributions (taking values in the set $[-\infty, \infty]$).

Let's denote by $a(b)$ a feature vector for condition $A(B)$, i.e., a vector formed by the PSD over all trials for a single frequency in condition $A(B)$. By conveniently swapping vectors a and b we can always assume that $amin = [\text{minimum of } a] \leq bmin = [\text{minimum of } b]$. The DP is defined as the capacity of the features vectors to distinguish the conditions A and B and is defined as:

$$DP = \frac{\text{card}\{a < bmin\} + \text{card}\{b > amax\}}{\text{card}\{a\} + \text{card}\{b\}} * 100$$

where $\text{card}\{\cdot\}$ stands for the number of elements in a set.

Under this definition, the DP denotes the percentage of trials of conditions A and B that will be correctly identified using as a separator the lines at the minimum value for class b and the maximum for class a . All the values lower than $bmin$ obviously belong to class A . Similarly, all values greater than $amax$ belong to class B . If the two maxima coincide then one class contains the other.

Through this analysis strategy we explicitly avoided choosing the brain voxels where we expected consistent modulation of OA with stimulus category. Instead, we used the discriminative power for feature selection and evaluated a posteriori the neurophysiological plausibility of the brain areas identified. Thus, based on the feature selection set, we selected for each subject the 150 most discriminative features sorted by their discriminative power independently of their spatial location or frequency range. Consequently, this procedure allows two types of images to be created: 1) spectral plots showing oscillations that strongly discriminate between stimulus classes (DP vs. frequency); and 2) spatial plots showing the brain voxels that better discriminate between classes (DP vs. voxel position). Subsequently, we evaluated the best 150 features over the validation set to categorize subjects' cognitive states.

Categorization of the cognitive state over the second half of the trials (validation set)

The decoding of cognitive states was based on linear supported vector machines [Vapnik, 1995]. A support vector machine (SVM) is a supervised learning algorithm that addresses the general problem of classification by constructing hyperplanes in a multidimensional space. We employed the Matlab toolbox OSU-SVM (Matlab toolbox OSU-SVM 3.0, available at http://www.eleceng.ohio-state.edu/?maj/osu_svm). To evaluate the possibilities of simultaneously decoding

multiple stimulus categories, we used a multiclass classification procedure based on the one-against-all decomposition [Ryan and Klautau, 2004]. Here the classification into four classes is decomposed into four binary problems, solved using supported vector machines. We also computed the classification results for all pairs of categories (binary classification). Binary classification was used to rule out the possibility of classification based on motor responses rather than on the subject's perceptual categorization. Since the index and middle fingers have different somatotopic cortical representations, it is important to exclude classification based on this irrelevant difference. This can be done by comparing classification values between categories using the same or different fingers for the responses. A second aspect that might influence classification is the amount of overlap between a possible motor component and the analysis window. We therefore compared the classification values corresponding to pairs of categories with similar mean reaction times (e.g., I vs. W) with categories producing very different mean reaction times (W vs. NW). If classification is influenced by differences in motor responses within the window, we would expect a much better binary classification in the latter case, where the difference in the amount of motor responses included is greater in comparison to the former case, where the amount of motor responses within the window should be comparable. Consequently, similar binary classifications for these pairs will be a strong indicator of actual classification of categorical decision.

A randomized cross-validation procedure was used in which decoding accuracy was defined as the percentage of correct classifications in the output of the SVM procedure (perfect = 100.0, random guessing = 25). Average accuracy was computed over 100 repetitions of a procedure in which 90% of the trials in the validation set (VS) were used to train the classifier. This classifier was then applied to the remaining 10% of the trials.

RESULTS

Discriminative Power of eLFP oscillations

Individual discriminative power plots revealed consistent patterns of OA over subjects that were modulated by the category of the visual stimuli. This information is summarized in Figure 2, which shows the average over subjects of the maximum discriminative power over all brain voxels. These plots provide information on the eLFP oscillations that best serve to decode the different categories in view of their task-related modulation. The NW-NI comparison showed the highest discriminative power over subjects with 30% of the samples decoded from a single frequency (6 Hz). However, we observed systematic peaks at different frequencies for the other categories, albeit with a lower decoding percentage. Specifically, we observed that all comparisons involving words showed a peak in the alpha band centered on 11 Hz. Moreover, all comparisons involving explicit categorization

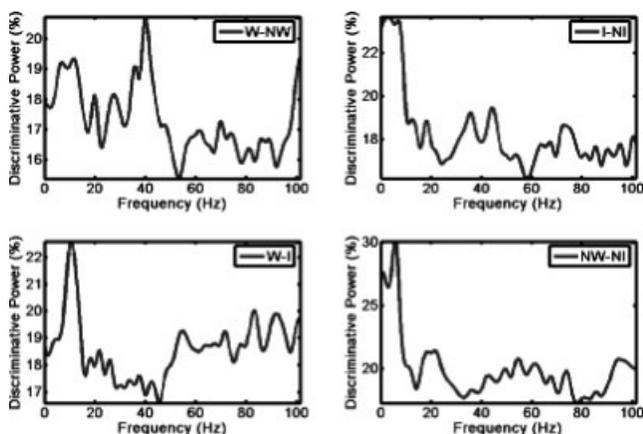


Figure 2.

Discriminative power (DP) as a function of frequency: plot of mean over subjects' DP as a function of frequency for all possible contrasts between pairs of visual stimuli. The DP is a measure between 0 and 100 indicating the amount of trials that can be correctly decoded using a single frequency and voxel in the brain. It therefore provides a minimal bound (classification results increase after combining multiple features) to decoding results.

of similar items (W-NW and I-NI) showed gamma band peaks (around 40 Hz).

Decoding the Subject's Cognitive States

Results for the multiclass classification are given in columns six to nine of Table I. Behavioral data, i.e., percentages of correct responses for each stimulus category, are given in the first four columns to facilitate comparison. The last four columns (10–13) of the table show, for each category, the ratio between the percentage of correctly decoded trials and the percentage of correct behavioral responses. This ratio is important since in this task two different types

of behavioral errors can occur: 1) erroneous categorization responses, and 2) incorrect manual responses (incorrect use of the responding finger).

Decoding results were above 90% for the W and NI categories for all subjects. In the I category, decoding results were lower than 90% in one single subject. The worst classification results were observed for the NW category, in which we also obtained the lowest rates of correct behavioral responses. Interestingly, the amount of correctly decoded trials surpasses, in some cases, the amount of correct responses as indicated by their ratio. This effect was more pronounced for subjects prone to commit behavioral errors in all categories (even those recognized easily) and is likely to be due to incorrect manual responses rather than categorization errors. As expected in a genuine "mind reading" procedure, the OA derived from eLFP was more accurate and less influenced by (voluntary or involuntary) subject errors.

An additional experiment was carried out to confirm that decoding results were independent of the finger used for the motor response. We selected pairs of categories in which the responding finger was the same, e.g., W-I and NW-NI, and we compared their binary classification rates with those of categories where different fingers were used, e.g., I-NI, W-NW. Paired classification results are given in Table II. Decoding rates for binary classification were better than for multiclass classification and totally independent of whether the same or different fingers were used for the response. A classification based on motor responses would predict better results for categories involving different fingers and worse results for identical fingers. In fact, we observed the opposite effect. The worst classification was obtained for the W-NW comparison, involving different fingers, while the best overall classification was obtained for the NW-NI condition that required identical finger responses. Of particular interest is the fact that, on average, the classification results for W-I were identical to those for W-NW. This occurred despite the comparable amount of motor responses included within the period of analysis for the W-I condition, which showed

TABLE I. Individual behavioral (C) and decoding results (D) for the four categories of visual stimuli

Subject	Behavior (% correct responses)				Decoding (% correct classification)				Ratio decoding/behavioral			
	CW	CNW	CI	CNI	DW	DNW	DI	DNI	RW	RNW	RI	RNI
1	95	70	100	100	94	81	98	98	0.99	1.16	0.98	0.98
2	95	47.5	82.5	90	95	78	96	96	1	1.64	1.16	1.06
3	95	82.5	87.5	97.5	96	93	96	97	1.01	1.13	1.09	0.99
4	95	80	92.5	85	96	91	98	98	1.01	1.13	1.05	1.15
5	100	72.5	97.5	97.5	98	74	97	98	0.98	1.02	0.99	1.00
6	97.5	97.5	85	100	97	96	85	98	0.99	0.98	1	0.98
7	95	67.5	90	100	95	74	92	96	1	1.09	1.02	0.96
8	95	90	100	97.5	93	91	96	96	0.98	1.01	0.96	0.98
9	97.5	87.5	92.5	100	97	86	94	98	0.99	0.98	1.01	0.98

The first four columns indicate the percentage of correct manual responses obtained for each subject during EEG recordings. The middle four columns give the percentage of correct classifications of the four different categories using the mind-reading approach described in the text. Classification is obtained for a time window of 500 ms length using the power spectral density (PSD) at the 150 features detected using the discriminative power measure. Last four columns give the ratio between the percentage of correct responses and the percentage of correctly decoded trials. Chance level in decoding is 25%.

TABLE II. Decoding results for pairs of visual stimuli

Subject	W-NW	I-NI	NW-NI	W-I
1	93.4	93.9	95.1	97.9
2	95.8	95.6	96.5	96.4
3	97.1	96.2	96.2	92.3
4	96	93.9	98.3	94.3
5	92.8	96.8	98.6	95.1
6	93.8	94.6	97.3	94.7
7	95.3	93.8	96	94.2
8	91.6	96.1	87.9	94
9	96.8	95.6	97.3	93.8
Mean	94.7	95.2	95.9	94.7

Values represent the percentage of trials attributed to the correct category in the paired classification. Classification is obtained using a linear support vector machine (SVM) over EEG windows of 500 ms length. Chance level is 50%.

similar reaction times. In contrast, in the W-NW comparison a larger proportion of responses were included for W than for NW (due to slower reaction times in the NW condition). Therefore, the similarity in classification values renders very unlikely an explanation in terms of the classification being due to a greater or lesser amount of motor responses. Furthermore, the best binary classification results were obtained for the two stimulus categories that induced the slowest responses (i.e., NW-NI) and in which the analysis window contained a smaller amount of motor responses. The results of the binary classification strongly support our suggestion that the technique is able to decode the content of visual perception without any influence of motor responses.

Discriminative Power of Different Brain Areas

The decoding results indicate that the temporal structure of the estimated ELFPs allows the visual stimuli to be reliably categorized. At the anatomic level, it is reasonable to expect that areas where OA modulation gave rise to higher rates of discrimination should coincide with the brain networks traditionally involved in the processing of such stimuli. To test this hypothesis, we constructed images where we assigned a value of 1 to voxels and frequencies belonging to the best 150 features used for classification. Averaging over all frequencies in the 0–100 Hz range resulted in individual images summarizing the contribution of each voxel to the discrimination between pairs of categories. Voxels that appear repeatedly (i.e., the same voxel that discriminates at more than one frequency) were considered only once. An example of this analysis is shown in Figure 3, which illustrates the individual images for the W-I comparison. With reference to major brain regions involved in language and reading (schematically drawn in the lower right panel), our analysis shows that, despite some interindividual variability, the best 150 voxels cluster together around regions that coincide well with classical language areas.

Paired categories were used to allow comparison with other functional neuroimaging techniques (e.g., fMRI) that usually rely on paired stimuli contrasts. To summarize infor-

mation over subjects we summed the individual images to code in color the amount of subjects displaying the specific effect at each area. Figures 4, 5, and 6 illustrate the results for the different comparisons (respectively, W with I, I with NI, and W with NW). In all panels the results are surface-rendered onto a canonical brain using the MRIcro software. A positive value was assigned if the average over trials of the OA was higher for the first category and a negative value when the mean spectral power was higher for the second.

Figure 4 shows brain voxels that best discriminated between W and I. Figure 4a shows the brain voxels where W elicited more activity than I and 4b the voxels where I elicited stronger activity than W. A left lateralization of discriminative voxels for W and right lateralization for I is observed. Most discriminative voxels for W are contained within regions that participate in the largely distributed language network [Fiez and Petersen, 1998; Turkeltaub et al., 2003; Vandenberghe et al., 1996] including the left superior temporal and middle temporal lobe, the inferior frontal operculum, the left angular, and the left supramarginal gyrus. A cluster of voxels was observed in the left fusiform gyrus (not shown on the cortical rendering). Best pixels for I are located at the occipital lobe and right middle frontal gyrus, with most dominant clusters located in bilateral parieto/occipital areas [Grill-Spector, 2003; Spiridon and Kanwisher, 2002]. Bilateral although somewhat right-dominant striate/extrastriate activation (not shown) was also observed.

Figure 5 shows the voxels that best discriminate between I and NI. Figure 5a shows the brain voxels where oscillations increased in response to images. Major differences were observed at occipito-parietal areas, with a strong cluster around the left angular gyrus and a second cluster in the inferior occipital gyrus [Grill-Spector et al., 1998; Tanaka, 1997]. A tendency toward a left hemisphere lateralization was observed in the occipital clusters, as well as the weak frontal clusters. The NI condition elicited oscillations (Fig. 5b) in the right and left lingual gyrus, the right superior temporal lobe, and the right parietal lobe. Globally, the number of discriminative voxels belonging to the NI category was smaller than for the I category, with little overlap over subjects.

The most discriminative voxels for the W-NW comparison are shown in Figure 6. When comparing NW to W (Fig. 6b), a bilateral enhancement of oscillations in striate (not shown) and extrastriate visual areas, as well as left occipito/temporal areas was observed. Words (Fig. 6a) induced a widespread increase in oscillations within the left temporal lobe (middle and inferior) and the inferior frontal opercularis area [Booth et al., 2004; Fiez and Petersen 1998; Simos et al., 2002]. Clusters of voxels were also observed in the left insula and the left hippocampus (not shown). Note that Figure 6a shares some resemblance with Figure 4a, which depicts the words vs. images comparison. However, the strong cluster in the left angular gyrus that appears in the W-I contrast is absent in the W-NW contrast.

The temporal structure of ELFP was modulated in a replicable manner over subjects at brain voxels known to partici-

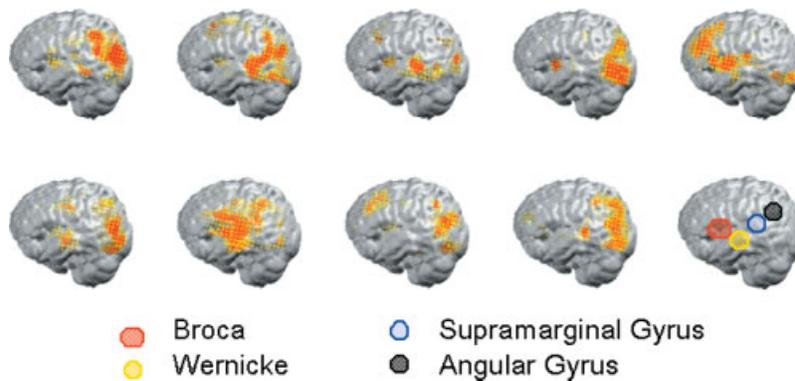


Figure 3.

Individual images showing the anatomical location of the best 150 voxels used for classification in the W-I condition. To construct the image, we assigned a value of one to voxels and frequencies belonging to the best 150 features used for classification and averaged over the whole range of 0–100 Hz. The last right-most panel

shows a schematic drawing of the major brain regions linked to language and reading. Note that despite some interindividual variability, the best 150 voxels cluster together around regions that coincide well with classical language areas.

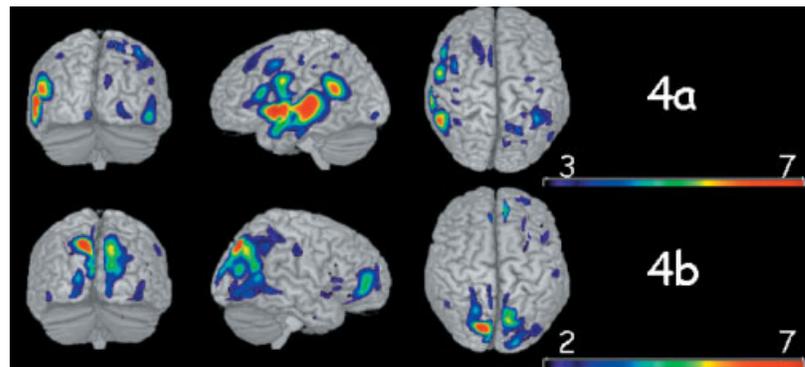


Figure 4.

Brain voxels that better discriminate between words (W) and images (I). **a:** Voxels where words elicited more oscillatory activity than images. **b:** Voxels where images elicited stronger oscillatory activity than words.

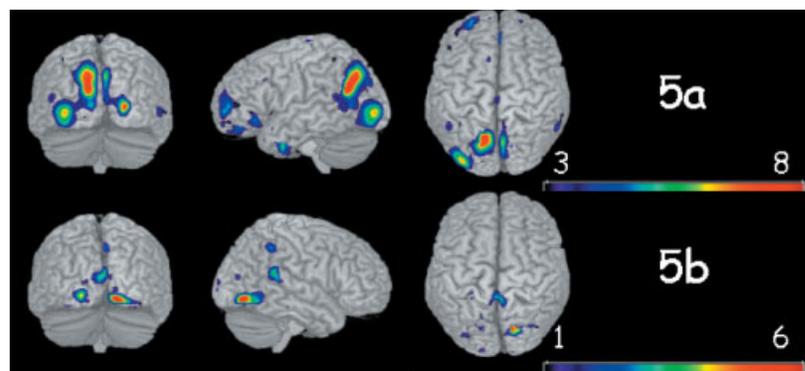


Figure 5.

Brain voxels that better discriminate between images and nonimages. **a:** Voxels where images elicited more oscillations than nonimages. **b:** Voxels where nonimages elicited more oscillations than images.

participate in the processing of these visual stimuli. More specifically, words and word-like recognition modulated oscillations to a greater extent within the left hemisphere language-related regions, while image recognition involved more bilateral striate/extrastriate regions. However, it is important to emphasize that these brain areas emerged here be-

cause of their importance in classifying the categories, and were thus not established a priori.

These results therefore confirm that the processing of visual stimuli elicits changes in the strength of neural oscillations in brain areas participating in the task. Moreover, modulation of OA is a physiologically reliable phenomenon

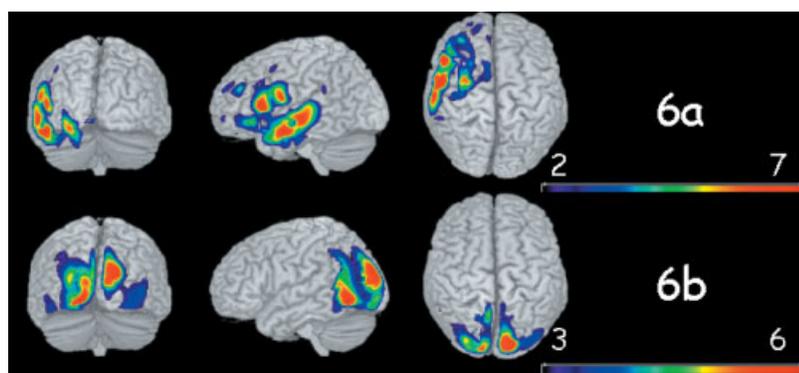


Figure 6.

Brain voxels that better discriminate between words (W) and scrambled words (NW). **a:** Voxels where words elicited more oscillations than nonwords. **b:** Voxels where nonwords elicited more oscillations than words.

that remains stable over trials and allows accurate decoding of cognitive states over subjects as identified by the large overlap observed in the most discriminative voxels.

DISCUSSION

We have shown that short EEG epochs of half a second can be transformed into intracranial LFP estimates that contain consistent information about the relative amplitude of neural oscillations elicited at different brain areas. In agreement with previous animal [Gray and Singer, 1989; Pesaran et al., 2002] and human experiments [Klimesch et al., 2004; Lachaux et al., 2005], we observed a stimulus sensitive modulation of eLFP oscillations that can be used efficiently to decode cognitive states. A number of different roles have been attributed to the neural oscillations occurring in different frequency bands. For example, gamma oscillations (30–90 Hz frequency band) have been linked to attention [Fries et al., 2001], to expectations and predictions about upcoming sensory stimuli [Engel et al., 2001], to binding of distributed representations of perceptual objects [Engel et al., 1991; Gray et al., 1990; Malsburg and Schneider, 1986], and to episodic [Sederberg et al., 2003] and working [Pesaran et al., 2002] memory. Likewise, alpha, theta, or even very high frequency oscillations over 100 Hz could encode different aspects of sensory-motor processes [Edwards et al., 2005]. While our results contribute little to the thorough understanding of the functional role of oscillatory activity, they are nonetheless consistent with some of the existing literature addressing such issues. In particular, our analysis showed a high discriminatory power for the gamma band when contrasting meaningful and meaningless categories was required, i.e., words vs. nonwords and images vs. nonimages (Fig. 2). These results are consistent with previous EEG findings that report differences in gamma oscillations for W-NW comparisons [Pulvermüller, 1995] or figure background segregation [Gail et al., 2000; Tallon-Baudry and Bertrand, 1999]. While we cannot rule out that certain oscillations are epiphenomenal and serve no clear function, our results reveal that at least part of the

OA can lead to a high level of decoding accuracy, thereby arguing against such a hypothesis. The excellent accuracy of decoding substantiates the reproducibility of our single trial estimates of OA as a predictor of subjects' mental states. The fact that highly discriminating voxels occur across subjects in brain regions similar (in terms of laterality and localization) to those uncovered by other functional techniques demonstrates that this method can efficiently estimate the frequency content of eLFP within each brain area.

With the help of this novel neuroimaging technique, we have been able to reproduce the decoding accuracy over short temporal windows such as that classically obtained with invasive recordings in animals. Consequently, these results pave the way for promising research perspectives in the field of neuroprosthetics, where real-time decoding of subjects' intentions using a noninvasive technique is needed. Importantly, the analysis procedure proposed here can be easily implemented on-line. Indeed, since the ELECTRA inverse solution is linear, the noninvasive estimation of the eLFPs amounts to a simple product of matrices. The subsequent steps, comprising the PSD estimation and classification, remain those already in use for analysis of invasive data.

The transformation of scalp-recorded EEG into intracranial field potential estimates passes through the resolution of the electromagnetic inverse problem, which is known to lack a unique solution. Theoretical studies on the use of linear inverse solutions to the bioelectromagnetic inverse problem [Grave de Peralta-Menendez et al., 1998] point to the fact that the lack of a unique solution must give rise to an incorrect estimation of instantaneous source amplitudes. This difficulty is circumvented in the analysis presented here by relying on the temporal structure of eLFP (oscillations strength) rather than on instantaneous amplitude estimates. While spectral estimates also depend on the source strength, the analysis procedure used to decode cognitive states relies on the comparison between categories of PSD estimates at each voxel. Since the underestimation of the source strength depends mainly on the distance between the source and the sensors, this factor is similar for the same voxel across stimulus categories. Therefore, the PSD differences between conditions should be mainly due to the spectral modulation produced by the different stimulus types.

Three reasons explain why the decoding accuracy compares well with the one obtained from invasive methods in animals. First, decoding of mental states in animals relies on oscillatory activity rather than on LFP amplitudes. Second, linear inverse solutions seem to provide adequate estimates of the temporal structure of the intracranial field potentials. Third, the inversion procedure incorporates information absent in the original EEG regarding 1) the mixing matrix (lead field), and 2) the source model and the intracranial field propagation laws (inverse solution).

Concerning the 500-ms time window used in this study, one might ask whether smaller periods could be used without losing decoding power. The answer obviously depends on both perceptual and methodological aspects. From the perceptual point of view, we require analysis windows, which are long enough to include the subject perceptual categorization. Considering that ERP differences between W and NW occur at around 300 ms and between I and NI at around 250 ms [Khateb et al., 2002], it can be argued that epochs of 300 ms would comprise processes of perceptual categorization and thus could still yield a comparable level of decoding accuracy. From the methodological point of view, we must take into account the fact that the use of short analysis windows produce a loss of spectral resolution, thus constituting a shortcoming if sharp frequency tuning is used to encode information about the stimulus. However, electrophysiological studies in animals and humans suggest that LFP tuning takes place over frequency bands rather than discrete frequencies. Therefore, the minimal length of the analysis window seems to be constrained by the speed of human perception and information processing rather than by methodological issues. The study of Gonzalez et al. [2006] supports this view by showing that decoding of the laterality of impending hand movements can be carried out on the basis of temporal windows of a duration shorter than 200 ms.

The methodology described here could have applications that extend beyond the problem of "mind reading." Indeed, it provides simultaneous information about: 1) the range of frequencies where OA maximally differ between pairs of experimental conditions as measured by the discriminative power; 2) the brain areas where OA differ over subjects using individual 3D images (although these images are of lower spatial resolution than those produced by fMRI, they have the advantage of reflecting the neural origin of oscillatory activity that cannot be studied by hemodynamic techniques); and 3) the consistency of the OA (evaluated by the classification) across single trials, which would approach chance if areas and oscillations were incorrectly estimated. Accordingly, this procedure exploits the full spectral information contained in single trials to construct a single statistical image of OA for each individual subject, thus allowing single subject (or patient) analysis. It therefore addresses a relevant point in current neuroscience: going from mere localization of function to actual brain functioning, which is inherently based on single stimuli processing rather than averages [Makeig et al., 2002].

ACKNOWLEDGMENTS

This article reflects the authors' views only and funding agencies are not liable for any use that may be made of the information contained herein.

REFERENCES

- Bishop CM (1995): Neural networks for pattern recognition. Oxford: Clarendon Press.
- Booth JR, Burman DD, Meyer JR, Gitelman DR, Parrish TB, Mesulam MM (2004): Development of brain mechanisms for processing orthographic and phonologic representations. *J Cogn Neurosci* 16: 1234–1249.
- Buckner RL, Bandettini PA, O'Craven KM, Savoy RL, Petersen SE, Raichle ME, Rosen BR (1996): Detection of cortical activation during averaged single trials of a cognitive task using functional magnetic resonance imaging. *Proc Natl Acad Sci U S A* 93: 14878–14883.
- Edwards E, Soltani M, Deouell LY, Berger MS, Knight RT (2005): High gamma activity in response to deviant auditory stimuli recorded directly from human cortex. *J Neurophysiol* 94:4269–4280.
- Engel AK, Kreiter AK, Konig P, Singer W (1991): Synchronization of oscillatory neuronal responses between striate and extrastriate visual cortical areas of the cat. *Proc Natl Acad Sci U S A* 88: 6048–6052.
- Engel AK, Fries P, Singer W (2001): Dynamic predictions: oscillations and synchrony in top-down processing. *Nat Rev Neurosci* 2:704–716.
- Fiez JA, Petersen SE (1998): Neuroimaging studies of word reading. *Proc Natl Acad Sci U S A* 95:914–921.
- Fries P, Neuenschwander S, Engel AK, Goebel R, Singer W (2001): Rapid feature selective neuronal synchronization through correlated latency shifting. *Nat Neurosci* 4:194–200.
- Gail A, Brinksmeier HJ, Eckhorn R (2000): Contour decouples gamma activity across texture representation in monkey striate cortex. *Cereb Cortex* 10:840–850.
- Gervasoni D, Lin SC, Ribeiro S, Soares ES, Pantoja J, Nicolelis MA (2004): Global forebrain dynamics predict rat behavioral states and their transitions. *J Neurosci* 24:11137–11147.
- Gonzalez Andino SL, Grave de Peralta Menendez R, Lantz CM, Blank O, Michel CM, Landis T (2001): Non-stationary distributed source approximation: an alternative to improve localization procedures. *Hum Brain Mapp* 14:81–95.
- Gonzalez Andino SL, Michel CM, Thut G, Landis T, Grave de Peralta R (2005): Prediction of response speed by anticipatory high-frequency (gamma band) oscillations in the human brain. *Hum Brain Mapp* 24:50–58.
- Gonzalez SL, Grave de Peralta R, Thut G, Millan JD, Morier P, Landis T (2006): Very high frequency oscillations (VHFO) as a predictor of movement intentions. *Neuroimage* (in press).
- Grave de Peralta-Menendez R, Gonzalez-Andino, Sara L (1998): A critical analysis of linear inverse solutions to the neuroelectromagnetic inverse problem. *IEEE Trans Biomed Eng* 45:440–448.
- Grave de Peralta Menendez R, Gonzalez Andino SL, Morand S, Michel CM, Landis T (2000): Imaging the electrical activity of the brain: ELECTRA. *Hum Brain Mapp* 9:1–12.
- Grave de Peralta Menendez R, Murray MM, Michel CM, Martuzzi R, Gonzalez Andino SL (2004): Electrical neuroimaging based on biophysical constraints. *Neuroimage* 21:527–539.
- Gray CM, Singer W (1989): Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Proc Natl Acad Sci U S A* 86:1698–1702.

- Gray CM, Engel AK, Konig P, Singer W (1990): Stimulus-dependent neuronal oscillations in cat visual cortex: receptive field properties and feature dependence. *Eur J Neurosci* 2:607–619.
- Grill-Spector K (2003): The neural basis of object perception. *Curr Opin Neurobiol* 13:159–166.
- Grill-Spector K, Kushnir T, Hendler T, Edelman S, Itzhak Y, Malach R (1998): A sequence of object-processing stages revealed by fMRI in the human occipital lobe. *Hum Brain Mapp* 6:316–328.
- Kamitani Y, Tong F (2005): Decoding the visual and subjective contents of the human brain. *Nat Neurosci* 8:679–685.
- Khateb A, Michel CM, Pegna AJ, Landis T, Annoni JM (2000): New insights into the Stroop effect: a spatio-temporal analysis of electric brain activity. *Neuroreport* 11:1849–1855.
- Khateb A, Pegna AJ, Michel CM, Landis T, Annoni JM (2002): Dynamics of brain activation during an explicit word and image recognition task: an electrophysiological study. *Brain Topogr* 14: 197–213.
- Kira K, Rendell LA (1992): A practical approach to feature selection. San Mateo, CA: Morgan Kaufmann.
- Klimesch W, Schack B, Schabus M, Doppelmayr M, Gruber W, Sauseng P (2004): Phase-locked alpha and theta oscillations generate the P1-N1 complex and are related to memory performance. *Brain Res Cogn Brain Res* 19:302–316.
- Lachaux JP, George N, Tallon-Baudry C, Martinerie J, Hugueville L, Minotti L, Kahane P, Renault B (2005): The many faces of the gamma band response to complex visual stimuli. *Neuroimage* 25: 491–501.
- Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A (2001): Neurophysiological investigation of the basis of the fMRI signal. *Nature* 412:150–157.
- Makeig S, Westerfield M, Jung TP, Enghoff S, Townsend J, Courchesne E, Sejnowski TJ (2002): Dynamic brain sources of visual evoked responses. *Science* 295:690–694.
- Malsburg C, Schneider W (1986): A neural cocktail-party processor. *Biol Cybern* 54:29–40.
- Mehring C, Rickert J, Vaadia E, Cardoso de Oliveira S, Aertsen A, Rotter S (2003): Inference of hand movements from local field potentials in monkey motor cortex. *Nat Neurosci* 6:1253–1254.
- Menon RS, Kim SG (1999): Spatial and temporal limits in cognitive neuroimaging with fMRI. *Trends Cogn Sci* 3:207–216.
- Niessing J, Ebisch B, Schmidt KE, Niessing M, Singer W, Galuske RA (2005): Hemodynamic signals correlate tightly with synchronized gamma oscillations. *Science* 309:948–951.
- Pal SK, Mitra P (2004): Pattern recognition algorithms for data mining: scalability, knowledge discovery and soft granular computing. Boca Raton, FL: Chapman & Hall/CRC.
- Pesaran B, Pezaris JS, Sahani M, Mitra PP, Andersen RA (2002): Temporal structure in neuronal activity during working memory in macaque parietal cortex. *Nat Neurosci* 5:805–811.
- Pulvermüller F, Lutzenberger W, Preissl H, Birbaumer N (1995): Spectral responses in the gamma-band: physiological signs of higher cognitive processes? *Neuroreport* 6:2059–2064.
- Ryan R, Klautau A (2004): In defense of one-vs-all classification. *J Mach Learn Res* 5:101–141.
- Sederberg PB, Kahana MJ, Howard MW, Donner EJ, Madsen JR (2003): Theta and gamma oscillations during encoding predict subsequent recall. *J Neurosci* 23:10809–10814.
- Simos PG, Breier JL, Fletcher JM, Foorman BR, Castillo EM, Papanicolaou AC (2002): Brain mechanisms for reading words and pseudowords: an integrated approach. *Cereb Cortex* 12:297–305.
- Spiridon M, Kanwisher N (2002): How distributed is visual category information in human occipito-temporal cortex? An fMRI study. *Neuron* 35:1157–1165.
- Tallon-Baudry C, Bertrand O (1999): Oscillatory gamma activity in humans and its role in object representation. *Trends Cogn Sci* 3: 151–162.
- Tanaka K (1997): Mechanisms of visual object recognition: monkey and human studies. *Curr Opin Neurobiol* 7:523–529.
- Thomson DJ (1982): Spectrum estimation and harmonic analysis. *Proc IEEE* 70:1055–1096.
- Thut G, Hauert CA, Blanke O, Morand S, Seeck M, Gonzalez SL, Grave de Peralta R, Spinelli L, Khateb A, Landis T et al. (2000): Visually induced activity in human frontal motor areas during simple visuomotor performance. *Neuroreport* 11:2843–2848.
- Turkeltaub PE, Gareau L, Flowers DL, Zeffiro TA, Eden GF (2003): Development of neural mechanisms for reading. *Nat Neurosci* 6: 767–773.
- Vandenbergh R, Price C, Wise R, Josephs O, Frackowiak RS (1996): Functional anatomy of a common semantic system for words and pictures. *Nature* 383:254–256.
- Vapnik VN (1995): The nature of statistical learning theory. New York: Springer.
- Wessberg J, Stambaugh CR, Kralik JD, Beck PD, Laubach M, Chapin JK, Kim J, Biggs SJ, Srinivasan MA, Nicolelis MA (2000): Real-time prediction of hand trajectory by ensembles of cortical neurons in primates. *Nature* 408:361–365.