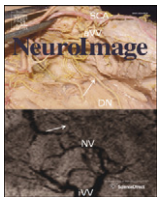
NeuroImage 63 (2012) 1162–1170

Contents lists available at SciVerse ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg



Decoding the perception of pain from fMRI using multivariate pattern analysis

Kay H. Brodersen [a,b,c,*,1], Katja Wiech [a,**,1], Ekaterina I. Lomakina [b,c], Chia-shu Lin [a], Joachim M. Buhmann [c], Ulrike Bingel [a,d], Markus Ploner [a,e], Klaas Enno Stephan [b,f], Irene Tracey [a]

[a] Centre for Functional Magnetic Resonance Imaging of the Brain (FMRIB), Nuffield Department of Clinical Neurosciences, Nuffield Division Anaesthetics, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, UK
[b] Translational Neuromodeling Unit (TNU), Institute for Biomedical Engineering, University of Zurich & ETH Zurich, Wilfriedstrasse 6, CH 8032 Zurich, Switzerland
[c] Department of Computer Science, ETH Zurich, Universitaetstrasse 6, CH 8092 Zurich, Switzerland
[d] NeuroImage Nord, Department of Neurology, University Medical Center Hamburg-Eppendorf, Martinistr. 52, 20246 Hamburg, Germany
[e] Department of Neurology, Technische Universität München, 81675 Munich, Germany
[f] Wellcome Trust Centre for Neuroimaging, University College London, UK

ARTICLE INFO

Article history:
Accepted 13 August 2012
Available online 18 August 2012

Keywords:
Pain
Decoding
Support vector machine
Permutation test
Classification accuracy

ABSTRACT

Pain is known to comprise sensory, cognitive, and affective aspects. Despite numerous previous fMRI studies, however, it remains open which spatial distribution of activity is sufficient to encode whether a stimulus is perceived as painful or not. In this study, we analyzed fMRI data from a perceptual decision-making task in which participants were exposed to near-threshold laser pulses. Using multivariate analyses on different spatial scales, we investigated the predictive capacity of fMRI data for decoding whether a stimulus had been perceived as painful. Our analysis yielded a rank order of brain regions: during pain anticipation, activity in the periaqueductal gray (PAG) and orbitofrontal cortex (OFC) afforded the most accurate trial-by-trial discrimination between painful and non-painful experiences; whereas during the actual stimulation, primary and secondary somatosensory cortex, anterior insula, dorsolateral and ventrolateral prefrontal cortex, and OFC were most discriminative. The most accurate prediction of pain perception from the stimulation period, however, was enabled by the combined activity in pain regions commonly referred to as the 'pain matrix'. Our results demonstrate that the neural representation of (near-threshold) pain is spatially distributed and can be best described at an intermediate spatial scale. In addition to its utility in establishing structure-function mappings, our approach affords trial-by-trial predictions and thus represents a step towards the goal of establishing an objective neuronal marker of pain perception.

Introduction

The perception of pain is a multi-factorial experience that comprises sensory, cognitive, and affective aspects. Accordingly, pain is thought to result from a complex interplay between many regions in the human brain, including the thalamus, insula, primary and secondary somatosensory, anterior cingulate cortex, and prefrontal cortex (Apkarian et al., 2005). The specific characteristics of regions underlying the perception of pain have been described in some detail using conventional univariate analysis methods for functional magnetic resonance imaging (fMRI). By contrast, there have been almost no attempts at examining the distributed representation of pain and how it is encoded jointly by activity within and across the set of regions commonly associated with pain.

Statistical methods for examining distributed coding schemes have undergone rapid progress over the past years. One particularly versatile approach, termed multivariate pattern analysis (MVPA), is based on the use of a classification algorithm to infer a perceptual or cognitive state from brain activity. The underlying multivariate decoding models differ in important ways from univariate encoding models, such as the general linear model (GLM). Univariate analyses have proven powerful for inference on structure-function mappings in the brain when activations are expressed in terms of local peaks or clusters of activity (Friston et al., 1995). However, they are less suitable for assessing the amount of information encoded in spatially distributed (multivoxel) patterns of activity underlying specific perceptual or cognitive states. This information can be estimated using multivariate decoding models (Friston et al., 2008; Haynes and Rees, 2006; Norman et al., 2006; O'Toole et al., 2007; Pereira et al., 2009). These models consider several voxels at the same time and may therefore be more sensitive than univariate models (for an analysis of the conditions under which this is the case, see Guyon and Elisseeff, 2003).

* Correspondence to: K.H. Brodersen, Department of Computer Science, ETH Zurich, Universitaetstrasse 6, CH 8092 Zurich, Switzerland. Fax: +41 4463 44 907.
** Correspondence to: K. Wiech, Centre for Functional Magnetic Resonance Imaging of the Brain (FMRIB), Nuffield Department of Clinical Neurosciences, Nuffield Division Anaesthetics, University of Oxford, Oxford, OX3 9DU, UK. Fax: +44 1865 234541.
E-mail addresses: kay.brodersen@inf.ethz.ch (K.H. Brodersen), kwiech@fmrib.ox.ac.uk (K. Wiech).
[1] These authors contributed equally.

1053-8119/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.neuroimage.2012.08.035

Decoding approaches are typically implemented in the form of classification algorithms. The results of such algorithms are often reported in terms of classification accuracies. It is important to remember, however, that in cognitive neuroscience the absolute accuracy is not of primary interest if, as we do here, one wishes to demonstrate the existence of a structure-function relationship in the brain, e.g., the relationship between measures of brain activity and a perceptual state (Friston et al., 2008). Such a relationship is evidenced by the *significance* with which the accuracy is above chance, not by its *magnitude*, since the significance takes into account both mean and variability in the group. This is different in engineering applications such as the design of brain-machine interfaces, where *substantive* significance, i.e., the magnitude of classification accuracy, is of interest. Thus, inferences in this paper are not based on accuracies but on the question of whether the reported accuracies are significantly above chance; similarly, comparisons are not based on absolute differences in accuracies but on the question of whether two accuracies differ significantly. We will revisit this distinction in the Discussion.

The utility of classification approaches has been demonstrated in many domains of systems neuroscience, but corresponding insights into the perception of pain have remained scarce. In one methodological study, the utility of Gaussian processes was illustrated using different levels of pain as well as graded responses to similar levels of pain (Marquand et al., 2010). Another technical study considered the temporal evolution of perception in response to prolonged noxious stimulation (Prato et al., 2011).

These studies have suggested that predicting pain from brain recordings may be feasible. However, it has remained unclear to what extent the extraction of pain-related information benefits from the simultaneous consideration of multiple brain regions. More specifically, it is not well understood which spatial scale[2] is optimal for decoding pain: individual voxels, single anatomical regions, combinations of regions, or whole-brain activity? Moreover, it is currently unknown what predictive capacity is enabled by those anatomical regions (and their combinations) that are typically associated with pain. Finally, there has been no investigation of pain encoding that assesses voxel-wise significances (e.g., *t*-scores) in a multivariate fashion.

In the present study, we addressed the above questions by analyzing the predictive capacity of individual and multiple brain regions in decoding the subjective experience of pain. Notably, we carry out this analysis in the setting of rather subtle (near-threshold) pain stimuli. This is challenging but important since decoding results may otherwise be dominated by physical differences in sensory stimulation rather than differences in subjective pain experience. First, we aimed to predict pain perception from whole-brain fMRI data on a trial-by-trial basis. Second, we examined which spatial level of description enabled the most accurate predictions of pain: single voxels, individual anatomical regions, combinations of regions, or whole-brain activity. For both questions, we trained and tested a linear support vector machine (SVM) on trial-specific correlates of whole-brain activity using a leave-one-session-out cross-validation scheme. Third, we evaluated SVM-based voxel weights with a permutation test to illustrate the spatial deployment of jointly informative voxels throughout the brain.

Methods

Participants

To study the multivariate nature of pain-related activity in the brain, we revisited a dataset that was originally analyzed using conventional univariate methods (Wiech et al., 2010). Here, we provide a summary of the underlying experimental design, focusing on those aspects that are relevant for the question addressed in the present paper. A group of 16 volunteers (age range 19–30 years, 11 females, all right-handed), with no history of neurological or psychiatric illnesses or chronic pain, participated in the study. All participants gave informed consent, and the study was approved by the local Research Ethics Committee.

Experimental design

Subjects were engaged in a sensory decision-making task consisting of carefully calibrated laser stimulation and an additional threat manipulation (Fig. 1). The experiment consisted of four sessions, each comprising 30 trials, totaling 120 trials per subject. On each trial, a near-threshold laser stimulus was applied to one out of six possible stimulation sites on the right foot. Following the laser pulse, participants were prompted to indicate by button press whether the stimulus had been perceived as painful or non-painful.

The design contained an additional factor which was of no interest in the present analysis, but whose details we briefly outline for completeness (see Wiech et al., 2010, for a full description). At three stimulation sites, participants were made to believe that the stimulation was safe and approved without reservations ('low threat' condition). At the remaining three sites, participants were told that the stimulation would still be performed but could only be approved with reservations, as a result of an assessment of skin properties prior to the experiment ('high threat' condition). Unknown to participants, the assignment of the six sites to the two conditions (low threat vs. high threat) was defined a priori and entirely unrelated to any actual skin properties. On each trial, a visual cue informed subjects whether

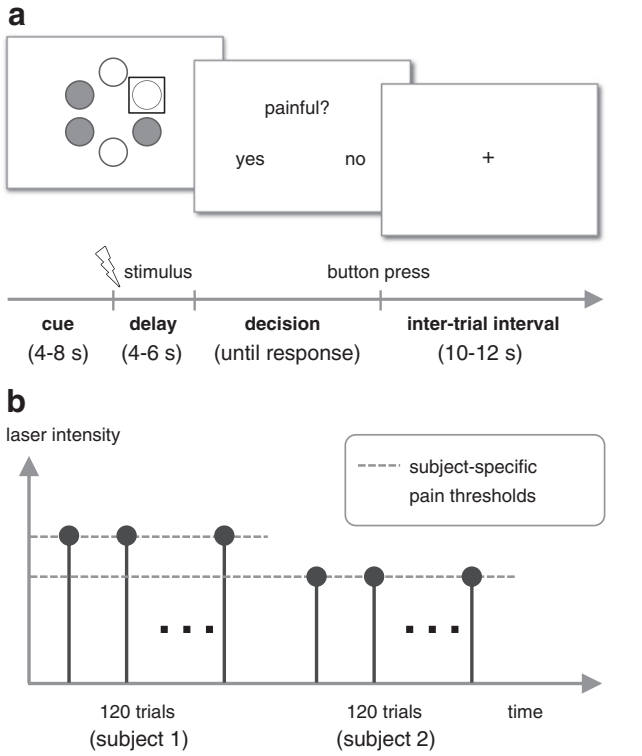


Fig. 1. Experimental design. Subjects were engaged in a simple perceptual decision-making task (Wiech et al., 2010). (a) At the beginning of each trial, a graphical representation of the 6 potential stimulation sites was shown before stimulus application. 'Fully approved' sites were shown in a different color than sites that were 'approved with reservations.' The site stimulated on the current trial was highlighted by a square. Following a brief laser stimulus, participants were prompted to indicate by a button press whether the stimulus had been perceived as painful (here: left button for 'pain', right button for 'no pain'). Assignment of buttons was randomized across all 120 trials. (b) Within each subject, the laser intensity was calibrated to match the individual pain threshold.

[2] It should be noted that the term 'scale' does not imply a physical scale parameter here but is used to refer to the spatial composition of the feature space, i.e., the size of the search volume used for classification.