PSYCHOLOGY & NEUROSCIENCE

Psychology & Neuroscience, 2014, 7, 3, 363 - 380
DOI: 10.3922/j.psns.2014.041

Automatic pain quantification using autonomic parameters

Steffen Walter[1], Sascha Gruss[1], Kerstin Limbrecht-Ecklundt[1], Harald C. Traue[1], Philipp Werner[2], Ayoub Al-Hamadi[2], Nicolai Diniz[3], Gustavo Moreira da Silva[3], and Adriano O. Andrade[3]

1- University of Ulm, Ulm, Germany

2- Otto-von-Guericke-University Magdeburg, Magdeburg, Germany

3- Universidade Federal de Uberlândia, Uberlândia, MG, Brazil

Abstract

The objective measurement of subjective, multi-dimensionally experienced pain is a problem for which there has not been an adequate solution. Although verbal methods (e.g., pain scales and questionnaires) are commonly used to measure clinical pain, they tend to lack objectivity, reliability, or validity when applied to mentally impaired individuals. Biopotential and behavioral parameters may represent a solution. Such coding systems already exist, but they are either very costly or time-consuming or have not been sufficiently evaluated. In this context, we collected a database of biopotentials to advance an automated pain recognition system, determine its theoretical testing quality, and optimize its performance. For this purpose, participants were subjected to painful heat stimuli under controlled conditions. One hundred thirty-five features were extracted from the mathematical groupings of amplitude, frequency, stationarity, entropy, linearity, and variability. The following features were chosen as the most selective: (1) electromyography corrugator peak to peak, (2) corrugator shannon entropy, and (3) heart rate variability slope RR. Individual-specific calibration allows the adjustment of feature patterns, resulting in significantly more accurate pain detection rates. The objective measurement of pain in patients will provide valuable information for the clinical team, which may aid the objective assessment of treatment (e.g., effectiveness of drugs for pain reduction, information on surgical indication, and quality of care provided to patients).
Keywords: pain quantification, heat, biopotentials, feature extraction and selection, calibration, support vector machines.

Received 11 August 2013; received in revised form 06 May 2014; accepted 10 July 2014. Available online 25 November 2014.

Introduction

Pain is a very personal sensation that is difficult to interpret without any communication from the patient. Consequently, a method for the objective measurement of pain would be beneficial, particularly in cases in which the patient is unable to describe the pain that he or she is experiencing, such as in neonates (Brahnam, Chuang, Shih, & Slack, 2006), somnolent patients, and patients who suffer from dementia (Basler et al., 2001; Zwakhalen, Hamers, Abu-Saad, & Berger, 2006; Herr, Bjoro & Decker, 2006). Under certain circumstances, little correlation exists between subjectively experienced pain and tissue lesions or other pathological changes. The pain may even be completely unrelated. Therefore, somatic pathology does not allow any conclusions to be drawn about subjectively experienced pain (Turk & Okifuji, 1999; Nilges & Traue, 2007). Children, older individuals, and patients who suffer from dementia have different pain thresholds and varying tolerance to pain relative to healthy adults (Lautenbacher, 2004; Soetanto, Chung, & Wong, 2004).

One central problem is the fact that no simple method can be used to measure pain directly. The examining physician must rely on the patient's qualitative description of the intensity, location, and nature of the pain. Quantifying pain is possible with the aid of the Visual Analog Scale or Numeric Rating Scale. However, these methods only work when the patient is sufficiently alert and cooperative, which is not always the case in the medical field (e.g., post-surgery phases). Overall, these methods are either considered inadequate or still in development (Lautenbacher, 2004). If conditions do not allow for a sufficiently valid measurement of the pain, then this may lead to cardiac stress in at-risk patients, under-perfusion of the operating field, or the development of chronic pain. For example, 30-70% of patients report moderate to severe pain after surgery (Wiebalck, Vandermeulen, Aken, & Vandermeersch, 1995). The measurement of biopotential via the autonomic nervous system may be a solution that would permit an objective, reliable, and variable diagnosis of pain.

In the area of pure research, many studies have been performed to determine correlations between the

Steffen Walter, Sascha Gruss, Kerstin Limbrecht-Ecklundt, and Harald C. Traue, Department of Psychosomatic Medicine and Psychotherapy, University of Ulm, Ulm, Germany. Philipp Werner, and Ayoub Al-Hamadi, Institute for Information Technology and Communications, Otto-von-Guericke-University Magdeburg, Magdeburg, Germany. Nicolai Diniz, Gustavo Moreira da Silva, and Adriano O. Andrade, Biomedical Engineering Laboratory (BioLab), Universidade Federal de Uberlândia, Uberlândia, Brazil. Correspondence regarding this article should be directed to: Steffen Walter – Email: steffen.walter@uni-ulm.de

autonomic nervous system (primary electrocardiogram and galvanic skin conductance) and pain stimulation (Colloca, Benedetti, & Pollo, 2006; Cortelli & Pierangeli, 2003; Jeanne, Logier, De Jonckheere, & Tavernier, 2009; Korhonen & Yli-Hankala, 2009; Ledowski, Ang, Schmarbeck, & Rhodes, 2009; Loggia, Juneau, & Bushnell, 2011; Schlereth & Birklein, 2008). However, these studies only examined the correlation between a single biopotential parameter (Treister, Kliger, Zuckerman, Goor Aryeh, & Eisenberg, 2012) and were not oriented toward applied research. To receive an objective, reliable, and valid diagnosis of pain, we need a combination of multi-parameter features. To the best of our knowledge, the study by Treister et al. (2012) was the first that took a multi-parameter biopotential approach. Tonic heat was applied to elicit pain for a duration of 1 min, with intensities of no pain, low pain, medium pain, and high pain. The pain intensities were calibrated individually. The following biopotential measurements were used: heart rate, heart rate variability-high frequency, skin conductance, number of skin conduction fluctuations, photoplethysmography, and a linear combination parameter. All of the features differed significantly in "no pain" and the other categories, only the linear combination of features significantly differentiated between all pain categories ($p < .001$ to .02). Additionally, a clinical study by the same research group (Ben-Israel, Kliger, Zuckerman, Katz, & Edry, 2013) reported similar results to those obtained with a linear regression and non-linear Random Forest regression based on the same six features used by Treister et al. (2012). Like Treister et al. (2012), the authors of the present study represent the scientific viewpoint that extracting only six features is insufficient for objective, reliable, and valid pain recognition. A clear statement about which features are the most innovative can only be made based on the simultaneous testing of a large collection of features. Furthermore, innovative applied pain recognition requires the use of modern machine learning classification methods (e.g., Neural Networks and Support Vector Machines [SVMs]).

Hence, the goal of the present study was to develop an extensive multimodal dataset (i.e., The BioVid Heat Pain Database; Walter et al., 2013a) in which varying levels of pain would be induced. This paper focuses only on the biopotentials of the multimodal dataset and not on the behavioral data. We plan to release the database for research purposes. We also used the pain heating model (e.g., Treister et al., 2012) because this model is computer-based and the best controlled pain model that can be found in the existing literature (Lautenbacher, 2004). The aim of the present study (see Figure 1A) was to select the feature patterns that contribute to the highest recognition rate for pain quantification.

The paper features several unique attributes: (1) highly controlled pain stimulation, (2) multimodal detection (i.e., simultaneous data collection on electromyogram [EMG] including zygomaticus, corrugator, and trapezius, skin conductance level [SCL], and electrocardiogram [ECG]), (3) extraction of features from the statistical groups of *amplitude*, *frequency*, *stationarity*, *entropy*, *linearity*, and *variability* (in this regard, a maximum number [$\Sigma = 135$] of features should be extracted), and (4) the selection of general statistical and individual automatic feature patterns that contribute to the highest recognition rate for pain quantification.

The overall hypothesis of the present study was that the distinction between pain quantification (baseline [*B*] *vs.* pain thresholds T_1 *vs.* T_2 *vs.* T_3 *vs.* T_4) would be significant with regard to signal features (i.e., *amplitude*, *frequency*, *stationarity*, *entropy*, *linearity*, and *variability*) of the EMG (zygomaticus, corrugator, and trapezius), ECG, and SCL. An explorative hypothesis was that there are reliable (> 80%) individual automatic (SVM) pain quantification rates (baseline [*B*] *vs.* pain thresholds T_1 *vs.* T_2 *vs.* T_3 *vs.* T_4) with regard to signal features (i.e., *amplitude*, *frequency*, *stationarity*, *entropy*, *linearity*, and *variability*) of the EMG (zygomaticus, corrugator, and trapezius), ECG, and SCL.

Methodology

Participants

A total of 90 subjects participated in the experiment, recruited from the following age groups: (1) 18-35 years ($n = 30$ years; 15 men, 15 women), (2) 36-50 years ($n = 30$; 15 men, 15 women), and (3) 51-65 years ($n = 30$; 15 men, 15 women). A total of 86 subjects were included in the final analysis because four subjects were excluded because of limited data quality with regard to the EMG. Recruitment was performed through notices posted at the university for the 18- to 35-year-old age group and through the press for the 36- to 65-year-old age group. Only healthy subjects were recruited. Pre-existing neurological conditions, chronic pain, cardiovascular disease, regular use of pain medication, and use of pain medication immediately before the experiment were applied as exclusion criteria. The subjects received an expense allowance. The study was conducted in accordance with the ethical guidelines set out in the WMA Declaration of Helsinki (ethical committee approval was granted: 196/10-UBB/bal).

Measured parameters

Biopotentials: A Nexus-32 amplifier (http://www.mindmedia.nl; accessed May 23, 2014) was used to record biopotential data (see Figure 1C) during the experiment. Biopotential and event data were recorded using Biotrace software. The following parameters were included in the classification (Walter et al., 2013a):

EMG: Electrical muscle activity is also an indicator of general psychophysiological stimulation in which increased muscle tone is associated with increasing activity of the sympathetic nervous system. A decrease in somatomotor activity reflects predominantly parasympathetic stimulation. We used two-channel EMGs for the zygomaticus, corrugator, and trapezius muscles. EMG responses via facial muscle regions such as the corrugator

supercilii, which draws the brow downward and medialward to form a frown, and the zygomaticus major, which elevates the corners of the mouth superiorly and posteriorly, are expected to be active during pain stimulation. The activity of the trapezius is an indication of a high stress level, which is also to be expected when pain is being experienced.

SCL: To measure the skin conductance level, two electrodes connected to the sensor were positioned on the index and ring fingers. Because the sweat glands are innervated exclusively sympathetically (i.e., without the influence of the parasympathetic nervous system), electrodermal activity is considered a good indicator of the "inner" tension of a person. This phenomenon can be reproduced particularly impressively by the observation of a rapid increase in skin conductance within 1-3 s due to a simple stress stimulus (e.g., deep breathing, emotional excitement, or mental activity).

ECG: We measured the average action potential of the heart on the skin using two electrodes, one on the upper right and one on the lower left of the body. Common features of the ECG signal are heart rate, interbeat interval, and heart rate variability (HRV). Heart rate variability refers to the oscillation of the interval between consecutive heartbeats. It has been used as an indication of mental effort and stress in adults (Kim & Andre, 2008).

EEG: We measured 21 EEG channels including two EOG (horizontal, vertical) channels. The EEG analysis is not presented in this paper.

Pain stimulation method

For pain elicitation we used a thermode (PATHWAY, http://www.medoc-web.com; accessed April 23, 2014) applied to the right arm (see Figure 1D). Throughout the entire experiment, the participants sat in a chair with their arms resting on the desk in front of them. With this kind of technology, eliciting quantified pain under highly controlled conditions is possible, without causing skin burns (Lautenbacher, 2004). A temperature of 50.5°C must not be exceeded.

Calibration of thresholds: At the beginning of the experiment, we tested pain (T_1) and tolerance thresholds (T_4) for every participant. During this process, the subjects sat in a chair, and a thermode was attached to their right forearm (see Figure 1D). In the left hand, they held a computer mouse. To measure T_1 and T_4, a temperature rise (10°C/s) was implemented, starting at a value of 32°C. When the threshold of T_1 and T_4 was reached, the subject clicked the right mouse button. Four measurements each for T_1 and T_4 were made for each subject. From these values, a specific average was calculated for T_1 and T_4 for each individual. Two other intermediate individual pain thresholds (T_2 and T_3) were determined mathematically:

$$T2 = \left(\frac{T4-T1}{3}\right) + T1 \text{ and } T3 = \left(\left(\frac{T4-T1}{3}\right) * 2\right) + T1.$$

Instruction for pain threshold: "Please press the stop button immediately when you experience a burning, stinging, piercing, or pulling sensation in addition to feeling heat."

Instruction for tolerance threshold: "Please press the stop button immediately when you can no longer tolerate the heat, taking into account the burning, stinging, piercing, or pulling sensation."

Pain stimulation: After the calibration procedure, we programmed the thermode software with the T_1 *vs.* T_2 *vs.* T_3 *vs.* T_4 separately for each individual for the stimulation experiment. For approximately 25 min,

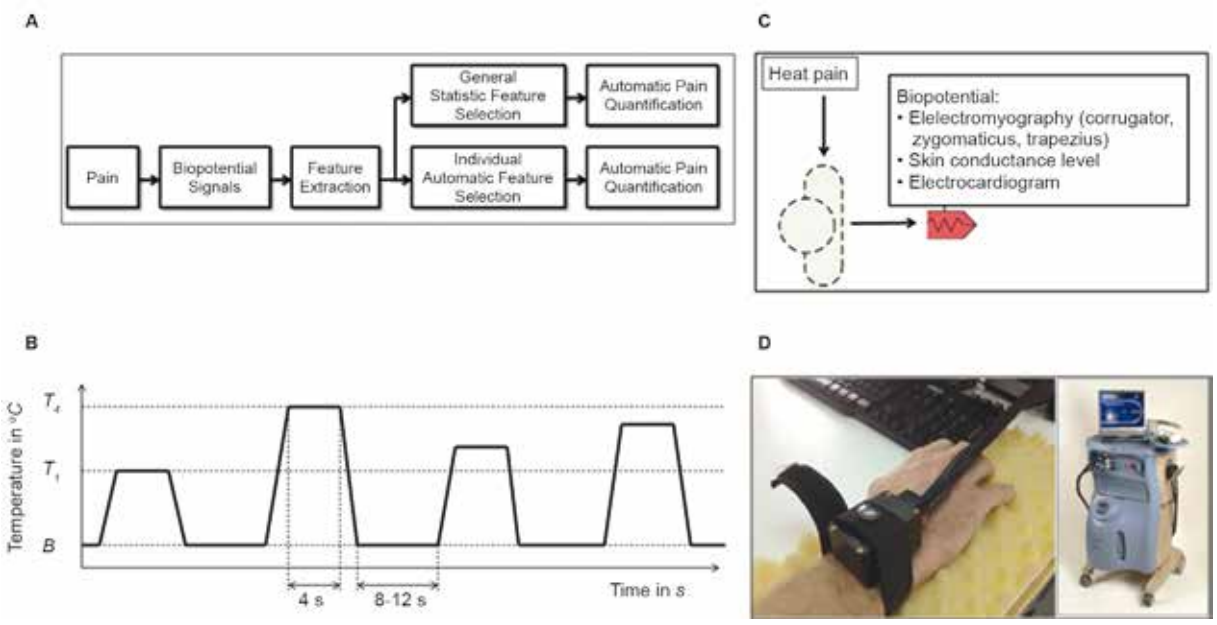


Figure 1. (A) Study procedure. (B) Heat signal with baseline *(B) vs*. pain thresholds (T_1, T_4, T_2, and T_3, from left to right). (C) Labor setting. (D) Thermode on the right arm.

we randomly stimulated the participants with the four individual specific thresholds of pain. The baseline (no pain) (*B*) was 32°C. Every pain level (T_1 *vs*. T_2 *vs*. T_3 *vs*. T_4) was applied 20 times, resulting in a total of 80 stimulations. Figure 1B illustrates a temperature plot of each stimulus and the subsequent pause. The maximum temperature for each pain threshold was maintained for 4 s. The pauses between stimuli were randomized between 8 and 12 s, and the serial heat stimulation was also randomized. The time until the thresholds (T_1 *vs*. T_2 *vs*. T_3 *vs*. T_4) were attained was proportionally equal. The subjects had the option to terminate the experiment immediately using an emergency stop button. After the experiment, the subject was asked to apply a cold pack to the site of the heat stimulation for at least 5 min.

Preprocessing

We performed the following biopotential preprocessing:

(1) We visualized the biopotentials to check the intensity of the noise and activity with regard to pain stimulation.
(2) We applied a Butterworth filter to the EMG (20-250 Hz) and ECG (.1-250 Hz) signals.
(3) We also applied an additional filter using the Empirical Mode Decomposition technique developed by Andrade, Kyberd, & Nasuto (2008).
(4) We quantified the pain level caused by the heat applied using four pain thresholds during the "pain window" (5.5 s) and with regard to the baseline during the "non-pain window" (see Figure 2).
(5) We detected bursts of activity via the EMG using the Hilbert Spectrum (Andrade, Nasuto, & Kyberd, 2007).

Feature extraction

We extracted features from the mathematical groups of (1) amplitude, (2) frequency, (3) stationarity, (4) entropy, (5) linearity, and (6) variability. To this end, the maximum information ($\Sigma = 135$) of the features was extracted systematically (Nakano, Ota, Ukai, Nakamura, & Fujita, 2002; Andrade, 2005; Cao & Slobounov, 2011; Chen, Zhuang, & Wang, 2009; Hua-Mei, Varshney, & Arora, 2003). Table 2 (see Appendix) provides a detailed overview of all of the feature information. Figure 3 contains a graph as an example for each feature group, one extreme example each for high and low manifestations.

Analysis

Pain stimulation thresholds

We used the Mann-Whitney U test to compare genders (female *vs*. male) and age groups (18-35 years *vs*. 36-50 years *vs*. 51-65 years) with regard to pain and tolerance thresholds. These analyses were performed to examine the extent to which the gender and age group comparisons were consistent with the literature (Lautenbacher, 2004; Zimmer, 2004; Basler, 2004).

Biopotential response via statistical results

All biopotentials were normalized separately for each individual signal feature. Generalized Linear Models (GLMs) were used to test the quantitative pain intensity with respect to all of the features (Nelder & Wedderburn, 1972). This model is based on the Wald χ^2 test (Wald, 1943) and the related *post hoc* test. For this purpose, a Wald χ^2 test for *B*, T_1, T_2, T_3, and T_4 (see Table 1) and five subsequent *post hoc* tests for *B vs*. T_1, T_1 *vs*. T_2, T_2 *vs*. T_3, T_3 *vs*. T_4, and *B vs*. T_4 (for details, see Table 3 in the Appendix) were performed across

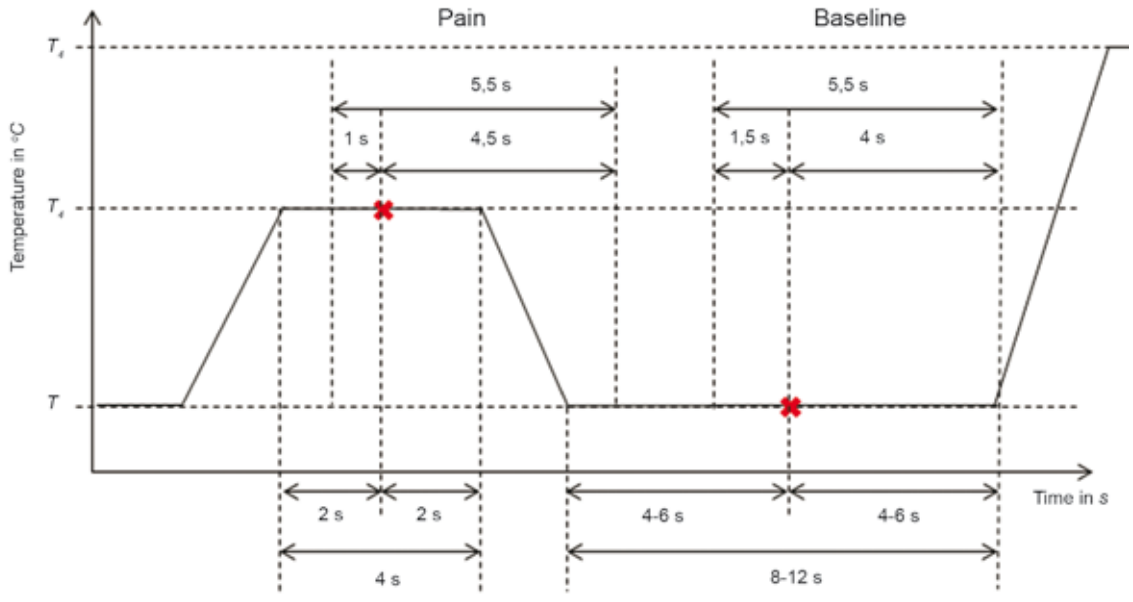


Figure 2. Pain quantification.

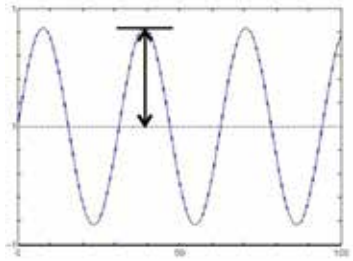
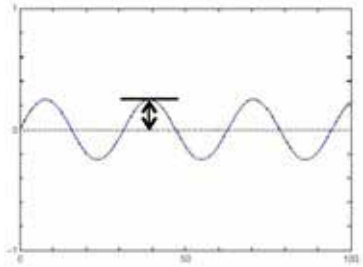
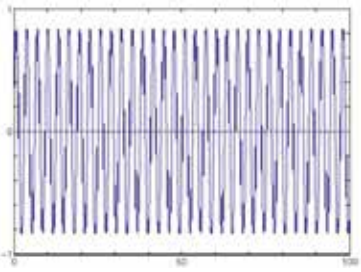
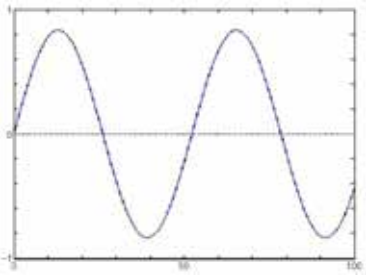
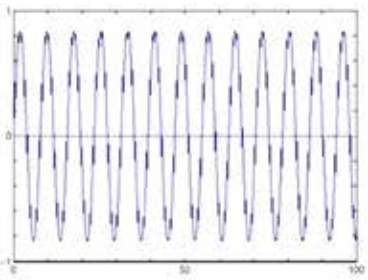
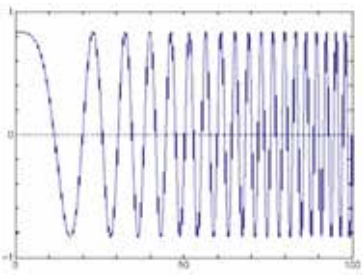
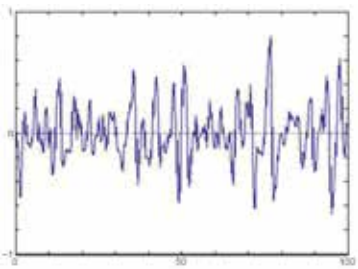
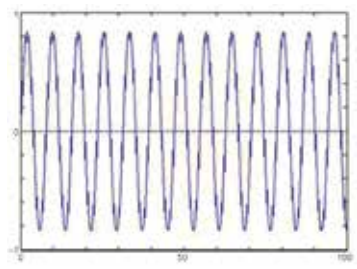
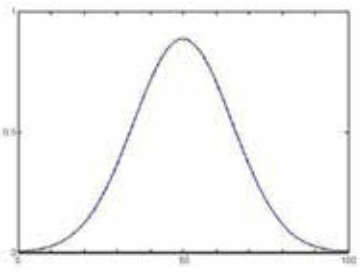
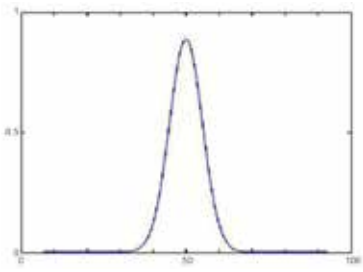
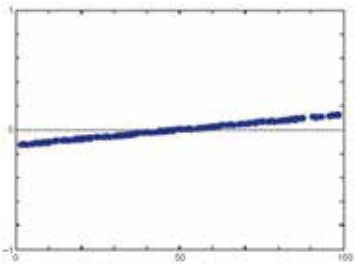
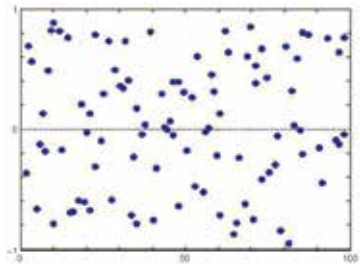
Feature category	high	low
The **amplitude** of a periodic variable is a measure of its change over a single period.		
Frequency is the number of occurrences of a repeating event per unit of time.		
Stationarity reflects some form of stable equilibrium. All statistical properties (e.g., mean, variance etc.) of a stationary time series are constant over time.		
Entropy measures the number of specific ways in which a system may be arranged, often taken to be a measure of disorder.		
Variability ("the state or characteristic of being variable") describes how spread out or closely clustered a set of data is.		
Linearity is the ability of a system always to respond to the change of a parameter with a proportional change of another parameter.		

Figure 3. Graphic examples of the feature groups: amplitude, frequency, stationarity, entropy, linearity, and variability. (Left) High expression. (Right) Low expression.

all thresholds, including the baseline. The maximum significant separation was determined for each feature (as in Treister et al., 2012). Because a total of 135 features were tested, Bonferroni correction was necessary so that only values of $p \leq .0001$ were considered significant.

Machine learning and classification

Machine learning systems are systems that learn from known data and try to recognize characteristic patterns in such data. After a learning phase, they return a model that can be used to map (i.e., classify) unknown input data into a category (Mitchell, 1997). For these classification tasks, there are several machine learners (classifiers), all of which work using different decision algorithms such as Neural Networks, Decision Trees, K-Nearest Neighbor, and SVMs.

For the classification of different pain thresholds, we chose SVMs because these have proven to be highly effective in other studies (Kapoor, Burleson, & Picard, 2007) and are capable of maintaining sufficient flexibility with regard to their main parameter optimization (Hsu, Chang, & Lin, 2003).

The goal of an SVM is to develop a predictive model from given training sets X_i and their associated class labels Y_i that can subsequently be applied to an unlabeled dataset to assign this set to a particular class. Thus, the SVM (Boser, Guyon, & Vapnik, 1992) searches for an optimal solution to the following problem:

Minimize with respect to: $w, b, \xi: \frac{1}{2} w^T w + C \sum_{i=1}^{m} \xi_i$

so that the following constraint applies:

$$y_i(w^T \phi(x_i) + b) \geq 1 - \xi_i, \forall\, 1 \leq i \leq m, \xi_i \geq 0.$$

By means of a kernel function, $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$ (in our case the radial basis function [RBF] kernel), the training sets are transformed into a higher dimensional space in which the SVM finds an optimal hyperplane with a maximum margin that separates the classes. The hyperplane then serves as a decision function for unlabeled data with unknown class allocations. For further details on the SVM, the reader may refer to Schoelkopf, Smola, Williamson, and Bartlett (2000).

Feature selection

Automatic pattern selection methods are used to further clarify recognition rates. Feature selection is a "method for selecting a subset of features providing optimal classification accuracy of the classification model" (Kolodyazhniy, Kreibig, Gross, Roth, and Wilhelm, 2011, p. 909). This is accomplished by means of a variety of feature selection (pattern configuration) methods including sequential backward search, sequential forward search, sequential floating forward search, Fisher projection, and hybrid methods, in combination with the classification (e.g., SVMs, Neuronal Networks, and K-Nearest Neighbor). The present article was limited to forward selection and backward selection.

Forward selection

"The forward selection algorithm starts with an empty set of features and adds in each round each unused feature of the given feature pool. For each added feature, the classification accuracy is estimated via cross-validation. Only the feature giving the highest increase of accuracy is added to the set. Then a new round is started with the modified set. The algorithm stops as soon as there is no increase anymore" (Akthar & Hahne, 2012).

Backward selection

"The backward selection algorithm starts with the full set of features and removes in each round each remaining feature of the given feature pool. For each removed feature, the classification accuracy is estimated via cross-validation. Only the feature giving the least decrease of accuracy is finally removed from the set. Then a new round is started with the modified set. The algorithm stops as soon as there is no increase anymore" (Akthar & Hahne, 2012).

Cross-validation

For every classification, cross-validation is necessary. Cross-validation is a "common approach for estimating the classification accuracy with unknown data. In this approach, the entire dataset is divided into N non-overlapping parts. Training and validation are performed repeatedly N times. At iteration k of the cross-validation, all parts of the data, with the exception of the kth, part are used for training, and the kth part of the data is used for validation" (Kolodyazhniy et al., 2011, p. 909). The N results from the N iterations are finally averaged to produce a single estimation. In the present paper, the performance of the individual classification procedure (Figure 4) is measured with 10-fold cross-validation, meaning that the data were partitioned into 10 parts of equal size. The performance of the general classification procedure (Figure 5) was measured with the "leave-one-subject-out" method (i.e., "in each iteration, all measurements corresponding to a particular participant are removed from the training set and used for validation"; Kolodyazhniy et al., 2011, p. 911).

Results

The following section contains the results of the (1) temperatures of the thresholds, including the relationships with gender and age, (2) biopotential response via the statistical results, and (3) biopotential response via the machine learning results.

Pain stimulation thresholds

The average temperature for the four thresholds was T_1 (M = 46.29°C, SD = 2.57°C), T_2 (M = 47.44°C, SD = 2.14°C), T_3 (M = 48.59°C, SD = 1.82°C), and T_4 (M = 49.74°C, SD = 1.73°C).

Group comparison for gender: No significant difference was found between female (M = 45.91°C,

SD = 2.59°C) and male (*M* = 46.70°C, *SD* = 2.51°C) subjects for T_1, but a significant difference was observed between female (*M* = 49.28°C, *SD* = 2.59°C) and male (*M* = 50.22°C, *SD* = .63°C) subjects for T_4 ($p \leq .001$).

Group comparison for age: A significant difference was found between age groups (18-35 years: *M* = 45.73°C, *SD* = 2.08°C; 36-50 years: *M* = 46.26°C, *SD* = 2.79°C; 51-65 years: *M* = 46.88°C, *SD* = 2.68°C) for T_1 ($p \leq .05$), but the *post hoc* tests indicated a significant difference only between the 18-35 and 51-65 groups ($p \leq .05$). No significant difference was found between age groups (18-35 years: *M* = 49.68°C, *SD* = 1.14°C; 36-50 years: *M* = 49.81°C, *SD* = 1.94°C; 51-65 years: *M* = 49.69°C, *SD* = 1.99°C) for T_4.

Biopotential response via statistical results

Table 1 summarizes the significant results ($p \leq .0001$, with Bonferroni correction) of the Wald χ^2 test for each of the 135 features. One hundred five features

Table 1. Significance test using general linear models (Wald-Chi-Quadrate) with Bonferroni correction.

	zygomaticus	corrugator	trapezius	SCL	ECG
amplitude	↗ peak ∗	↗ peak ∗	↗ peak ∗	↗ peak ∗	-
amplitude	↗ p2p ∗	↗ p2p ∗	↗ p2p ∗	↗ p2p ∗	-
amplitude	↗ rms ∗	↗ rms ∗	↗ rms ∗	↗ rms ∗	-
amplitude	↗ mlocmaxv ∗	↗ mlocmaxv ∗	↗ mlocmaxv ∗	↗ mlocmaxv ∗	-
amplitude	↙ minlocminv ∗	↙ minlocminv ∗	↙ minlocminv ∗	↗ minlocminv ∗	-
amplitude	↗ mav ∗	↗ mav ∗	↗ mav ∗	↗ mav ∗	-
amplitude	↗ mavfd ∗	↗ mavfd ∗	↗ mavfd ∗	↗ mavfd ∗	-
amplitude	↗ mavfdn ∗	↗ mavfdn ∗	↗ mavfdn ∗	↙ mavfdn ∗	-
amplitude	↗ mavdsd ∗	↗ mavdsd ∗	↗ mavdsd ∗	↗ mavdsd ∗	-
amplitude	↗ mavsdn ∗	↗ mavsdn ∗	↗ mavsdn ∗	↙mavsdn ∗	-
frequency	↗zc ∗	↗ zc ∗	↗ zc ∗	zc	-
frequency	↗ fmode ∗	↗fmode ∗	↗ fmode ∗	fmode	-
frequency	bw	bw	bw	bw	-
frequency	cf	cf	cf	cf	-
frequency	↗ fmean ∗	↗fmean ∗	↗ fmean ∗	fmean	-
frequency	↗ fmed ∗	↗fmed ∗	↗fmed ∗	↗fmed ∗	-
stationarity	↗median ∗	↗median ∗	median	median	-
stationarity	↗feqpond ∗	↗feqpond ∗	feqpond	feqpond	-
stationarity	↙area ∗	↙area ∗	area	↗area ∗	-
stationarity	↙area pond ∗	↙area pond	area pond	↗area pond ∗	-
stationarity	↗me ∗	↗me ∗	↗me ∗	↗me ∗	-
stationarity	↗sd ∗	↗sd ∗	↗sd ∗	↗sd ∗	-
entropy	↗aprox ∗	↗aprox ∗	↗aprox ∗	↙aprox ∗	-
entropy	fuzzy	fuzzy	fuzzy	↙fuzzy ∗	-
entropy	sample	sample	sample	↙sample ∗	-
entropy	↗shannon ∗	↗shannon ∗	↗shannon ∗	↗shannon ∗	-
entropy	↗spectral ∗	↗spectral ∗	spectral	↗spectral ∗	-
linearity	pldf	pldf	pldf	↗pldf ∗	-
linearity	↙ldf ∗	↙ldf ∗	ldf	ldf	-
variability	↗var ∗	↗var ∗	↗var ∗	↗var ∗	-
variability	↗std ∗	↗std ∗	↗std ∗	↗std ∗	-
variability	↗range ∗	↗range ∗	↗range ∗	↗range ∗	-
variability	↗intrange ∗	↗intrange ∗	↗intrange ∗	↗intrange ∗	-
variability	-	-	-	-	↙HRV_meanrrz ∗
variability	-	-	-	-	↙HRV_rmssdz ∗
variability	-	-	-	-	↙HRV_sloperrz ∗

HRV, heart rate variability; SCL, skin conductance level; ECG, electrocardiogram; ∗$p \leq .0001$, significant (with Bonferroni correction);
↗ baseline, minimum, level *T4*, maximum; ↙ baseline, maximum, level *T4*, minimum.

could be considered significant with regard to pain differentiation. Some features separated ascending (B ≙ (is equivalent) minimum, T_4 ≙ maximum), whereas others separated descending (B ≙ maximum, T_4 ≙ minimum). With regard to the frequency of significant separation (see Table 3 in the Appendix, last column [total]), the features of (1) Corrugator_Amplitude_p2p, (2) Corrugator_Entropy_shannon, and (3) HRV_slopeRR were the most selective (five significant pain differentiations) in distinguishing between pain thresholds.

With regard to the four significant pain differentiations, all 10 amplitude features of the zygomaticus (Zygomaticus_Variance: var, std, range, intrage), and all 10 amplitude features of the corrugator except for p2p (Corrugator_Variance: var, std, range, intrage; Corrugator_Frequency_zc; Corrugator_Stationarity_sd; SCL_Amplitude: mavfdn, mavsdn; SCL_Stationarity: me, sd; SCL_Entropy: aprox, fuzzy, sample; SCL_Variability_range) were selected. In total, we found 41 top features.

No calculations were possible with regard to the features of SCL_Frequency (zc, fmode, bw, cf).

Biopotential response via machine learning results

Figure 4 presents the mean of all individual (10-fold cross-validation) automatic classification results via automatic feature selection for B *vs.* T_1, T_1 *vs.* T_2, T_2 *vs.* T_3, T_3 *vs.* T_4, and B *vs.* T_4. The detection rates were 88.79-94.73% for forward selection and 59.44-81.75% for backward selection.

We found highly significant results ($p \leq .0001$) in the comparison (Wilcoxon signed-rank test) between forward selection recognition and backward selection recognition rates via every threshold contrast: B *vs.* T_1, T_1 *vs.* T_2, T_2 *vs.* T_3, T_3 *vs.* T_4, and B *vs.* T_4.

Because of the extremely high level of individual specificity of the patterns, providing a frequency diagram of the forward and backward selection was not possible because this would clearly exceed the scope of this article.

Figure 5 presents the general automatic classification (leave-one-subject-out method) results of 52.41-74.59% for the general statistical features (1) Corrugator_Amplitude_p2p, (2) Corrugator_Entropy_Shannon, and (3) HRV_slopeRR. For the top 41 features, we found classification results of 53.49-77.05%. A comparison between the top three *vs.* top 41 analyses was not possible because there were no individual means.

Discussion and conclusion

We have presented a newly collected multimodal dataset (BioVid Heat Pain Database; Walter, 2013a) to facilitate advances in the reliable recognition of pain intensity. The higher-level pragmatic orientation of this research ultimately allows the objective, reliable,

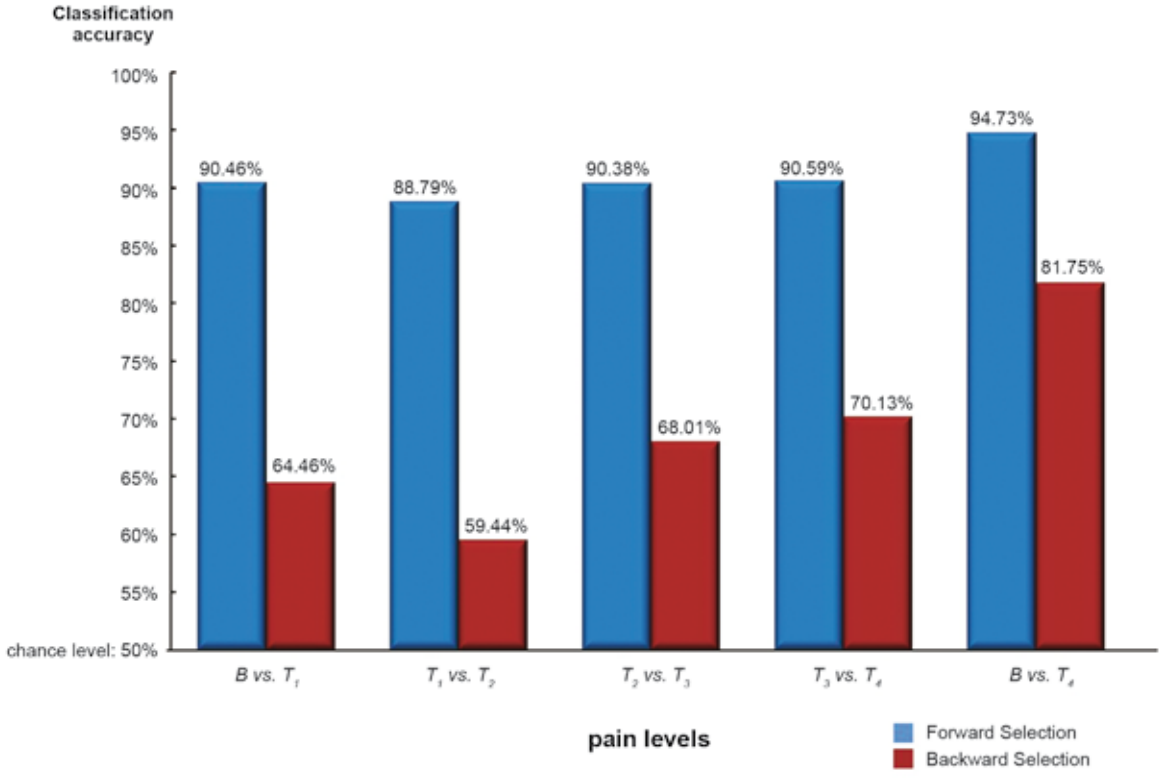


Figure 4. Mean automatic individual (10-fold cross-validation) classification results across baseline and four thresholds (B *vs.* T_1, T_1 *vs.* T_2, T_2 *vs.* T_3, T_3 *vs.* T_4, and B *vs.* T_4) for amplitude, frequency, stationarity, entropy, linearity, and variability features ($\sum = 135$) with automatic feature selection. Blue ≙ forward selection; brown, ≙ backward selection.

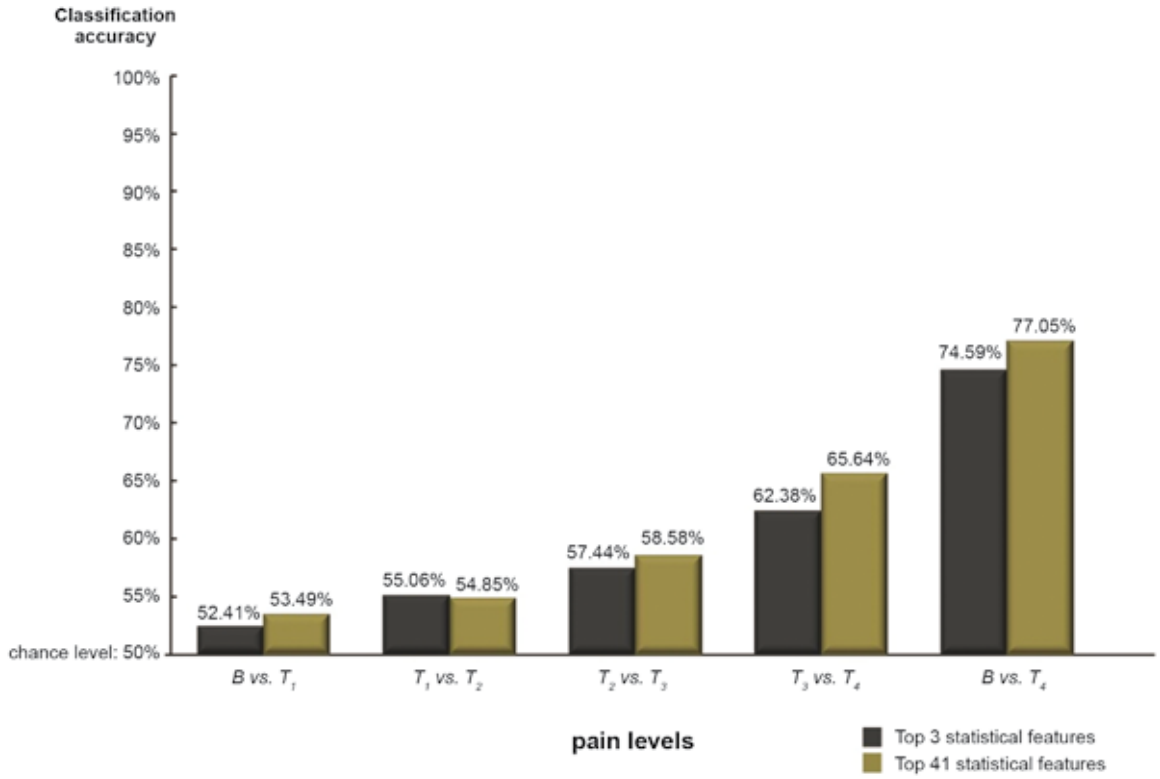


Figure 5. Automatic general classification (leave-one-subject-out method) results across baseline and four thresholds (*B vs. T_1*, *T_1 vs. T_2*, *T_2 vs. T_3*, *T_3 vs. T_4*, and *B vs. T_4*); Black ≙ only for (1) Corrugator_ Amplitude_ p2p, (2) Corrugator_Entropie_Shannon, and (3) HRV_slopeRR; Olive ≙ 41 top features.

and valid recognition of pain in infants, people who suffer from dementia, and people with limited verbal communication skills. The authors of the present article consider the approach of Treister et al. (2012) and Ben-Israel et al. (2013) as straightforward but insufficient in terms of objective, reliable, and valid clinic pain recognition. Therefore, we extracted features of the highly complex statistical groups *amplitude*, *frequency*, *stationarity*, *entropy*, *linearity*, and *variability* and selected the feature patterns (general statistic and individual automatic) that contributed to the highest recognition rate for pain quantification.

Discussion of results

Pain stimulation thresholds: The thresholds T_1, T_2, T_3, and T_4, including the effects of age and gender, are consistent with the results reported in the literature (Lautenbacher, 2004; Zimmer, 2004; Basler, 2004).

Biopotential response via statistical results: A very low p level ($p \leq .0001$) was used for the Bonferroni correction and selection criteria by taking into account the most selective features. With regard to our statistical procedure, the majority of features are generally suitable for ensuring the quantification of pain. Similar to Treister et al. (2012), we used the frequency of significant separation between pain thresholds as a selection criterion. The features Corrugator_ Amplitude_p2p, Corrugator_ Entropy_shannon, and HRV_slopeRR can be regarded as the most selective, in which they significantly distinguished among five pain differentiations. Furthermore, the features that resulted from the mathematical groupings of amplitude and variability in conjunction with zygomaticus and corrugator tended to be suitable. In the SCL, selectivity with respect to pain quantification and mathematical grouping was more complex and less clear. The features in the areas of linearity, stationarity, variability, and frequency can only be regarded as satisfactory. These features can be assumed to have greater significance with regard to the duration or nature of pain (e.g., stabbing, pulling, throbbing, sharp, tearing, etc.).

Biopotential response via machine learning results: Using automatic feature selection (forward selection and backward selection) with SVMs, we tested the extent to which individual-specific automatic feature selection is beneficial. The pattern configurations are evidently extremely individual-specific, accompanied by very high recognition rates, especially in forward selection (> 88%). Precisely distinguishing between all four thresholds is possible. The high individual-specific pattern configuration is consistent with intense individual-specific stress regulation according to fundamental research (e.g., Stemmler & Wacker, 2010).

Although we are aware of the advantages and disadvantages of our chosen feature selection algorithms, we presently have no adequate explanation why forward selection outperforms backward selection. A frequency

diagram of typical individual-specific patterns was not productive because the patterns significantly differed from each other.

The recognition rates regarding the general classification with only three features (Features Corrugator_Amplitude_p2p, Corrugator_Entropy_shannon, and HRV_slopeRR) were obviously less compared with individual rates. Pain tolerance and baseline could be recognized by 74.59% (top 3) and 77.05% (top 41), meaning we are high about chance level. In a two class problem (what we have used) it means always about 50%). With regard to our calculated automatic classification rates, there are currently no comparable studies in the area of automatic pain recognition. However, our results are in line with comparisons of high vs. low arousal in the Affective Computing research area (Kim & Andre, 2008; Walter et al., 2011; Walter, Kim, Hrabal, Crawcour, Kessler, & Traue, 2013b). The comparison of the top 3 vs. top 41 features shows more about 3 top features make the results not really relevant better.

Comparison: statistics and machine analysis—summary

We are unaware of any studies in which conservative statistical methods were compared with modern automatic classification algorithms with regard to empirical pain induction, indicating a lack of "state-of-the-art" methods. We sought to make this comparison and will pursue it further. The EMG features appeared to make a significant contribution to the quantification of pain.

In terms of the statistical results, a general feature pattern was detected, but the individual-specific classification rates showed that detection rates can be significantly improved through individual-specific calibration.

For pain recognition in clinical practice, future pain recognition algorithms may have initial default features, such as (1) Corrugator_Amplitude_p2p, (2) Corrugator_Entropie_Shannon, and (3) HRV_slopeRR. Individual-specific calibration allows for an adjustment of feature patterns, resulting in significantly more accurate pain detection rates.

Outlook

Numerous additional analyses will be performed using the described dataset (Walter et al., 2013a). Specifically, a data fusion of biopotentials with video signals (i.e., facial expressions and gestures) that have been recorded three-dimensionally (Walter et al., 2013a) is planned. Early, intermediary, and late fusions are being tested for the data fusion (Schwenker, Dietrich, Thiel, & Palm, 2006; Schwenker, Dietrich, Kestler, Rieder, & Palm, 2003. The features presented in the present article must be investigated using other models in terms of the duration and type of pain induction.

There are plans for a clinical project in which detection will occur postoperatively in humans. Multimodal signals with biomedical, visual, and paralinguistic (e.g., sighing) parameters will be measured. Highly complex pain logs will be created to allow for pain quantification. Pain detection will be further clarified by means of data fusion.

Generally, we would like to point out that the development of technology for the detection of pain always requires a multimodal approach with a maximum dimensionality of features. Within this context, it is crucial that the extracted feature configurations are logically comprehensible and clearly structured. Methodological benchmarks are urgently needed.

Acknowledgements

This research was part of the DFG/TR233/12 "Advancement and Systematic Validation of an Automated Pain Recognition System on the Basis of Facial Expression and Psychobiological Parameters" project, funded by the German Research Foundation, FAPEMIG, CNDq, CAPES, and the Brazilian government.

References

Andrade, A. O, Kyberd, P., & Nasuto, S. J. (2008). The application of the Hilbert spectrum to the analysis of electromyographic signals. *Information Sciences*, 2176–2193.

Andrade, A. O., Nasuto, S. J., & Kyberd, P. (2007). Extraction of motor unit action potentials from electromyographic signals through generative topographic mapping. *Journal of the Franklin Institute*, *344*(3-4), 154-179.

Andrade, A. O., *Decomposition and Analysis of Electromyographic Signals*, 3 Doutorado, University of Reading, UR-England 2005.

Akthar, F., & Hahne, C. (2012). RapidMiner 5 Operator Reference. Retrieved from http://rapidminer.com/documentation/; accessed April 23, 2014.

Basler, H. D., Bloem, R., Casser, H. R., Gerbershagen, H. U., Grießinger, N., Hankemeier, U., Weiß, L. (2001). Ein strukturiertes Schmerzinterview für geriatrische Patienten. *Der Schmerz*, *15*(3), 164-171.

Ben-Israel, N., Kliger, M., Zuckerman, G., Katz, Y., & Edry, R. (2013). Monitoring the nociception level: a multi-parameter approach. *Journal of Clinical Monitoring and Computing*, *27*(6), 659-668.

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). *A training algorithm for optimal margin classifiers*. In D. Haussler, editor, 5th Annual ACM Workshop on COLT, 144-152, Pittsburgh, PA, ACM Press.

Brahnam, S., Chuang, C. F., Shih, F., & Slack, M. (2006). SVM classification of neonatal facial images of pain. In: I. Bloch, A. Petrosino, & A. B. Tettamanzi (Eds.), *Fuzzy Logic and Applications: 6th International Workshop, WILF 2005, Crema, Italy, September 15-17, 2005: revised selected papers* (series title: *Lecture notes in computer science*, vol. 3849; 121-128. Berlin: Springer.

Cao, C., & Slobounov, S. (2011). Application of a novel measure of EEG non-stationarity as 'Shannon entropy of the peak frequency shifting' for detecting residual abnormalities in concussed individuals. *Clinical Neurophysiology*, *122*, 1314-1321.

Chen, W. T., Zhuang, J., Yu, W. X., & Wang, Z. Z. (2009). Measuring complexity using FuzzyEn, ApEn, and SampEn. *Medical Engineering and Physics*, *31*(1), 61-68.

Colloca, L., Benedetti, F., & Pollo, A. (2006). Repeatability of autonomic responses to pain anticipation and pain stimulation. *European Journal of Pain*, *10*(7), 659-665.

Cortelli, P., & Pierangeli, G. (2003). Chronic pain-autonomic interactions. *Neurological Sciences*, *24*(Suppl. 2), S68-S70.

Herr, K., Bjoro, K., & Decker, S. (2006). Tools for assessment of pain in nonverbal older adults with dementia: a state-of-the-science review. *Journal of Pain Symptom Management*, *31*(2), 170-192.

Hofmann, T., Scholkopf, B., & Smola, A. J. (2008). Kernel methods in machine learning. *Annals of Statistics*, *36*(3), 1171-1220.

Hsu, C. W., Cang, C. C., & Lin, C. J. (2003). A practical guide to support vector classification. Online in Internet: http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf

Hua-Mei, C., Varshney, P. K., & Arora, M. K. (2003). Performance of mutual information similarity measure for registration of multitemporal remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, *41*(11), 2445-2454.

Jeanne, M., Logier, R., De Jonckheere, J., & Tavernier, B. (2009). Validation of a graphic measurement of heart rate variability to assess analgesia/nociception balance during general anesthesia. *Conference Proceedings - IEEE Engineering in Medicine and Biology Society, 2009*, 1840-1843.

Kapoor, A., Burleson, W., Picard, R. W. (2007). Automatic prediction of frustration. *International Journal of Human-Computer Studies*, 65(8), 724-736.

Kim, J., & Andre, E. (2008). Emotion recognition based on physiological changes in music listening. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *30*(12), 2067-2083.

Kolodyazhniy, V., Kreibig, S. D., Gross, J. J., Roth, W. T., & Wilhelm, F. H. (2011). An affective computing approach to physiological emotion specificity: toward subject-independent and stimulus-independent classification of film-induced emotions. *Psychophysiology*, *48*(7), 908-922.

Korhonen, I., & Yli-Hankala, A. (2009). Photoplethysmography and nociception. *Acta Anaesthesiologica Scandinavica*, *53*(8), 975-985.

Ledowski, T., Ang, B., Schmarbeck, T., & Rhodes, J. (2009). Monitoring of sympathetic tone to assess postoperative pain: skin conductance vs surgical stress index. *Anaesthesia, 64*(7), 727-731.

Loggia, M. L., Juneau, M., & Bushnell, M. C. (2011). Autonomic responses to heat pain: Heart rate, skin conductance, and their relation to verbal ratings and stimulus intensity. *Pain, 152*(3), 592-598.

Lautenbacher, S. (2004). Schmerzmessung. In: H.D. Basler, C. Franz, B. Kröner-Herwig, & H.P. Rehfisch (Eds.), *Psychologische schmerztherapie* (271-288). Berlin: Springer.

Mitchell, T. (1997). *Machine learning*. London: McGraw-Hill.

Nakano, K., Ota, Y., Ukai, H., Nakamura, K., & Fujita, H. (2002, 24-28 June). *Frequency Detection Method Based on Recursive DFT Algorithm.* Paper presented at the 14th PSCC, Sevilla.

Nelder, J. A., & Wedderburn, (1972). Generalized Linear Models. *Journal of the Royal Statistical Society Series A (General), 135*(3), 370-384.

Nilges, P., & H.C., T. (2007). Psychologische Aspekte des Schmerzes. *Verhaltenstherapie & Verhaltensmedizin, 28*(3), 302-322.

Schlereth, T., & Birklein, F. (2008). The sympathetic nervous system and pain. *Neuromolecular Med, 10*(3), 141-147.

Schoelkopf, B., Smola, A. J., Williamson, R. C., & Bartlett, P. L. (2000). New Support Vector Algorithms. *Neural Computation, 12*(5), 1207-1245.

Schwenker, F., Dietrich, C., Thiel, C., & Palm, G. (2006). Learning of decision fusion mappings for pattern recognition. *Journal on Artificial Intelligence and Machine Learning (AIML), 6*, 17-21.

Schwenker, F., Dietrich, C., Kestler, H. A., Riede, K., & Palm, G. (2003). Radial basis function neural networks and temporal fusion for the classification of bioacoustic time series. *Neurocomputing, 51*(0), 265-275.

Soetanto, A. L., Chung, J. W., & Wong, T. K. (2004). Gender differences in pain perception: a signal detection theory approach. *Acta Anaesthesiologica Taiwanica*, *42*(1), 15-22.

Stemmler, G., & Wacker, J. (2010). Personality, emotion, and individual differences in physiological responses. *Biological Psychology*, *84*(3), 541-551.

Treister, R., Kliger, M., Zuckerman, G., Goor Aryeh, I., & Eisenberg, E. (2012). Differentiating between heat pain intensities: the combined effect of multiple autonomic parameters. *Pain*, *153*(9), 1807-1814.

Turk, D. C., & Okifuji, A. (1999). Assessment of patients' reporting of pain: an integrated perspective. *Lancet*, *353*(9166), 1784-1788.

Wald, A. (1943). On a statistical generalization of metric spaces. *Proceedings of the National Academy of Sciences of the United States of America*, *29*(6), 196-197.

Walter, S., Gruss, S., Ehleiter, H., Tan, J., Traue, H. C., Werner, P., Al-Hamadi, A., Crawcour, S., Andrade, A. O. & Moreira da Silva, G. (2013a). *The BioVid Heat Pain Database: data for the advancement and systematic validation of an automated pain recognition system*. Paper presented at the IEEE International Conference on Cybernetics, Lausanne, Switzerland, June 13-15, 2013.

Walter, S., Kim, J., Hrabal, D., Crawcour, S. C., Kessler, H., & Traue, H. C. (2013b). Transsituational individual-specific biopsychological classification of emotions. *IEEE Transactions on Systems Man Cybernetics-Systems*, *43*(4), 988-995.

Walter, S., Scherer, S., Schels, M., Glodek, M., Hrabal, D., Schmidt, M., Böck, R., Limbrecht, K., Traue, H. & Schwenker, F. (2011). Multimodal emotion classification in naturalistic user behavior. *Human-Computer Interaction: Towards Mobile and Intelligent Interaction Environments, Part III*, *6763*, 603-611.

Wiebalck, A., Vandermeulen, E., Aken, H. V., & Vandermeersch, E. (1995). Ein konzept zur verbesserung der postoperativen schmerzbehandlung. *Der Anaesthesist*, *44*(12), 831-842.

Zimmer, C. (2004). Schmerz und geschlecht. In: H. D. Basler, C. Franz, B. Kröner-Herwig, & H. P. Rehfisch (Eds.), *Psychologische schmerztherapie* (203-215). Berlin: Springer.

Zwakhalen, S. M., Hamers, J. P., Abu-Saad, H. H., & Berger, M. P. (2006). Pain in elderly people with severe dementia: a systematic review of behavioural pain assessment tools. *BMC Geriatrics*, *6*, 3.

Appendix

Table 2. Feature information and mathematical equation.

Number	Mathematical group	Feature	Equation / Description
1	amplitude	peak	peak = max(signal); index(max(signal))
2	amplitude	p2p	p2p = max(signal) - min(signal)
3	amplitude	rms	rms = rms(signal)
4	amplitude	mlocmaxv	maxlocmaxv = mean(locmax(signal))
5	amplitude	minlocminv	minlocMinV = mean(locmix(signal))
6	amplitude	mav	mav = mav(signal)
7	amplitude	mavfd	mavfd = mavfd(signal)
8	amplitude	mavfdn	mavfdn = mavfdn(signal)
9	amplitude	mavsd	mavsd = mavsd(signal)
10	amplitude	mavsdn	mavsdn = mavsdn(signal)
11	frequency	zc	Calculated by comparing each point of the signal with the next; if there is a crossing by zero, then it is accounted.
12	frequency	fmode	This fast Fourier transformation equation is valid for this and the following frequency features $X(k) = \sum_{j=1}^{N} x(j)\omega_N{}^{(j-1)(k-1)}$ Where $\omega_N = e^{(\frac{-2\pi i}{N})}$ To find the mode, find the maximum value of X.
13	frequency	bw	To obtain the bandwidth of a signal, find the first and the last frequencies where the spectral density values, X(kl) and X(kh), are approximately .707*X(kmax), where X(kmax) is the maximum value of X. Finally, the bandwidth value is the subtraction of the frequency of kh(fh) by the frequency of kl(fl).
14	frequency	cf	The central frequency is simply the mean of the frequencies that delimit the bandwidth. $cf = \frac{fh - fl}{2}$
15	frequency	fmean	$\frac{\sum_{k=1}^{NFFT} X(k).f(k)}{\sum X(k)}$
16	frequency	fmed	To obtain the median frequency, find the value of the frequency that bisects the area below the X waveform.
17	stationarity	median	$DS = \frac{1}{T}\int_0^T (1 - \frac{H(\omega,t)}{h(\omega)/T})^2 dt$ where $H(\omega, t)$ is the value of the spectrogram for frequency ω and time t, and $h(\omega)$ is the spectral density for frequency ω.
18	stationarity	freqpond	
19	stationarity	area	
20	stationarity	area ponderada	
21	stationarity	me	Given the signal x, split it into x1, x2, ... xn where $n = \frac{T}{Ti}$, with T as the total time length of the signal, which is $\frac{N}{Fs}$, and Ti the time of each part xi. For each xi, compute the mean, then the standard deviation of the resultant mean vector.
22	stationarity	sd	Use the same split logic as in the previous feature. For each xi, compute the standard deviation, then the standard deviation of the resultant standard deviation vector.

Table 2. Feature information and mathematical equation.

Number	Mathematical group	Feature	Equation / Description
23	entropy	aprox	For a temporal series with N samples {u(i): 1 ≤ i ≤ N} given *m,* create vectors X_j^m, for each X_{N-m+1}^m as X_j^m = {u (i), u (i + 1),..., u (i + m - 1)}, i = 1,..., N - m +1. where *m* is the number of points to group together for the comparison. For each k ≤ *N* – *m* + 1 groups, do C_k^m (r) which is the number of times the groups had distance less than tolerance r. Then compute the value $\varnothing^m$ as $\emptyset^m(r) = \frac{\sum_{i=1}^{N-m+1} ln\ C_i^m\ (r)}{(N-m+1)}$ The Approximated Entropy is: ApEn(m,r) = $\lim_{N\to\infty}$ [$\varnothing^m$ (*r*) – $\varnothing^{m+1}$ (*r*)]
24	entropy	fuzzy	$Saen(m,s,d) = ln\left[\frac{Com(s)}{Co(m+1)(s)}\right]$ where: 'm' is window size. 's' is the similarity standard. 'd' signal. It is calculated in a very similar way to the Sample Entropy. The only similarity between the groups is computed by means of a Fuzzy membership function.
25	entropy	sample	$Saen(m,s,d) = ln\left[\frac{Cm(s)}{C(m+1)(s)}\right]$ where: 'm' is window size. 's' is the similarity standard. 'd' signal. 'Cm' is the regularity or frequency of similar windows in a given set of windows 'd' with length 'm', obeying 's' tolerance.
26	entropy	shannon	$H = -\sum P_k\ \log P_k$ where Pk is the probability of a value for each value present in a signal.
27	entropy	spectral	S = $\Sigma\ p_k \log p_k / \log (N)$ where pk is the spectral density estimation of each fk frequency.
28	linearity	pldf	$t\sqrt{R^2_{o\,(k)1,\ldots K-1.}}$ $R^2_{0(k)1,\ldots K-1} = \frac{SQR_{1,\ldots K-1} - SQR(1,\ldots k)}{SQR_{1,\ldots k-1}}$
29	linearity	ldf	$FDD_k = t\sqrt{R^2_{o\,(k)}}$ $R^2_{0(k)} = \frac{SQT_0 - SQR_{(k)}}{SQT_0}$
30	variability	var	$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N-1}$
31	variability	std	$S = \sqrt{\sigma^2}$
32	variability	range	*R* = *MAX* (*U*) – *MIN* (*U*)
33	variability	intrange	$SI = \frac{Q_3 - Q_1}{2}$
34	variability	HRV_ meanRR	*meanRR = mean(hr_RR_vector)*
35	variability	HRV_mssd	$rmssd = \sqrt{\frac{1}{N-1}(\sum_{i-1}^{N-1}(RR_i - RR_{i-1})^2)}$
36	variability	HRV_slopeRR	*slopeRR = regression(x,hr_RR_vector)*

Table 3. Significance test of general linear models with *post hoc* test and mean (*MD*) and standard deviations (*SD*) with *z* values for pain levels: *B*, T_1, T_2, T_3, and T_4 with Bonferroni correction, maximum significant tests ≙ total.

Signal	Group	Feature	Pain levels														
			B	*B*	T_1	T_1	T_2	T_2	T_3	T_3	T_4	T_4	*B*, T_{1-4}	*B vs.* T_1	T_1 *vs.* T_2	T_2 *vs.* T_3	T_3 *vs.* T_4	*B vs.* T_4	Total
			MD	*SD*	*MD*	*SD*	*MD*	*SD*	*MD*	*SD*	*MD*	*SD*	Wald-χ^2	*post hoc p* level					
zygomaticus	a	peak	-.200	.727	-.144	.921	-.086	.828	.087	.959	.714	1.410	.000	.094	.088	.000	.000	.000	4
zygomaticus	a	p2p	-.200	.730	-.148	.914	-.085	.828	.086	.974	.737	1.431	.000	.129	.067	.000	.000	.000	4
zygomaticus	a	rms	-.188	.752	-.139	.929	-.098	.789	.088	1.009	.706	1.412	.000	.152	.238	.000	.000	.000	4
zygomaticus	a	mlocmaxv	-.177	.758	-.127	.944	-.096	.787	.081	1.012	.658	1.415	.000	.150	.370	.000	.000	.000	4
zygomaticus	a	minlocminv	.177	.757	.127	.950	.100	.782	-.078	1.003	-.669	1.425	.000	.154	.424	.000	.000	.000	4
zygomaticus	a	mav	-.179	.760	-.130	.946	-.101	.783	.079	1.019	.672	1.404	.000	.158	.404	.000	.000	.000	4
zygomaticus	a	mavfd	-.178	.764	-.132	.931	-.103	.781	.076	1.003	.661	1.413	.000	.181	.389	.000	.000	.000	4
zygomaticus	a	mavfdn	-.115	1.010	-.099	.981	-.034	.961	.095	.953	.264	.920	.000	.619	.048	.000	.000	.000	4
zygomaticus	a	mavsd	-.175	.771	-.133	.924	-.102	.786	.074	1.001	.643	1.405	.000	.217	.366	.000	.000	.000	4
zygomaticus	a	mavsdn	-.105	1.000	-.099	.966	-.032	.953	.092	.972	.221	.955	.000	.859	.044	.000	.000	.000	4
zygomaticus	f	zc	-.155	.945	-.109	.993	-.042	.932	.060	.937	.493	1.087	.000	.168	.045	.002	.000	.000	3
zygomaticus	f	fmode	-.102	1.010	-.073	.950	-.035	.939	.082	.962	.190	.978	.000	.373	.258	.000	.001	.000	3
zygomaticus	f	bw	-.037	.973	-.020	.984	-.016	.973	.053	1.030	.070	1.031	.003	.621	.910	.042	.623	.002	0
zygomaticus	f	cf	-.037	.973	-.020	.984	-.016	.973	.053	1.030	.070	1.031	.003	.621	.910	.042	.623	.002	0
zygomaticus	f	fmean	-.102	.979	-.088	.963	-.041	.921	.104	.997	.213	1.003	.000	.670	.154	.000	.001	.000	3
zygomaticus	f	fmed	-.106	.990	-.101	.960	-.034	.936	.101	.992	.185	.980	.000	.881	.043	.000	.012	.000	3
zygomaticus	s	median	-.066	1.062	-.083	.989	.003	.973	.029	.896	.105	.952	.000	.625	.010	.434	.023	.000	2
zygomaticus	s	feqpond	-.079	1.075	-.092	.992	.006	.951	.037	.905	.137	.930	.000	.706	.003	.349	.003	.000	2
zygomaticus	s	area	.047	1.077	.029	.972	-.037	.907	-.041	.852	-.164	.573	.000	.568	.028	.897	.000	.000	3
zygomaticus	s	area pond	.038	1.057	.006	.901	-.033	.898	-.039	.848	-.126	.429	.000	.312	.180	.825	.003	.000	2
zygomaticus	s	me	-.104	.773	-.043	.994	-.072	.829	.074	1.075	.256	1.153	.000	.066	.375	.000	.000	.000	3
zygomaticus	s	sd	-.181	.772	-.138	.906	-.097	.804	.103	.992	.698	1.437	.000	.208	.237	.000	.000	.000	3
zygomaticus	e	aprox	-.077	1.006	-.076	1.001	-.004	.995	.049	.948	.244	.940	.000	.976	.030	.113	.000	.000	3
zygomaticus	e	fuzzy	.011	1.055	-.005	1.039	.014	1.002	-.033	.914	-.030	.850	.472	.646	.581	.159	.937	.217	0
zygomaticus	e	sample	.012	1.061	-.012	1.016	.016	1.000	-.035	.916	-.025	.863	.452	.459	.399	.123	.748	.264	0
zygomaticus	e	shannon	-.158	1.017	-.119	1.018	-.048	.996	.106	.924	.452	.877	.000	.242	.030	.000	.000	.000	4
zygomaticus	e	spectral	-.099	1.054	-.066	1.008	-.001	.908	.085	.920	.153	.897	.000	.311	.049	.008	.040	.000	2

zygomaticus	l	pldf	.035	1.129	-.055	.706	-.051	.814	-.015	.884	-.011	1.056	.054	.008	.904	.281	.921	.173	0
zygomaticus	l	ldf	.073	1.020	.088	.961	.022	.968	-.081	.994	-.136	.972	.000	.611	.050	.002	.099	.000	2
zygomaticus	v	var	-.140	.698	-.103	.909	-.094	.698	.054	1.047	.584	1.580	.000	.297	.811	.000	.000	.000	4
zygomaticus	v	std	-.192	.746	-.146	.920	-.098	.790	.094	1.010	.709	1.413	.000	.178	.169	.000	.000	.000	4
zygomaticus	v	range	-.204	.714	-.154	.901	-.084	.830	.089	.974	.740	1.434	.000	.142	.040	.000	.000	.000	4
zygomaticus	v	intrange	-.157	.770	-.107	.974	-.090	.809	.060	1.009	.564	1.408	.000	.147	.630	.000	.000	.000	4
corrugator	a	peak	-.258	.689	-.169	.848	-.068	.876	.194	1.017	.920	1.510	.000	.011	.004	.000	.000	.000	4
corrugator	a	p2p	-.270	.655	-.177	.826	-.069	.884	.206	1.027	.930	1.491	.000	.000	.002	.000	.000	.000	5
corrugator	a	rms	-.257	.721	-.161	.874	-.089	.841	.194	1.021	.892	1.500	.000	.006	.039	.000	.000	.000	4
corrugator	a	mlocmaxv	-.244	.760	-.153	.881	-.092	.857	.179	1.002	.853	1.509	.000	.010	.088	.000	.000	.000	4
corrugator	a	minlocminv	.246	.745	.152	.880	.089	.863	-.181	1.013	-.854	1.506	.000	.008	.072	.000	.000	.000	4
corrugator	a	mav	-.249	.743	-.154	.883	-.093	.837	.179	.998	.861	1.526	.000	.007	.086	.000	.000	.000	4
corrugator	a	mavfd	-.243	.748	-.146	.894	-.092	.829	.173	.998	.848	1.529	.000	.006	.128	.000	.000	.000	4
corrugator	a	mavfdn	-.203	.994	-.100	.988	-.024	.976	.130	.941	.345	.959	.000	.002	.022	.000	.000	.000	4
corrugator	a	mavsd	-.238	.756	-.141	.900	-.088	.831	.167	.995	.834	1.531	.000	.006	.140	.000	.000	.000	4
corrugator	a	mavsdn	-.203	.974	-.095	.988	-.018	.972	.119	.949	.357	.991	.000	.001	.020	.000	.000	.000	4
corrugator	f	zc	-.208	.892	-.153	.926	-.020	.951	.139	.957	.545	1.120	.000	.096	.000	.000	.000	.000	4
corrugator	f	fmode	-.155	.969	-.065	1.010	.012	1.044	.069	.924	.240	.884	.000	.007	.019	.083	.000	.000	3
corrugator	f	bw	-.034	1.003	.032	1.050	.001	1.003	-.006	.943	.032	.920	.233	.048	.354	.835	.258	.048	0
corrugator	f	cf	-.034	1.003	.032	1.050	.001	1.003	-.006	.943	.032	.920	.233	.048	.354	.835	.258	.048	0
corrugator	f	fmean	-.157	.975	-.062	1.003	-.008	.991	.072	.938	.328	.952	.000	.004	.102	.017	.000	.000	3
corrugator	f	fmed	-.158	1.001	-.078	1.021	-.002	.978	.093	.938	.338	.953	.000	.017	.022	.005	.000	.000	3
corrugator	s	median	-.038	1.007	-.021	1.068	-.015	1.002	.076	.944	.197	.971	.000	.632	.854	.008	.000	.000	3
corrugator	s	feqpond	-.056	1.022	-.030	1.060	-.005	1.004	.074	.929	.228	.952	.000	.429	.469	.020	.000	.000	3
corrugator	s	area	.024	1.049	.028	1.017	-.043	.837	-.049	.841	-.094	.783	.000	.874	.021	.857	.142	.000	2
corrugator	s	area pond	.018	1.080	.010	1.001	-.039	.790	-.034	.827	-.057	.740	.056	.787	.109	.869	.445	.014	0
corrugator	s	me	-.108	.938	-.032	1.008	-.052	.870	.067	1.025	.362	1.351	.000	.035	.584	.001	.000	.000	3
corrugator	s	sd	-.246	.674	-.157	.850	-.072	.826	.188	1.047	.897	1.519	.000	.011	.015	.000	.000	.000	4
corrugator	e	aprox	-.166	.988	-.099	.981	.003	.957	.103	.964	.237	1.001	.000	.044	.002	.003	.000	.000	3
corrugator	e	fuzzy	-.084	.998	-.025	1.015	-.007	.962	.030	.936	-.021	.984	.015	.074	.599	.261	.123	.058	0
corrugator	e	sample	-.086	.997	-.031	1.001	-.016	.955	.032	.947	-.003	1.025	.010	.101	.660	.155	.304	.013	0
corrugator	e	shannon	-.215	.982	-.149	.972	-.001	.954	.167	.931	.443	.919	.000	.043	.000	.000	.000	.000	5
corrugator	e	spectral	-.103	1.075	-.059	1.073	.014	.953	.061	.883	.214	.826	.000	.187	.027	.147	.000	.000	3
corrugator	l	pldf	.021	1.050	.013	1.081	.008	.973	.008	1.091	-.059	.772	.178	.830	.896	.993	.067	.029	0

Table 3. Significance test of general linear models with *post hoc* test and mean (*MD*) and standard deviations (*SD*) with *z* values for pain levels: *B*, T_1, T_2, T_3, and T_4 with Bonferroni correction, maximum significant tests ≙ total.

Signal	**Group**	**Feature**	**Pain levels**														
			B	***B***	T_1	T_1	T_2	T_2	T_3	T_3	T_4	T_4	***B***, T_{1-4}	***B vs.*** T_1	T_1 ***vs.*** T_2	T_2 ***vs.*** T_3	T_3 ***vs.*** T_4	***B vs.*** T_4	**Total**
			MD	***SD***	***MD***	***SD***	***MD***	***SD***	***MD***	***SD***	***MD***	***SD***	**Wald-χ^2**	***post hoc p* level**					
corrugator	l	ldf	.124	1.046	.042	1.020	.008	.978	-.058	.927	-.324	.983	.000	.015	.312	.050	.000	.000	3
corrugator	v	var	-.203	.642	-.134	.853	-.097	.788	.125	1.004	.810	1.732	.000	.060	.304	.000	.000	.000	4
corrugator	v	std	-.264	.714	-.171	.870	-.090	.836	.200	1.026	.910	1.504	.000	.008	.021	.000	.000	.000	4
corrugator	v	range	-.275	.652	-.185	.825	-.071	.881	.209	1.023	.950	1.493	.000	.009	.001	.000	.000	.000	4
corrugator	v	intrange	-.225	.745	-.142	.877	-.095	.827	.150	.962	.771	1.597	.000	.020	.190	.000	.000	.000	4
trapezius	a	peak	-.102	.804	-.128	.806	-.044	.931	.069	1.054	.331	1.240	.000	.437	.012	.001	.000	.000	3
trapezius	a	p2p	-.101	.806	-.126	.807	-.039	.929	.067	1.053	.316	1.235	.000	.462	.010	.001	.000	.000	3
trapezius	a	rms	-.113	.849	-.107	.885	-.040	.955	.058	1.029	.301	1.195	.000	.878	.045	.004	.000	.000	3
trapezius	a	mlocmaxv	-.118	.850	-.092	.908	-.035	.960	.057	1.035	.289	1.191	.000	.446	.090	.007	.000	.000	3
trapezius	a	minlocminv	.108	.862	.094	.904	.031	.978	-.041	.999	-.285	1.210	.000	.677	.067	.034	.000	.000	3
trapezius	a	mav	-.113	.849	-.099	.904	-.035	.960	.048	1.013	.308	1.205	.000	.687	.059	.014	.000	.000	3
trapezius	a	mavfd	-.118	.846	-.107	.887	-.046	.944	.047	1.002	.324	1.215	.000	.747	.070	.006	.000	.000	3
trapezius	a	mavfdn	-.082	1.000	-.033	.990	-.037	.939	.045	.992	.183	1.020	.000	.146	.911	.015	.000	.000	3
trapezius	a	mavsd	-.119	.847	-.110	.879	-.052	.936	.046	.990	.335	1.225	.000	.781	.084	.003	.000	.000	3
trapezius	a	mavsdn	-.066	.992	-.029	.972	-.042	.935	.032	.999	.155	1.036	.000	.269	.710	.029	.000	.000	3
trapezius	f	zc	-.056	.964	-.097	.954	-.039	.953	.046	1.023	.290	1.082	.000	.225	.088	.012	.000	.000	3
trapezius	f	fmode	-.017	1.053	-.033	.962	-.043	.957	.015	.914	.115	.975	.000	.630	.765	.081	.003	.000	2
trapezius	f	bw	-.008	.978	-.037	.931	.000	.961	.025	1.081	.047	1.025	.123	.394	.268	.472	.502	.100	0
trapezius	f	cf	-.008	.978	-.037	.931	.000	.961	.025	1.081	.047	1.025	.123	.394	.268	.472	.502	.100	0
trapezius	f	fmean	-.023	.999	-.044	.978	-.053	.957	.025	.932	.130	1.007	.000	.527	.794	.019	.002	.000	2
trapezius	f	fmed	-.022	.988	-.039	.983	-.056	.955	.026	.969	.103	1.009	.000	.595	.618	.014	.021	.000	2
trapezius	s	median	-.006	.976	.013	.942	-.006	1.019	-.018	1.056	.019	1.046	.816	.570	.578	.731	.286	.462	0
trapezius	s	feqpond	-.018	.973	.005	.951	-.010	1.009	-.009	1.061	.065	1.046	.098	.497	.665	.986	.03	.015	0
trapezius	s	area	-.021	.888	.054	1.106	.030	1.087	-.035	.951	-.076	.916	.001	.027	.494	.052	.231	.106	0
trapezius	s	area pond	-.033	.791	.045	1.163	.025	1.087	-.020	.979	-.031	.9428	.069	.022	.566	.190	.739	.941	0
trapezius	s	me	-.066	.914	.017	1.063	-.010	.923	.017	1.029	.124	1.221	.000	.018	.438	.442	.003	.000	2
trapezius	s	sd	-.110	.860	-.096	.835	-.017	1.038	.061	.984	.282	1.160	.000	.692	.018	.020	.000	.000	3

trapezius	e	aprox	-.064	.986	-.037	.992	-.015	.966	.040	1.001	.195	1.034	.000	.425	.514	.108	.000	.000	3
trapezius	e	fuzzy	-.054	.986	-.004	1.025	.021	1.016	.002	.956	.054	.957	.026	.132	.453	.572	.127	.001	0
trapezius	e	sample	-.045	1.008	-.007	1.021	.022	1.019	.003	.969	.053	.954	.058	.265	.394	.568	.135	.004	0
trapezius	e	shannon	-.099	1.008	-.080	.995	-.012	.970	.071	.973	.300	.999	.000	.568	.045	.013	.000	.000	3
trapezius	e	spectral	-.013	.976	-.038	.983	-.019	.951	.005	1.039	.087	.987	.002	.462	.568	.472	.015	.003	0
trapezius	l	pldf	.035	1.175	-.046	.813	-.032	.773	.000	1.010	.024	1.121	.114	.026	.714	.371	.504	.763	0
trapezius	l	ldf	.000	.993	.032	.989	.063	.927	-.005	.974	-.060	1.025	.005	.346	.345	.042	.100	.074	0
trapezius	v	var	-.080	.831	-.092	.794	-.027	.922	.032	1.041	.212	1.254	.000	.711	.050	.081	.000	.000	3
trapezius	v	std	-.108	.868	-.113	.879	-.035	.958	.054	1.023	.298	1.190	.000	.890	.020	.009	.000	.000	3
trapezius	v	range	-.099	.819	-.131	.799	-.035	.930	.063	1.044	.315	1.233	.000	.337	.004	.003	.000	.000	3
trapezius	v	intrange	-.096	.875	-.091	.899	-.018	.991	.031	1.005	.282	1.195	.000	.878	.032	.152	.000	.000	3
SCL	a	peak	-.168	.995	-.124	1.012	-.099	.967	-.012	.973	.241	1.054	.000	.194	.465	.01	.000	.000	3
SCL	a	p2p	-.171	.869	-.172	.904	-.123	.962	.019	1.040	.504	1.326	.000	.973	.167	.000	.000	.000	3
SCL	a	rms	-.122	1.001	-.105	1.011	-.093	.985	-.024	.992	.070	.985	.000	.616	.709	.044	.005	.000	2
SCL	a	mlocmaxv	-.132	.982	-.099	1.064	-.104	.987	-.001	.963	.064	.983	.000	.331	.872	.002	.056	.000	2
SCL	a	minlocminv	-.134	.972	-.089	1.061	-.088	.978	-.028	.987	.022	.976	.000	.185	.962	.077	.146	.000	2
SCL	a	mav	-.122	1.001	-.105	1.012	-.093	.985	-.024	.993	.067	.983	.000	.629	.710	.042	.007	.000	2
SCL	a	mavfd	-.138	.955	-.106	1.005	-.105	.947	-.001	1.051	.271	1.105	.000	.346	.996	.003	.000	.000	3
SCL	a	mavfdn	.190	1.044	.195	1.172	.115	1.024	-.028	.969	-.226	.912	.000	.901	.022	.000	.000	.000	4
SCL	a	mavsd	-.158	.945	-.086	1.023	-.099	.961	-.015	1.032	.167	1.055	.000	.035	.704	.014	.000	.000	3
SCL	a	mavsdn	.172	1.047	.209	1.180	.117	1.023	-.025	.963	-.253	.907	.000	.289	.008	.000	.000	.000	4
SCL	f	zc	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
SCL	f	fmode	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
SCL	f	bw	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
SCL	f	cf	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
SCL	f	fmean	.900	.397	.923	.620	.897	.367	.895	.361	.905	.378	.317	.115	.08	.865	.494	.735	0
SCL	f	fmed	-.084	.940	-.043	.945	-.072	.813	.004	.986	.294	1.436	.000	.250	.424	.034	.000	.000	2
SCL	s	median	.010	.975	.043	1.006	.008	1.001	.006	.999	-.038	1.031	.219	.329	.306	.95	.196	.160	0
SCL	s	feqpond	.012	.979	-.001	1.106	.016	.965	-.003	.949	-.019	1.001	.861	.715	.615	.578	.636	.372	0
SCL	s	area	-.034	1.113	-.076	.958	-.040	.860	-.013	.855	.287	1.141	.000	.215	.291	.417	.000	.000	3
SCL	s	area pond	-.044	1.134	-.064	.803	-.054	.875	-.027	.877	.312	1.161	.000	.565	.771	.442	.000	.000	3
SCL	s	me	-.175	.843	-.128	.943	-.114	.965	.012	1.023	.432	1.315	.000	.174	.702	.000	.000	.000	4
SCL	s	sd	-.202	.898	-.108	.899	-.056	.932	.133	1.027	.615	1.334	.000	.007	.14	.000	.000	.000	4

Table 3. Significance test of general linear models with *post hoc* test and mean (*MD*) and standard deviations (*SD*) with *z* values for pain levels: *B*, T_1, T_2, T_3, and T_4 with Bonferroni correction, maximum significant tests ≙ total.

Signal	**Group**	**Feature**	**Pain levels**														
			B	***B***	T_1	T_1	T_2	T_2	T_3	T_3	T_4	T_4	***B***, T_{1-4}	***B vs.*** T_1	T_1 ***vs.*** T_2	T_2 ***vs.*** T_3	T_3 ***vs.*** T_4	***B vs.*** T_4	**Total**
			MD	***SD***	***MD***	***SD***	***MD***	***SD***	***MD***	***SD***	***MD***	***SD***	**Wald-χ^2**	***post hoc p* level**					
SCL	e	aprox	.169	1.058	.183	1.134	.108	1.010	-.061	1.090	-.209	.856	.000	.691	.037	.000	.000	.000	4
SCL	e	fuzzy	.220	1.066	.194	1.124	.117	1.027	-.042	.992	-.227	.926	.000	.455	.028	.000	.000	.000	4
SCL	e	sample	.196	1.073	.188	1.099	.124	1.058	-.059	1.024	-.221	.897	.000	.808	.069	.000	.000	.000	4
SCL	e	shannon	-.177	.972	-.207	1.025	-.178	.968	-.083	.955	.182	.994	.000	.373	.39	.005	.000	.000	3
SCL	e	spectral	-.067	.953	-.039	.929	-.070	.822	.007	1.058	.288	1.437	.000	.432	.393	.034	.000	.000	3
SCL	l	pldf	-.103	.831	-.109	.771	-.089	.860	-.011	1.036	.317	1.407	.000	.864	.579	.028	.000	.000	3
SCL	l	ldf	-.031	.997	-.018	1.071	.031	.871	.021	1.017	.021	1.114	.335	.705	.161	.788	.987	.135	0
SCL	v	var	-.110	.810	-.061	.985	-.058	.973	.009	1.151	.358	1.398	.000	.183	.948	.069	.000	.000	3
SCL	v	std	-.166	.846	-.130	.944	-.104	.966	.001	1.096	.418	1.306	.000	.307	.461	.003	.000	.000	3
SCL	v	range	-.179	.844	-.160	.933	-.114	.947	.018	1.057	.479	1.304	.000	.580	.188	.000	.000	.000	4
SCL	v	intrange	-.147	.842	-.104	.947	-.089	1.004	-.008	1.125	.295	1.278	.000	.228	.675	.024	.000	.000	3
ECG	v: HRV	meanRR	.154	1.028	.129	.905	.024	1.000	-.084	.982	-.308	.979	.000	.447	.002	.001	.000	.000	3
ECG	v: HRV	rmssd	.163	1.017	.121	.906	.021	.998	-.092	.983	-.305	.984	.000	.203	.003	.001	.000	.000	3
ECG	v: HRV	slopeRR	.058	.955	.135	.983	-.002	.948	-.117	.996	-.338	1.080	.000	.023	.000	.000	.000	.000	5

SCL, ≙ skin conductance level; ECG, ≙ electrocardiogram; HRV, ≙ heart rate variability; a, ≙ amplitude; f, ≙ frequency; s, ≙ stationarity; e, ≙ entropy; l, ≙ linearity; v, ≙ variance.