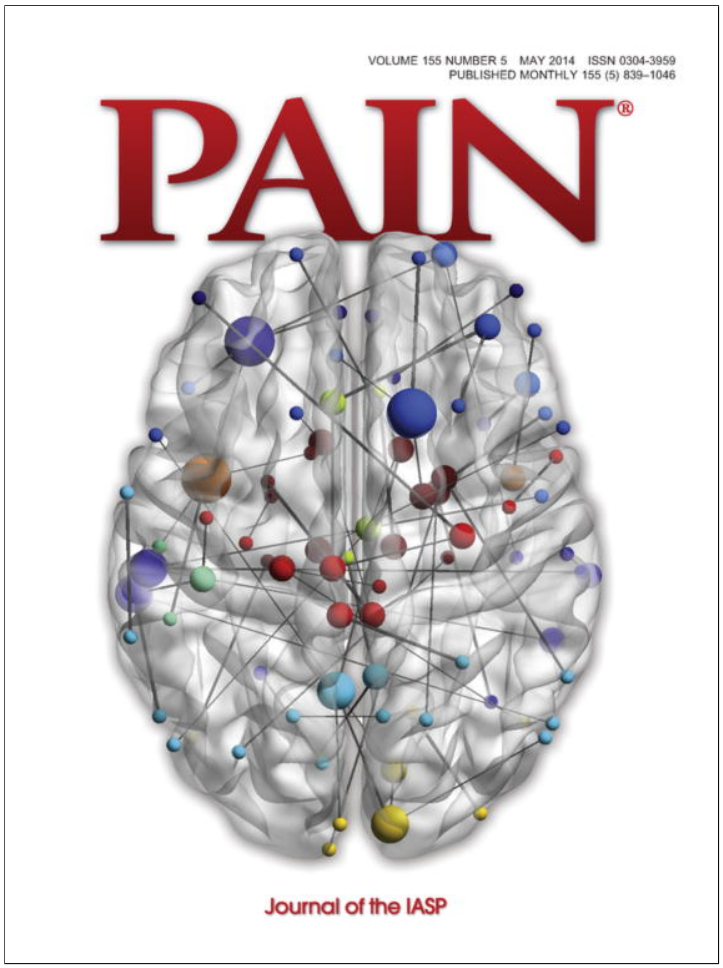

PAIN® 155 (2014) 994–1001

www.elsevier.com/locate/pain

Parametric trial-by-trial prediction of pain by easily available physiological measures

Stephan Geuter [a,*], Matthias Gamer [a], Selim Onat [a], Christian Büchel [a,b]

[a] Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, Hamburg 20246, Germany
[b] Department of Psychology, Stanford University, Stanford, CA 94305, USA

Sponsorships or competing interests that may be relevant to content are disclosed at the end of this article.

ARTICLE INFO

Article history:
Received 6 November 2013
Received in revised form 29 January 2014
Accepted 5 February 2014

Keywords:
Pain measurement
Skin conductance
Pupil dilation
Principal component regression

ABSTRACT

Pain is commonly assessed by subjective reports on rating scales. However, in many experimental and clinical settings, an additional, objective indicator of pain is desirable. In order to identify an objective, parametric signature of pain intensity that is predictive at the individual stimulus level across subjects, we recorded skin conductance and pupil diameter responses to heat pain stimuli of different durations and temperatures in 34 healthy subjects. The temporal profiles of trial-wise physiological responses were characterized by component scores obtained from principal component analysis. These component scores were then used as predictors in a linear regression analysis, resulting in accurate pain predictions for individual trials. Using the temporal information encoded in the principal component scores explained the data better than prediction by a single summary statistic (ie, maximum amplitude). These results indicate that perceived pain is best reflected by the temporal dynamics of autonomic responses. Application of the regression model to an independent data set of 20 subjects resulted in a very good prediction of the pain ratings demonstrating the generalizability of the identified temporal pattern. Utilizing the readily available temporal information from skin conductance and pupil diameter responses thus allows parametric prediction of pain in human subjects.

1. Introduction

Pain is a subjective experience and is therefore assessed by subjective reports, which are commonly mapped to a numerical rating scale or a visual analogue scale [29]. However, in many circumstances an objective pain assessment is important. For example, there are situations in which a patient is not able to understand the rating scales (eg, in children), or the report may be biased in other scenarios. The latter is relevant, because subjects may form expectations about the desired study outcome. This might actually be the case in most of the studies on pain, for example, when testing a drug or evaluating some behavioral intervention.

Candidates for an objective auxiliary measure of pain are autonomic nervous system responses that are related to pain perception [9,10], for example, skin conductance [10,18,22,23,34] and pupil diameter [8,10,13,16,27]. Changes in skin conductance levels correlate with pain ratings of heat pain stimuli [23], and pupil dilation amplitudes reflect electric stimulation intensity [8].

Studies investigating these measures often characterize physiological responses by a single summary statistic (eg, maximum amplitude) and relate that parameter to pain reports [8,13,16,22,23,27,34]. As autonomic responses typically entail both phasic and tonic components, this approach may neglect the information present in the full time-course. One way to use this information is principal component regression (PCR): individual trial time-courses are represented by scores of the most important principal components [15], thereby providing a more accurate representation of the time-course than a summary statistic. The main objective of the current study was therefore to investigate whether pain prediction can be improved by incorporating temporal information present in autonomic recordings.

Ideally, an objective measure of pain would be available at minimal costs, both monetary and with regard to the experimental design. Additionally, a signal that is reliable enough to work on individual trials is more helpful than a measure aggregated across trials. As the perceived intensity of painful stimuli varies considerably across studies and subjects, a parametric prediction

* Corresponding author. Address: Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, Martinistr. 52, Hamburg D-20246, Germany. Tel.: +49 40 7410 57890.
E-mail address: s.geuter@uke.de (S. Geuter).

http://dx.doi.org/10.1016/j.pain.2014.02.005
0304-3959/

of different pain intensity levels at the trial level seems desirable instead of relying on binary pain vs no-pain reports [31]. Furthermore, a prediction model would optimally be able to predict unseen data from an independent test data set without re-fitting subject specific parameters. In order to test an objective, parametric marker of pain intensity, predictive at the trial level across subjects, we recorded autonomic responses to a set of different cutaneous heat pain stimuli. The predictive performance of the PCR model utilizing detailed temporal information was then compared to a prediction based on summary statistics (ie, amplitudes) extracted from the physiological responses.

A recent study reported that pain prediction by a combination of several physiological measures outperforms prediction by single measures [34]. We therefore recorded 2 autonomic measures, skin conductance and pupil diameter, and combined these for parametric pain prediction.

2. Materials and methods

2.1. Samples

Thirty-seven healthy male subjects participated for payment in this study. All except one subject were right-handed. Subjects reported no history of neurological, psychiatric, or skin diseases and had not taken any medication during the last 48 hours prior to the experiment. One subject had to be excluded because of no evident skin conductance response. Two further subjects had to be excluded because of poor eye-tracking data quality. The final sample thus consisted of 34 subjects, aged 21-37 years (mean age: 25.8 years). In order to evaluate the generalizability of the PCR model, we used a second independent data set. This sample included 20 additional male subjects (mean age: 27.5 years; range: 22-39 years) participating in an experiment very similar to the main experiment. The Ethics committee of the Medical Chamber Hamburg approved the study and all subjects gave written informed consent.

2.2. Procedure

Subjects were individually tested in an eye-tracking laboratory with controlled equal illumination in all sessions. First, subjects were informed about the study and provided informed consent. Subjects then washed their hands with warm water, but without soap, to optimize skin conductance recording. Skin conductance recording electrodes were attached to the subject's left hand. To calibrate the pupil diameter properly, 2 epochs of 5 seconds duration were recorded with artificial pupils of 5 and 10 mm diameter. Fake pupils were fixed over the closed right eye while the subjects placed their head in the headrest of the eye tracker. Subsequently, a 9-point gaze calibration was performed with the subject's head positioned in the headrest. Pain thresholds (mean: 45.7°C, SD: 3.0°C) were then determined according to the method of limits (1°C/s slope). The main experiment consisted of 32 cutaneous heat pain trials, split into 4 blocks of 8 trials each. The thermode was moved to a different patch on the volar forearm after completion of a block to prevent sensitization. During heat pain stimulation, skin conductance and pupil diameter were recorded.

The heat pain stimulation paradigm consisted of 8 different stimuli, each repeated 4 times. Each stimulus occurred once per block. Stimulation order was pseudo-randomized across subjects such that each stimulus was equally often at the first position within a block. Stimulus temperatures were 45, 46, 47, and 47.5°C and lasted 8 or 20 seconds, resulting in 8 different combinations of temperature and duration. Stimulus duration included a ~1.5-second ramp-up and -down period and a plateau lasting 5 (short trials) or 17 seconds (long trials), respectively. Each trial (Fig. 1A) consisted of an anticipation period of 13–17 seconds, heat pain stimulation, a 5-second delay, pain rating on a visual analogue scale (VAS), and a 12-second intertrial interval. Subjects were asked to fixate a crosshair at the center of the screen and to refrain from blinking during the anticipation and stimulation periods. Pain stimulation was not cued, that is, the crosshair remained unchanged during both periods. Pain ratings were completed on a VAS anchored "no pain" and "unbearable pain." The delay between heat stimulation and rating was introduced to prevent contamination of autonomic recordings by button presses [20]. During the intertrial interval, a blank screen was presented and subjects were allowed to move their eyes freely. After each block, subjects could rest for a few minutes. The whole procedure lasted about 45 minutes.

2.3. Data acquisition

Response logging, thermode triggering, and synchronization with the physiological recordings were controlled by Presentation software (Neurobehavioral Systems, Berkeley, CA, USA). A 3 × 3 cm Peltier thermode (Pathway ATS; Medoc Advanced Medical Systems, Ramat Yishai, Israel) was used to deliver thermal stimulation on the left volar forearm. Skin conductance was recorded using a constant voltage (0.5 V) Biopac MP100 system (Biopac Systems, Inc., Goleta, CA, USA) at a sampling rate of 250 Hz. Ag/AgCl recording electrodes filled with 0.05 M NaCl electrolyte were placed on thenar and hypothenar eminences of the left hand. Pupil diameter was recorded at a sampling rate of 1000 Hz using an Eyelink 1000 system (SR Research Ltd., Mississauga, ON, Canada).

2.4. Data analysis

All data analyses were completed using MATLAB v7.12 (Mathworks, Natick, MA, USA) and SPSS 19 (IBM, Armonk, NY, USA). After preprocessing the physiological data, we predicted pain ratings using 2 different sets of predictors: (1) using summary statistics describing the physiological response of each trial (ie, amplitude) and (2) utilizing the temporal information of the physiological responses for prediction. The former will be referred to as summary statistic regression and the latter will be referred to as PCR [15]. According to a recent study [34], simultaneous use of multiple autonomic measures differentiates best between pain stimulus levels. Hence, we predicted pain simultaneously by skin conductance and pupil measures. Results are summarized in Table 1. For completeness, we report results on separate regression models (ie, only skin conductance or pupil dilation plus intercept) in Table 2.

The effect of stimulus duration and temperature on pain ratings was tested using a 2 × 4 repeated-measure analysis of variance.

2.4.1. Skin conductance

Skin conductance traces were down-sampled to 25 Hz, low-pass filtered at a cutoff frequency of 1 Hz, and epochs from 3 seconds before to 5 seconds after stimulus presentation were selected for further analyses. Trials with recording artifacts were removed from further analyses ($n = 22$; 2%). To quantify the skin conductance response with a summary statistic, we extracted 3 different parameters. The skin conductance response (SCR) amplitude was calculated by subtracting the local minimum at the onset of the first SCR from the first peak. The SCR onset was required to occur between 1 and 5 seconds after stimulus onset. We chose this interval because the thermode needs 1–2 seconds to reach its target temperature. SCR amplitudes below 0.02 μS were set to zero. Skin conductance level (SCL) was computed by averaging the skin conductance trace from stimulus onset until 5 seconds after stimulus

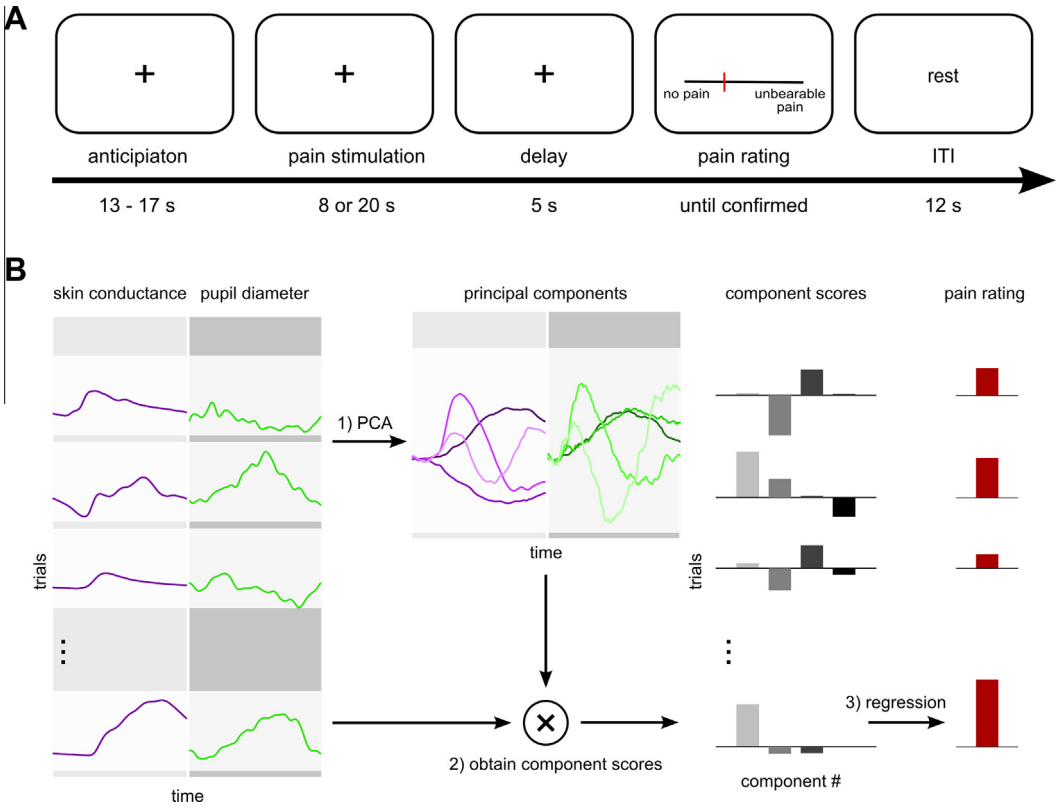

Fig. 1. Trial structure and principal component regression. (A) Each trial consisted of anticipation, heat pain stimulation, delay, visual analogue scale pain ratings and intertrial interval (ITI). The anticipation period randomly varied between 13 and 17 seconds. Pain stimulation was either 8 or 20 seconds, with a temperature of 45, 46, 47, or 47.5°C. During ITI, subjects were allowed to freely move their eyes. They were asked to fixate on a central crosshair during anticipation, pain stimulation, and the delay. (B) Outline of the principal component regression (PCR) analysis. In a first step, the most important principal components were computed using principal component analysis (1). The original data matrix was then multiplied by the principal components to obtain component scores for each trial (2). Finally, these component scores were used as predictors of pain ratings in linear regression analysis (3). Please note that no trial variable was used as a predictor in the regression analysis. Only an intercept and 4 component scores per trial were included in the model. The 4 principal components computed for the long trials are shown in the second panel.

Table 1
Model fits for summary statistic and principal component regression (PCR).

	Short trials		Long trials	
	Summary statistic	PCR	Summary statistic	PCR
BIC	2556.2	2545.2	2606.9	2541.2
RMSE	21.0	20.4	25.8	23.4
r	0.60	0.63	0.62	0.70
Adj. R^2	0.36	0.39	0.38	0.49

BIC, Bayesian information criterion; RMSE, root mean squared error; r = Pearson's correlation between observed and predicted pain ratings.

Table 2
Model fits for principal component regression based on single autonomic variables.

	Short trials		Long trials	
	Skin conductance	Pupil diameter	Skin conductance	Pupil diameter
BIC	2592.8	2606.6	2635.0	2591.9
RMSE	21.6	21.9	26.4	24.9
r	0.57	0.55	0.60	0.65
Adj. R^2	0.32	0.30	0.35	0.42

BIC, Bayesian information criterion; RMSE, root mean squared error; r = Pearson's correlation between observed and predicted pain ratings.

offset and subtracting the average amplitude of baseline immediately 1 second before stimulus onset. The third measure was the skin conductance global amplitude (SCG), which is the difference between the maximum during the stimulation phase (1 second after stimulus onset to 5 seconds after stimulus offset) and the 1-second baseline before stimulus onset. The SCR and SCG values were log-transformed before averaging [35].

2.4.2. Pupil diameter

Eye-tracking data were recorded from 3 seconds before to 5 seconds after heat stimulation, resulting in single epochs for each trial. Periods of 100 ms before and after blinks were removed from the pupil diameter traces and the whole blink period was linearly interpolated. Trials in which 50% or more of samples had to be replaced were discarded (256 trials, 23%). Excluded trials were equally distributed across stimulation conditions, and exclusion of these trials did only marginally change predictive accuracy. Pupil data were then down-sampled to 25 Hz. Pupil diameter was linearly transformed from arbitrary values to mm by using the calibration recordings with the artificial pupils of 5 and 10 mm diameter. Analogous to the skin conductance parameters (SCR, SCL, SCG), we calculated summary statistics for the pupil diameter. The pupil dilation response (PDR) was computed from the first local minimum after stimulus onset to the first local maximum. Pupil dilation level (PDL) was computed as the average diameter during the period from stimulation onset to 5 seconds after stimulus offset minus baseline (1 second before stimulus onset). The pupil diameter global peak (PDG) was defined as the global maximum during the same period as used for the PDL minus

prestimulus baseline. As with the skin conductance data, PDR and PDG were log-transformed.

2.4.3. Regression analyses

To evaluate the bivariate relationships between stimulus features and subjects' behavioral and physiological responses, we computed correlations of temperature and pain ratings with summary statistics (SCR, SCL, SCG, PDR, PDL, and PDG) separately for short and long trials. Pearson's correlation coefficients were computed within each subject, Fisher *z*-transformed, averaged across subjects, and inverse transformed. For both short and long trials, the bivariate correlation between pain ratings and the global amplitude (SCG and PDG) was highest. In order to predict pain ratings from physiological data, we specified linear regression models separately for long and short trials. In total, 4 regression models were specified, 2 for the summary statistics approach (one for short and one for long trials) and 2 for the PCR approach (for short and long trials, respectively). To predict pain by summary statistics, SCG, PDG, and an intercept were used as predictors in linear regression analyses. The PCR models (see below) included 5 predictors – an intercept and 4 principal component scores. All analyses involving physiological measures are based on a total of 815 valid trials (417 short trials and 398 long trials).

In order to utilize the information present in the time-course of the physiological measures, we adopted a PCR approach [15,28,36,37]. The original time-courses organized in matrix $\boldsymbol{X}$ (with n trials $\times$ m data samples) cannot be used as pain predictors for the following reasons: Since subsequent data samples within time-series data have nonzero autocorrelation, they are dependent on each other. This leads to highly correlated columns $(1 \ldots m)$ within the predictor matrix, and multicollinearity poses a problem for regression analysis. Therefore, regression weights become highly unstable with increasing numbers of predictors. To overcome these problems, PCR first extracts principal components from the set of predictors (here, data points from pupil diameter and skin conductance) and then uses the component scores of each trial as predictors of the pain ratings (Fig. 1B). By using only scores of the most important components, the number of predictors is reduced and multicollinearity is removed because principal components are orthogonal to each other. Hence, PCR can be seen as a feature selection technique that uses only the most important components of the physiological time-courses.

In order to predict pain by component scores, we range corrected (0-1) pupil and skin conductance traces within subjects and concatenated both physiological traces of each trial. Using principal component analyses, we extracted the first 4 principal components across trials. The 4 principal components accounted for about 85% of the variance explained by all components. We chose to extract 4 principal components because this number of components achieved the lowest Bayesian information criterion (BIC) value combined for short and long stimulus durations. The component scores $\boldsymbol{S}$ are then given by $\boldsymbol{S} = \boldsymbol{XP}$, where $\boldsymbol{X}$ is the original data matrix, and $\boldsymbol{P}$ the $m \times 4$ matrix of principal components. This resulted in a dramatic reduction of dimensionality where each single trial can be expressed by a vector of 4 weights. Component scores for each single trial ($\boldsymbol{S}$) were then used to predict pain ratings using standard regression analysis, producing regression coefficients ($\boldsymbol{\beta}$) for the intercept and the principal components. Please note that due to the concatenated physiological data, this regression model combines skin conductance and pupil diameter as predictors (Fig. 1B).

By multiplying the principal components $\boldsymbol{P}$ with the obtained regression coefficients $\boldsymbol{\beta}$, the latter were projected to the time space for interpretation ($\boldsymbol{w} = \boldsymbol{P\beta}$). The vector $\boldsymbol{w}$ contains prediction weights for each data sample of the pupil and skin conductance recordings. The weights in $\boldsymbol{w}$ can also be used to predict pain scores from new physiological recordings without conducting a new principal component analysis. Although we used an intercept in the regression models, $\boldsymbol{w}$ was calculated without the intercept. In order to predict pain ratings from new data sets, the intercept has to be added again.

Although the data are hierarchically structured (trials nested within subjects), we chose a single-level regression model for the following reasons: We aimed at identifying prediction weights that can be applied to new data sets. Therefore, we assumed fixed intercepts and slopes across subjects and data sets. This may degrade the model fit in the training data. To estimate performance differences between fixed and random intercept models, we also estimated a regression model with random subject intercepts. However, with intraclass correlations of $\rho_{short} = 0.176$ and $\rho_{long} = 0.113$, the sample intercepts for short and long trials changed only slightly compared to the single-level model (short: 22.0 vs 21.8; long: 46.2 vs 45.7). Applying the sample intercept and slopes from the random intercept model (long trials) to an independent test data set (see Section 2.4.4) did not change performance of this cross-validation (ΔRMSE = −0.034; Δr = −0.007).

2.4.4. Model evaluation

In order to compare the 2 regression models against each other, we computed BIC values [32]. The BIC gives an overall estimate of the model fit in relationship to the number of predictors and sample size. This allows for direct comparison of model fits independent of the number of model parameters. For the case of linear regression, BIC values are given by

$$\mathrm{BIC} = n\log(\hat{\sigma}^2) + k\log(n)$$

where n is the sample size, $\hat{\sigma}^2$ is the maximum likelihood estimator of the error variance, and k is the number of predictors [19]. Lower BIC values indicate better model fit. In addition to the BIC, we computed the root mean squared error (RMSE), the Pearson's correlation coefficient r between observed and predicted pain ratings, and the adjusted R^2 value for the regression models. Correlations between observed and predicted data for PCR and summary statistic regression were compared using a *z*-test [24].

By projecting the regression coefficients $\boldsymbol{\beta}$ back to the time-domain, a vector of prediction weights $\boldsymbol{w}$ can be obtained. By applying the prediction weights $\boldsymbol{w}$ to a new data set, one can obtain pain predictions for an independent study. In order to evaluate the generalizability of the PCR model, we used the weight vector $\boldsymbol{w}$ and applied it to data from an independent validation sample. This data set consisted of pupil and skin conductance recordings from 20 additional subjects. The experiment was a heat pain stimulation paradigm in which subjects fixated a central crosshair. Subjects rated heat pain stimuli of 20 seconds duration and temperatures of 45, 46, 47, and 47.5°C and did not complete any additional task. Heat stimulation was also applied to the left forearm, and each stimulus was applied 4 times distributed over 2 blocks. All other aspects of data acquisition and data processing were identical to the main sample. We used the weights $\boldsymbol{w}$ from the long trial PCR model. Each data sample from the validation set was then multiplied by the respective weight from the PCR. We then computed the RMSE and the correlation between predicted and observed pain ratings to evaluate the PCR model generalizability.

In order to estimate the significance of individual sample weights, we performed a permutation analysis. By randomly shuffling the pain ratings with respect to the physiological data 10,000 times, we computed a null distribution of prediction weights $\boldsymbol{w}$, separately for short and long trials. For each permutation, we estimated regression weights and projected them to the time space. Based on these null distributions, we could determine *P*-values for weights at each data point. False discovery rate [3] was used

to control for multiple comparisons at a significance level of $P < 0.05$.

As pain ratings were not standardized, the resulting regression weights are therefore also unstandardized, that is, their magnitude scales with the dependent variable and the duration of the physiological epochs. Hence, weight magnitudes cannot be directly compared across trial lengths.

3. Results

3.1. Stimuli and pain ratings

First, we evaluated the dependence of pain ratings on stimulus features (ie, temperature and duration). A repeated-measure analysis of variance revealed significant main effects for temperature [$F(3,99) = 526.7$; $P < 0.001$] and duration [$F(1,33) = 553.9$; $P < 0.001$], as well as a significant interaction [$F(3,99) = 56.4$; $P < 0.001$]. Evaluation of the mean pain ratings revealed that pain increased with stimulus duration and temperature, as would be expected. This increase was more pronounced for long, as compared to short, trials (Fig. 2).

3.2. Evoked responses

Inspecting the evoked skin conductance and pupil responses (Fig. 3A) revealed several findings. First, the evoked responses discriminate between the different stimulus temperatures. Second, the initial phasic response is very similar across temperatures and durations (see first local maximum in the evoked responses plotted in Fig. 3A). This initial response is commonly quantified as SCR or PDR and it presumably constitutes an unspecific response to the onset of the heat stimulus. Third, the evoked responses of skin conductance and pupil dilation look similar at first glance, but the initial response of the pupil dilation has shorter latencies and decreases faster than the skin conductance responses.

To further evaluate the relationship between physiological measures and stimulus intensity, we quantified 3 measures for each signal modality; the initial response (SCR and PDR), the global measures (SCG and PDG), and the average response level relative to baseline (SCL and PDL). Fig. 3B displays these summary statistics

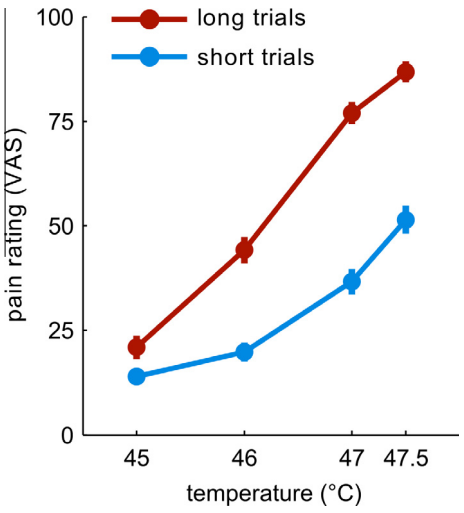

Fig. 2. Grand means of pain ratings for each stimulus for the training data set. Pain ratings were lower for the short trials (blue) and increased with temperature in both duration conditions. This increase was stronger for the long trials (red). VAS, visual analogue scale. Error bars indicate standard error of the mean (SEM).

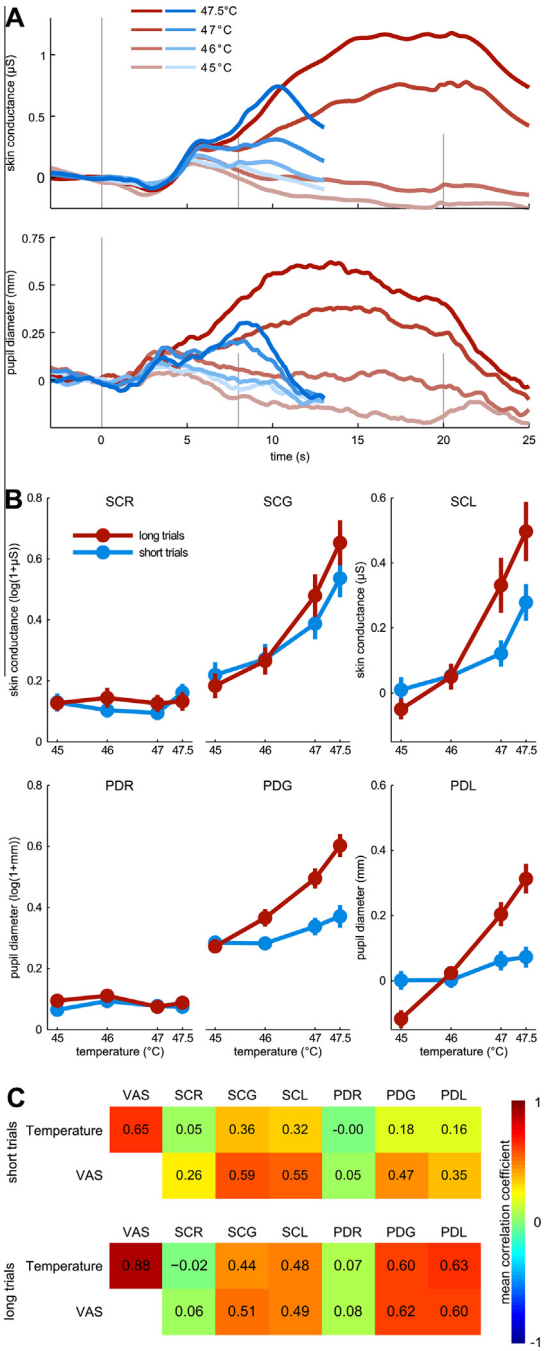

Fig. 3. Pain ratings and evoked responses. (A) Physiological recordings aligned to the stimulus onset at $t = 0$ seconds. The top panel shows skin conductance responses, the bottom panel shows pupil diameter recordings. Both measures clearly discriminate stimulus temperatures. Short stimuli lasted 8 seconds (blue) and long stimuli lasted 20 seconds (red). Darker colors indicate higher temperatures. The gray lines at 8 and 20 seconds denote stimulus offsets. The mean of a 1-second baseline prior to stimulus onset was subtracted from traces before plotting. (B) Plots of summary statistics against stimulus temperatures are shown for short (blue) and long (red) trials. The top panels show skin conductance variables and the bottom panels show measures of pupil diameter. Skin conductance response (SCR) and pupil dilation response (PDR) describe the initial peak after stimulus onset, skin conductance global amplitude (SCG) and pupil diameter global peak (PDG) describe the maximum amplitude, and skin conductance level (SCL) and pupil dilation level (PDL) are the average levels over the stimulation period. Error bars indicate standard error of the mean (SEM). (C) Averaged correlations of summary statistics, temperature (Temp.), and pain ratings (visual analogue scale [VAS]) are shown for short and long trials in the upper and lower section, respectively.

for all stimulus conditions. The SCR and PDR variables describe the first peak amplitude in the evoked responses and do not scale with stimulus temperature. The 2 variables describing the maximum amplitude during the long pain stimuli, SCG and PDG, differentiate between temperatures. For the short stimuli, PDG does not clearly differentiate between temperatures. A similar pattern was observed for the level statistics (SCL and PDL).

As the focus of this study is on the prediction of pain by physiological measures, we did not use factorial tests for effects of stimulus features on physiological responses. Instead we computed bivariate correlations between the physiological variables, pain ratings, and stimulus temperatures separately for long and short trials (Fig. 3C). Interestingly, neither the SCR nor the PDR amplitudes correlated with stimulus intensity and pain rating (all $|\bar{r}| \leqslant 0.08$, except for short trials SCR and VAS $\bar{r} = 0.26$). The other summary statistics (SCG, PDG, SCL, and PDL) were linearly related to pain ratings and less to temperature (mean difference between temperature and ratings of $\Delta\bar{r} = 0.24$ for short trials and $\Delta\bar{r} = 0.02$ for long trials; Fig. 3C). The global amplitude measures correlated with reported pain in short trials (SCG: $\bar{r} = 0.59$; PDG: $\bar{r} = 0.47$) and long trials (SCG: $\bar{r} = 0.51$; PDG: $\bar{r} = 0.62$). The SCL correlated slightly less with pain reports (long stimuli: $\bar{r} = 0.49$; short stimuli: $\bar{r} = 0.55$). The variables correlating highest with pain ratings were SCG and PDG, respectively. We therefore chose these variables to predict pain ratings by summary statistics.

3.3. Pain prediction

We then turned to pain prediction by regression models. First, the SCG and PDG amplitudes (together with an intercept) were used as predictors in a linear regression model. Separate regression models were used for long and short trials, respectively. The correlation between observed and pain ratings predicted by the 2 regression models were $r = 0.60$ for short trials and $r = 0.62$ for long trials (all model fit parameters are shown in Table 1). We next evaluated the model fit using BIC, thereby controlling for the number of predictors. BIC values for the summary statistic regression models were 2556.2 for short trials and 2606.9 for the long trials. Model fits of the summary statistics regression were then compared to the PCR using BIC values. The PCR model represents individual trial time-courses by using principal components (Methods and Fig. 1B). BIC values of the PCR models were lower than for the summary statistics approach, indicating better model fit by the PCR ($BIC_{short} = 2545.2$; $BIC_{long} = 2541.2$). All other fit parameters (ie, RMSE, r, and adjusted R^2) favored the PCR models over the summary statistic models (Table 1). The correlation between observed pain ratings and PCR predictions in long trials was $r = 0.70$ (Fig. 4A) and $r = 0.63$ for short trials. Interestingly, prediction performance based on our physiological recordings is comparable to the performance of $r = 0.74$ achieved by prediction from pain processing brain areas [37].

Comparing correlations between observed and predicted ratings obtained from PCR and summary statistic models [24] revealed higher correlations for the PCR model for long trials ($z = 3.37$; $P < 0.001$). For short trials, no significant difference in correlations between predicted and observed ratings was found ($z = 1.03$; $P = 0.30$).

Having identified better prediction by incorporating temporal information for long stimuli, we next evaluated the contribution of data points over time. This is possible by projecting the component score regression weights back to the temporal dimension, that is, onto the individual data points of skin conductance and pupil diameter (Fig. 4B). For both measures, skin conductance and pupil diameter, weight values showed an initial dip and a peak during

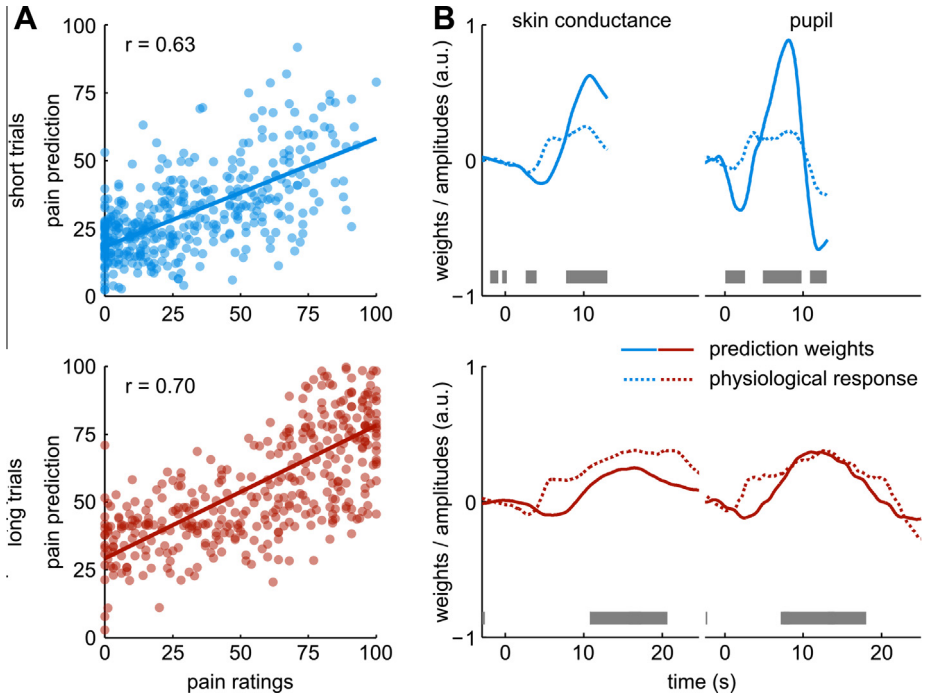

Fig. 4. Prediction and prediction weights. (A) Observed pain ratings plotted against the predicted pain ratings from the principal component regression model for short trials (top) and long trials (bottom). The correlation between observed and predicted pain ratings was $r = 0.63$ for short trials and $r = 0.70$ for long trials. Each dot corresponds to one trial. (B) Prediction weights (solid lines) for individual data points are plotted for short trials (top) and long trials (bottom). The grand mean of normalized physiological responses are plotted on the same axes (dotted lines), scaled by a factor of 4. Skin conductance is plotted on the left side, pupil diameter on the right. All plots are aligned to heat stimulus onset at t = 0 seconds. The gray bars denote data points at which the prediction weights are significantly different from 0 at false discovery rate (FDR) < 0.05. Please note that the weight values depend on the scaling of the pain ratings and the number of data points in the physiological recordings. The weights of short and long trials cannot be directly compared to each other.

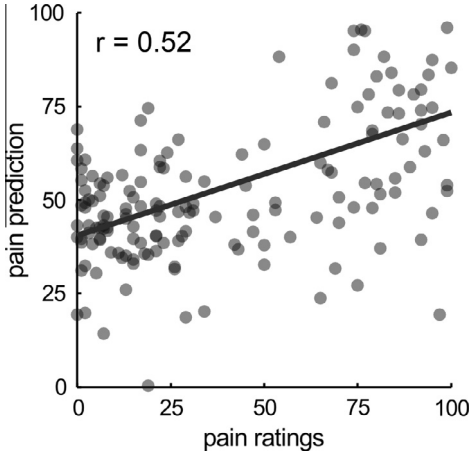

Fig. 5. Validation in an independent test data set. Prediction weights of the long-trial principal component regression model were applied to skin conductance and pupil diameter recordings obtained from 20 independent subjects.

late heat stimulation. Significant and high prediction weights were located around the peak of the physiological response, indicating positive association of those samples with pain reports. The pattern was similar for short and long trials (Fig. 4B). Please note that the weights cannot be directly compared between the 2 trial types because they are unstandardized. Weights need to be rescaled before applying prediction weights to a dataset of different length. The weight curve holds information about how to compute a weighted and scaled average of the physiological response to predict the subjective pain and which parts of the response are more closely related to the pain report.

In order to validate the generalizability of the best-fitting regression model, we used an independent test data set from 20 additional subjects who also experienced long heat pain stimuli (20 seconds). Projecting the regression weights of the component scores back to the time-space resulted in prediction weights for individual data points and thus allowed for pain prediction in the new data set. The correlation between observed and predicted pain ratings was high ($r = 0.52$, $P < 0.001$; Fig. 5). The RMSE increased to 31.9. This increase may be related to the difference in average pain ratings between the 2 data sets (M = 40.2 for the test set and M = 57.3 for the training data). Application of prediction weights to new data set resulted in model fit parameters (RMSE and r) comparable to the summary statistic regression in the training data set.

4. Discussion

Using temporal information about skin conductance and pupil diameter responses, we identified a physiological signature of pain perception. The signature is easy to record and predicts pain in single trials on a continuous rating scale ranging from 0 to 100 across different subjects and data sets. The temporally informed prediction by principal component scores outperformed prediction by a standard summary statistic, maximum amplitude. BIC values of the PCR models were lower for both short and long trials. For long heat pain stimuli, a z-test on the correlation coefficients between predicted and observed ratings revealed a better fit of the PCR model.

Pain ratings have rarely been parametrically predicted by biological measures of pain [17,37]. Instead, a categorical differentiation has often been used in similar studies [5,21,22,31,34]. However, it is desirable not only to determine whether a subject is currently in pain, but also to determine the pain intensity. Electroencephalography and functional magnetic resonance imaging (fMRI) have recently been proposed for this purpose [17,37]. Both studies achieved good predictive accuracy, and the study by Wager et al. [37] comprehensively demonstrated the specificity of their marker for pain. These 2 studies and the present one evaluated the relationship of a physiological pain response to pain reports on a rating scale. It would be highly interesting to see how the 3 physiological signatures relate to each other and at the same time to pain reports. Some studies evaluated bivariate relationships between brain responses and pain ratings [1,4,6], brain responses and skin conductance [11,33], or between skin conductance and pain ratings [7,10,23,25,34], but an investigation of all 3 variables at a time is missing. Brain and autonomic markers might be highly correlated, and pain reports could, in principal, be less correlated to any of the physiological variables. A hint toward this hypothesis comes from the observation that both our PCR prediction and an fMRI-based prediction [37] overestimate low pain ratings (see Fig. 4A in this study and Fig. 1B in [37]). This may be due to the intrinsic properties of the physiological responses or due to a nonlinear mapping of pain onto rating scales. Although subjects in the current study were instructed to rate every nonpainful stimulus at scale minimum, anecdotal reports during debriefing suggest that some subjects also used the lower part of the scale to rate stimuli around or below the pain threshold. It may also be the case that physiological pain markers work only for painful stimuli and not for innocuous stimuli [37]. With regard to the relationship of different physiological pain markers to pain reports, it could also be that physiological measures are too noisy to perfectly predict pain reports (or pain ratings may be noisy).

An anatomical link between nociceptive stimulation and the observed autonomic responses might include the hypothalamus and locus coeruleus (LC). Both structures receive nociceptive input via spinohypothalamic and spinoreticular pathways, respectively. The LC also projects back to the spinal cord dorsal horn [14] and indirectly controls pupil dilation via the Edinger-Westphal nucleus [8,12,26]. Autonomic nervous system activity and, particularly, measures of its sympathetic division, like skin conductance and pupil dilation, are at least partially controlled by the hypothalamus and correlate very well with LC activity [8,10,26]. These projections to and from hypothalamus and LC might link the observed autonomic measures to pain processing in the brain.

As the goal of the current study was to investigate easily available pain markers, we used skin conductance and pupil diameter. As mentioned above, electroencephalography and fMRI have previously been used for accurate pain prediction [17,37]. Nevertheless, these methods seem to be rather expensive and impracticable for routine use in experimental or clinical studies because they require shielded recording environments and impose several restrictions on experimental design and subjects. Peripheral physiological markers, on the other hand, impose only minimal restrictions on experimental design and are affordable for many studies. We therefore turned to 2 autonomic variables – skin conductance and pupil diameter – and showed that these measures achieve a predictive accuracy similar to that of fMRI-based markers (correlation between observed and predicted pain: $r_{autonomic} = 0.7$ vs $r_{fMRI} = 0.74$ [37]).

An important aspect of our prediction model is the use of a single parameter set for all subjects, that is, we did not account for the hierarchical structure of the data (see Chapman et al. [7]). This may decrease prediction accuracy within the training sample, but allows generalization, for example, application to new data. As shown by the validation on the independent data set, the prediction model generalizes well to new data sets, although the mean pain level was lower in the test set.

Furthermore, pain markers and reports may differ between different types of pain. Until now, the proposed pain markers were tested with only experimental heat pain stimuli. To extend the current findings to other methods of pain induction or even to clinical types of pain, further translational research is needed [18,30].

Inspection of the prediction weights assigned to individual data points revealed a strong influence of late data points around the global maximum of the response. This is in line with the observation that summary statistics considering the whole epoch (eg, SCG) are well correlated with pain as reported in previous studies [2,7,18,23,34]. The difference between summary statistics and the PCR approach became stronger with longer stimuli and longer physiological recordings. Interestingly, the data points around the initial physiological response received negative weights. The amplitude of the initial peak of the autonomic recordings was unrelated to pain and stimulus temperature in terms of a linear correlation. When this initial response is analyzed in isolation, it is thus unrelated to pain perception. When all time points are considered simultaneously, its amplitude is de-weighted relative to the later global peak.

Taken together, we identified an objective physiological predictor of pain ratings that can be easily integrated into experimental studies to corroborate subjective pain ratings of individual trials. This is particularly helpful when subjects' responses may be biased or when subjects are not able to provide ratings. However, we would like to stress that the absence of a physiological response does not mean the absence of pain in a given subject. There may be factors inhibiting a physiological response, although the current predictor seems to slightly overestimate ratings in low pain trials.

Conflicts of interest statement

The authors declare no conflict of interest.

Acknowledgements

We thank Birgit Popp and Adrian Rothers for great help with data collection. C.B. and S.G. were supported by the DFG, SFB 936 (A06).

References

[1] Apkarian AV, Bushnell MC, Treede R-D, Zubieta J-K. Human brain mechanisms of pain perception and regulation in health and disease. Eur J Pain 2005;9:463–84.
[2] Aslaksen PM, Myrbakk IN, Høifødt RS, Flaten MA. The effect of experimenter gender on autonomic and subjective responses to pain stimuli. PAIN® 2007;129:260–8.
[3] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B Methodol 1995;57:289–300.
[4] Bornhövd K, Quante M, Glauche V, Bromm B, Weiller C, Büchel C. Painful stimuli evoke different stimulus–response functions in the amygdala, prefrontal, insula and somatosensory cortex: a single-trial fMRI study. Brain 2002;125:1326–36.
[5] Brodersen KH, Wiech K, Lomakina EI, Lin C, Buhmann JM, Bingel U, Ploner M, Stephan KE, Tracey I. Decoding the perception of pain from fMRI using multivariate pattern analysis. Neuroimage 2012;63:1162–70.
[6] Büchel C, Bornhövd K, Quante M, Glauche V, Bromm B, Weiller C. Dissociable neural responses related to pain intensity, stimulus intensity, and stimulus awareness within the anterior cingulate cortex: a parametric single-trial laser functional magnetic resonance imaging study. J Neurosci 2002;22:970–6.
[7] Chapman CR, Nakamura Y, Donaldson GW, Jacobson RC, Bradshaw DH, Flores L, Chapman CN. Sensory and affective dimensions of phasic pain are indistinguishable in the self-report and psychophysiology of normal laboratory subjects. J Pain 2001;2:279–94.
[8] Chapman CR, Oka S, Bradshaw DH, Jacobson RC, Donaldson GW. Phasic pupil dilation response to noxious stimulation in normal volunteers: relationship to brain evoked potentials and pain report. Psychophysiology 1999;36:44–52.
[9] Colloca L, Benedetti F, Pollo A. Repeatability of autonomic responses to pain anticipation and pain stimulation. Eur J Pain 2006;10:659–65.
[10] Donaldson GW, Chapman CR, Nakamura Y, Bradshaw DH, Jacobson RC, Chapman CN. Pain and the defense response: structural equation modeling reveals a coordinated psychophysiological response to increasing painful stimulation. PAIN® 2003;102:97–108.
[11] Dubé A-A, Duquette M, Roy M, Lepore F, Duncan G, Rainville P. Brain activity associated with the electrodermal reactivity to acute heat pain. Neuroimage 2009;45:169–80.
[12] Einhäuser W, Stout J, Koch C, Carter O. Pupil dilation reflects perceptual selection and predicts subsequent stability in perceptual rivalry. Proc Natl Acad Sci U S A 2008;105:1704–9.
[13] Ellermeier W, Westphal W. Gender differences in pain ratings and pupil reactions to painful pressure stimuli. PAIN® 1995;61:435–9.
[14] Fields HL, Basbaum AI, Heinricher MM. Central nervous system mechanisms of pain modulation. In: McMahon SB, Koltzenburg M, editors. Wall and Melzack's textbook of pain. London: Churchill Livingstone; 2006. p. 125–42.
[15] Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference, and prediction. 2nd ed. Berlin: Springer; 2009.
[16] Höfle M, Kenntner-Mabiala R, Pauli P, Alpers GW. You can see pain in the eye: pupillometry as an index of pain intensity under different luminance conditions. Int J Psychophysiol 2008;70:171–5.
[17] Huang G, Xiao P, Hung YS, Zhang ZG, Hu L. A novel approach to predict subjective pain perception from single-trial laser-evoked potentials. Neuroimage 2013;81:283–93.
[18] Hullett B, Chambers N, Preuss J, Zamudio I, Lange J, Pascoe E, Ledowski T. Monitoring electrical skin conductance: a tool for the assessment of postoperative pain in children? Anesthesiology 2009;111:513–7.
[19] Hurvich CM, Tsai C-L. The impact of model selection on inference in linear regression. Am Stat 1990;44:214–7.
[20] Kietzmann TC, Geuter S, König P. Overt visual attention as a causal factor of perceptual awareness. PLoS One 2011;6:e22614.
[21] Ledowski T, Ang B, Schmarbeck T, Rhodes J. Monitoring of sympathetic tone to assess postoperative pain: skin conductance vs surgical stress index. Anaesthesia 2009;64:727–31.
[22] Ledowski T, Bromilow J, Wu J, Paech MJ, Storm H, Schug SA. The assessment of postoperative pain by monitoring skin conductance: results of a prospective study. Anaesthesia 2007;62:989–93.
[23] Loggia ML, Juneau M, Bushnell MC. Autonomic responses to heat pain: heart rate, skin conductance, and their relation to verbal ratings and stimulus intensity. PAIN® 2011;152:592–8.
[24] Meng X, Rosenthal R, Rubin D. Comparing correlated correlation coefficients. Psychol Bull 1992;111:172–5.
[25] Nakamura Y, Donaldson GW, Kuhn R, Bradshaw DH, Jacobson RC, Chapman CR. Investigating dose-dependent effects of placebo analgesia: a psychophysiological approach. PAIN® 2012;153:227–37.
[26] Nieuwenhuis S, De Geus EJ, Aston-Jones G. The anatomical and functional relationship between the P3 and autonomic components of the orienting response. Psychophysiology 2011;48:162–75.
[27] Oka S, Chapman CR, Kim B, Nakajima I, Shimizu O, Oi Y. Pupil dilation response to noxious stimulation: effect of varying nitrous oxide concentration. Clin Neurophysiol 2007;118:2016–24.
[28] Onat S, Nortmann N, Rekauzke S, König P, Jancke D. Independent encoding of grating motion across stationary feature maps in primary visual cortex visualized with voltage-sensitive dye imaging. Neuroimage 2011;55:1763–70.
[29] Price DD, McGrath PA, Rafii A, Buckingham B. The validation of visual analogue scales as ratio scale measures for chronic and experimental pain. PAIN® 1983;17:45–56.
[30] Rantanen M, Yli-Hankala A, van Gils M, Yppärilä-Wolters H, Takala P, Huiku M, Kymäläinen M, Seitsonen E, Korhonen I. Novel multiparameter approach for measurement of nociception at skin incision during general anaesthesia. Br J Anaesth 2006;96:367–76.
[31] Schulz E, Zherdin A, Tiemann L, Plant C, Ploner M. Decoding an individual's sensitivity to pain from the multivariate analysis of EEG data. Cereb Cortex 2012;22:1118–23.
[32] Schwarz G. Estimating the dimension of a model. Ann Stat 1978;6:461–4.
[33] Seifert F, Schuberth N, De Col R, Peltz E, Nickel FT, Maihöfner C. Brain activity during sympathetic response in anticipation and experience of pain. Hum Brain Mapp 2013;34:1768–82.
[34] Treister R, Kliger M, Zuckerman G, Aryeh IG, Eisenberg E. Differentiating between heat pain intensities: the combined effect of multiple autonomic parameters. PAIN® 2012;153:1807–14.
[35] Venables PH, Christie MJ. Electrodermal activity. In: Martin I, Venables PH, editors. Techniques in psychophysiology. Chichester, UK: Wiley; 1980. p. 3–67.
[36] Wager TD, Atlas LY, Leotti LA, Rilling JK. Predicting individual differences in placebo analgesia: contributions of brain activity during anticipation and pain experience. J Neurosci 2011;31:439–52.
[37] Wager TD, Atlas LY, Lindquist MA, Roy M, Woo C-W, Kross E. An fMRI-based neurologic signature of physical pain. N Engl J Med 2013;368:1388–97.