Twofold-Multimodal Pain Recognition with the X-ITE Pain Database

Philipp Werner, Ayoub Al-Hamadi
Neuro-Information Technology group
Otto von Guericke University
Magdeburg, Germany
{Philipp.Werner, Ayoub.Al-hamadi}@ovgu.de

Sascha Gruss, Steffen Walter
Medical Psychology group
University Clinic Ulm
Ulm, Germany
{Sascha.Gruss, Steffen.Walter}@uni-ulm.de

***Abstract*—Automatic pain recognition has great potential to improve pain management. In this work, we investigate multimodality in pain recognition in two regards. First, we compare and combine multiple sensor modalities, which capture both behavioral and physiological pain responses. Second, we compare and distinguish the heat and the electrical pain stimulus modalities in both phasic (short) and tonic (long) variants. Experiments show that (1) pain intensity can be recognized automatically in all stimulus variants, (2) that pain of different qualities (heat and electical stimuli) can be distinguished, (3) that electrodermal activity (EDA) is the best performing single modality, and (4) that fusion with modalities can improve results further.**

***Index Terms*—pain, assessment, recognition, intensity, multimodal, modality, fusion, heat stimuli, electrical stimuli**

I. Introduction

Automatic pain recognition may complement current methods of clinical pain assessment one day, especially for patients who cannot utter on their pain experience, such as infants, adults with cognitive impairment, or unconscious persons [1], [2]. Currently, the pain of those patients is assessed by humans by observing behavior and physiological signals. Automatic systems are promising, because they facilitate continuous monitoring of pain, which is not possible with human observers. Further, they may be more objective than humans, who are influenced by personal factors, such as the relationship to the sufferer [3] and the patient's attractiveness [4].

A. Related Work

Most works on automatic pain recognition focus on facial expression and use the UNBC-McMaster Shoulder Pain Database [5]–[7]. However, many recent results showed that other sensor modalities are very promising as well and that fusion of multiple modalities can improve recognition performance and flexibility (e.g. by relying on other modalities if the face is occluded). Tsai et al. [8] demonstrated that pain intensity can be assessed from audio recorded during triage interviews in an emergency department. In another work, they fused audio and facial expression, outperforming each single modality [9]. Head pose is another behavioral signal that has been shown to be useful for pain recognition [10], [11]. Among the physiological signals, electrodermal activity (EDA), surface electromyography (EMG), and electrocardiogram (ECG) are the most widely used modalities. Werner et al. [12] and Kächele et al. [13]–[15] conducted experiments with these physiological signals, head pose, and facial expression showing some strengths and weaknesses of the single modalities and that multimodal fusion can improve recognition in many cases. Thiam et al. [16] additionally included audio and respiration signals in their work. Aung et al. [17] analyzed facial expression (with a camera) and (separately) body movement (with EMG and mocap) for recognizing pain behaviors of chronic low back pain patients. The body movement based recognition was developed and validated further by Olugbade et al. [18]. A clinical study with infants was conducted by Zamzmi et al. [19], in which pain was recognized using facial expression, body movement, audio, physiological signals, and their fusion.

For comparing and advancing the recognition methods, benchmark datasets are very beneficial. Table I summarizes pain recognition databases that are publicly available now or announced to be published soon. In this work we use the X-ITE Pain Database [25]. Of the databases known to the authors, X-ITE comprises the most sensor modalities (also see Sec. II-E). The SenseEmotion Database [23] comes with a similar set of modalities, but in contrast to X-ITE it does *not* include facial EMG, body movement video, and thermal video. In terms of pain stimulus modalities, X-ITE is the first database that includes stimuli of four different types: phasic (short duration) heat, tonic (long duration) heat, phasic electrical, and tonic electrical stimuli. The other databases only include one type of stimulus (BioVid: phasic heat, BP4D+: tonic cold pressor test, EmoPain: physical exercises in a therapy scenario, SenseEmotion: phasic heat, MIntPAIN: phasic electical). X-ITE also surpasses the other databases regarding the number of samples and subjects.

Next to intensity of pain, which is widely covered in literature on automatic pain recognition, pain quality is another relevant characteristic that is commonly assessed in clinical practice [26]. The pain quality refers to the description or type of pain, e.g. whether it is sharp, dull, crushing, burning, tearing, etc. and whether it is intermittent, constant, or throbbing. It has been shown that the reported pain quality differs between

This work was funded by the German Research Foundation (DFG, no. AL 638/3-2) and the Federal Ministry of Education and Research (BMBF, no. 03ZZ0470). The sole responsibility for the content lies with the authors.

TABLE I
PAIN DATABASES

	Sensor Modalities						Stimulus Modalities / Types					Size		
Database	Video	Audio	EMG	ECG	EDA	Other	Thermal	Electrical	Movement	Phasic	Tonic	Subjects	Painful Stimuli	Medical Condition
UNBC-McMaster [5]	✓								✓			25	200	shoulder pain
BioVid [20], [21]	✓		✓	✓	✓		✓			✓		ca. 90	14k	healthy
BP4D+ [22]	✓			✓	✓	✓	✓				✓	140	140	healthy
EmoPain [17]	✓	✓	✓			✓			✓			50	?	chronic low back pain
SenseEmotion [23]	✓	✓	✓	✓	✓	✓	✓			✓		45	8k	healthy
MIntPAIN [24]	✓					✓		✓		✓		20	2k	healthy
X-ITE [25]	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	134	25k	healthy

experimental pain models [27] that are widely used in research. For instance, phasic electric pain stimulation was described as pricking, flickering, sharp, and pinching; longer lasting (tonic) ischemia pain as dull, pinching, hot, annoying, spreading, and tight [28]. To the authors' knowledge, distinguishing pain of different qualities has not been addressed in automatic pain recognition yet.

B. Contribution

This work features several novelties: (1) It is the first work reporting results on the new X-ITE Pain Database (which is described in Sec. II). (2) To our best knowledge, it is the first work addressing multimodality in terms of sensors and pain stimulation at the same time. (3) It compares recognition of four types of stimuli that differ in duration (phasic/tonic) and the used experimental pain model (heat/electrical). (4) We report on experiments on using facial video (facial expression and head pose), audio, three EMGs (including facial EMG), EDA, and ECG individually and combining the modalities. (5) Results show that it is possible to distinguish two types of stimuli that are associated with a different quality of pain (heat/electrical). In Sec. III we describe the used recognition method. Sec. IV reports on our experiments and Sec. V discusses the results and draws conclusions.

II. X-ITE PAIN DATABASE

In this section we give an overview of the new multimodal Experimentally Induced Thermal and Electrical (X-ITE) Pain Database. For more details, refer to Gruss et al. [25].

A. Participants

A total of 134 healthy adults (67 men and 67 women) aged between 18 and 50 years participated in the experiment. The average age of all subjects was 31.4 (SD = 9.7), of all men = 33.4 (SD = 9.3), and of all women = 32.9 (SD = 10.2) years. None of them suffered from chronic pain, depression, or had a history of psychiatric disorders; none had neurological conditions, headache syndrome, or cardiovascular disease; none had taken pain medication or used painkillers directly before the experiment.

B. Pain Stimulation

Heat pain was stimulated at the participants forearm using the Medoc PATHWAY Model ATS thermal stimulator. Electrical pain was stimulated with electrodes attached to the index and middle finger using the electrical stimulator Digitimer DS7A. Both, heat pain (H) and electrical pain (E) were applied in two variants: phasic (short) stimuli of 5 seconds duration (P) and tonic (long) stimuli of 60 seconds duration (T). Each of the four stimulus types (HP, HT, EP, ET) was stimulated in three intensities.

C. Stimulus Calibration Phase

In order to handle differences in pain sensitivity, each participant underwent a stimulus calibration procedure prior to the main experiment. The calibration comprised four parts, one for each stimulus type. In each part, the stimulus intensity was gradually increased while asking the participant to report the felt pain intensity to determine the person-specific pain threshold and tolerance. Afterwards, the six heat stimulation temperatures (three intensities for phasic and three for tonic stimulation) and six electrical stimulation currents (three phasic and three tonic) were calculated based on the pain thresholds and tolerances.

D. Main Stimulation Phase

After the calibration procedure, the participant laid down on an examination couch and underwent the main stimulation phase, which took about 90 minutes. The phasic stimuli of each modality (heat and electrical pain) and intensity were repeated 30 times in randomized order with pauses of 8-12 seconds. The 1-minute tonic stimuli were only applied once per intensity, i.e. there were six tonic stimuli per participant, each followed by a pause of five minutes. They were applied in three phases: The highest intensity tonic heat and electrical pain stimuli were applied at the end of the experiment. The other two phases, which each contained one tonic heat and one tonic electrical stimulus of lower intensity, were randomly started during the phasic stimulation period.

E. Data Recording

Synchronously to the applied pain stimulation, several sensors were used to collect multimodal pain response data:

- RGB video of the face (frontal and side view) for analyzing facial expression and head pose,
- audio signal for analyzing para-linguistic responses,
- electrocardiogram (ECG) for analyzing heart rate and its variability,

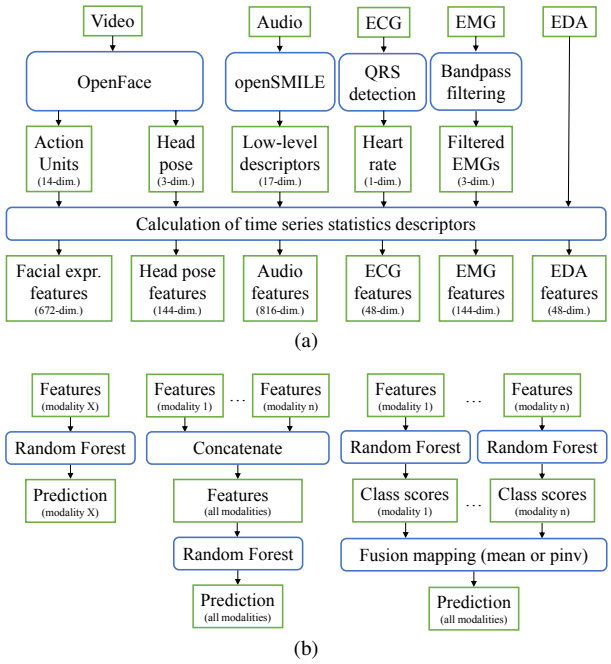

Fig. 1. Recognition method comprising (a) feature extraction and (b) classification of single modality (left), multimodal feature fusion (middle), and multimodal decision fusion (right).

- surface electromyography (EMG) to measure activity of three muscles, the trapezius (neck/shoulder), corrugator supercilii (close to eyebrows), and zygomaticus major (at the cheeks),
- electrodermal activity (EDA) to measure sweating,
- video of the body to analyze body movements, and
- thermal video of the face.

III. Recognition Approach

This section describes the recognition approach used for the experiments. It was adopted from prior work, because this work focuses on providing and analyzing baseline results for the new X-ITE Pain Database rather than proposing a new recognition approach. Fig. 1 gives an overview of the used method. We processed facial RGB video, audio, ECG, EMG, and EDA data as detailed in the subsections. This yielded one or multiple time series per modality, which were condensed in time-series statistics descriptors. The descriptors were used as features for recognizing the pain stimulus intensity and stimulus modality using Random Forest classification. We applied the classification for each sensor modality individually and also fused all modalities using three approaches: feature fusion, decision fusion with mean-score mapping, and decision fusion with pseudoinverse mapping.

A. Processing of Sensor Signals

Video: The frontal facial RGB video was processed with OpenFace [29] to extract facial expression and head pose information. As frame-level expression features we used the 14 FACS Action Unit intensities provided by OpenFace without post-processing, i.e. the outputs were *not* clipped to the AU intensity range of 0 to 5. The head pose was described using the provided yaw, pitch, and roll head orientation angles.

Audio: The audio signal was processed with openSMILE [30] using frames of 25 ms length extracted in steps of 10 ms. For each frame we extracted a 17-dimensional low-level descriptor comprising the logarithmic signal energy, the voicing probability, the pitch (F_0), the F_0 envelope, and 13 Mel Frequency Cepstral Coefficients (MFCCs). The low-level descriptor time series were smoothed with a moving average filter over three frames.

ECG: We applied the QRS-detection algorithm by Hamilton and Tompkins [31] in order to find the R-peaks in the ECG signal. Afterwards, the heart rate was calculated from the R-to-R intervals. Finally, we interpolated the heart rate signal linearly to match the sampling of the EMG and EDA (500 Hz).

EMG: The three EMG channels were preprocessed with a zero-phase 3rd-order Butterworth band-pass filter with cut-off frequencies of 20 and 250 Hz. Further, they were downsampled from 1,000 Hz to 500 Hz to speed-up processing.

EDA: The 1-dimensional EDA time series was downsampled to 500 Hz as the other biopotentials, but was not processed any further.

B. Time Series Statistics Descriptor

The time series of the above-mentioned sensor modalities were segmented based on the applied stimulation, i.e. the samples were temporally aligned with the pain stimulation. For the phasic stimuli, we extracted time windows with a duration of 6 s, each starting with the stimulus. As phasic baseline samples we selected 6 s time windows following phasic heat stimuli of lowest intensity. For the tonic stimuli, time windows of 60 s were extracted, temporally matching the applied pain stimulus. The respective baseline samples of the same length were cut from the pause following the tonic heat stimuli of lowest intensity.

We calculated a feature vector for each sample and sensor modality as proposed by Werner et al. [11] in the context of facial activity. Each time series was summarized by several statistics of the time series itself and of its first and second derivative, including mean, maximum, range, time of maximum, and others, yielding a 48-dimensional descriptor per time series. The smoothing proposed by Werner et al. (1 Hz first-order Butterworth low-pass filtering) was only applied for the video time series, because further noise reduction was not necessary for the other modalities. We applied a person-specific standardization of the features [11] in order to focus this work on the within-subject response variation rather than the differences between subjects.

C. Classification and Fusion

The samples were classified using Random Forests (RF) [32] with 100 trees and a maximum depth of 10 nodes. To compare the usefulness of the sensor modalities alone, we first trained and tested RFs using the features of each modality individually. Second, we concatenated the feature vectors of all

TABLE II
CROSS-VALIDATION ACCURACY OF **PHASIC STIMULI 2-CLASS TASKS** (COLUMNS) AND MODALITIES / FUSION APPROACHES (ROWS).

	No Pain vs H. Pain			No Pain vs E. Pain			H. Pain vs H. Pain			E. Pain vs E. Pain			H. Pain vs E. Pain			Mean	
	B/H1	B/H2	B/H3	B/E1	B/E2	B/E3	H1/H2	H2/H3	H1/H3	E1/E2	E2/E3	E1/E3	H1/E1	H2/E2	H3/E3	modality	category
Facial expression	52.2	58.2	75.2	52.8	59.1	75.9	55.0	68.7	73.9	55.2	71.2	75.4	52.1	59.3	73.4	63.9	
Head pose	51.2	54.9	70.7	51.8	56.6	70.0	53.9	67.1	70.9	54.5	67.4	71.0	51.4	55.5	67.4	61.0	61.4
Audio	50.5	53.5	67.8	50.2	54.5	69.8	54.0	65.1	68.1	53.3	67.8	70.3	51.0	52.4	62.2	59.4	
ECG	50.4	53.9	65.6	52.2	60.4	72.9	52.5	62.9	65.0	58.9	66.4	73.5	51.4	60.6	68.6	61.0	
EMG	51.4	59.0	78.0	64.3	74.9	87.8	57.0	71.1	77.3	63.3	75.1	82.2	63.4	74.6	82.7	70.8	67.9
EDA	59.3	64.1	79.1	66.6	79.9	91.1	56.8	68.6	72.9	64.8	75.0	83.6	61.5	71.7	81.6	71.8	
Feature fusion	58.4	63.8	82.2	68.0	81.0	92.4	57.1	72.1	78.6	66.4	79.0	86.9	63.3	75.5	85.4	74.0	
Decision fusion (mean)	58.6	63.4	79.7	68.6	78.7	88.2	58.3	71.8	77.9	66.2	76.9	84.6	63.1	77.5	86.3	73.3	74.8
Decision fusion (pinv)	55.7	65.9	83.3	72.3	83.7	94.3	60.3	72.7	80.4	70.7	82.2	90.3	71.7	81.1	90.5	77.0	
Mean	54.2	59.6	75.8	60.8	69.9	82.5	56.1	68.9	73.9	61.5	73.4	79.8	58.8	67.6	77.6	68.0	
Mean (category)		63.2			71.0			66.3			71.6			68.0			

B: Baseline (no pain) H: Heat (pain) E: Electrical (pain) Hx: Heat pain of intensity x (1 = lowest, 3 = highest) Ex: Electrical pain of ...

TABLE III
CROSS-VALIDATION ACCURACY OF **TONIC STIMULI 2-CLASS TASKS** (COLUMNS) AND MODALITIES / FUSION APPROACHES (ROWS).

	No Pain vs H. Pain			No Pain vs E. Pain			H. Pain vs H. Pain			E. Pain vs E. Pain			H. Pain vs E. Pain			Mean	
	B/H1	B/H2	B/H3	B/E1	B/E2	B/E3	H1/H2	H2/H3	H1/H3	E1/E2	E2/E3	E1/E3	H1/E1	H2/E2	H3/E3	modality	category
Facial expression	59.6	58.0	69.2	59.8	62.4	71.3	58.1	66.5	73.4	55.9	69.0	70.7	52.6	58.1	65.6	63.3	
Head pose	58.8	60.0	68.4	65.0	60.0	65.6	61.8	65.3	73.8	50.6	63.3	69.5	55.9	52.8	59.2	62.0	62.0
Audio	58.8	55.9	68.0	56.9	53.9	59.9	60.6	64.1	70.2	57.9	62.9	68.7	58.3	51.6	62.0	60.6	
ECG	53.5	59.6	72.1	61.0	56.3	75.3	63.0	67.3	76.2	56.7	73.4	69.9	56.7	65.9	63.2	64.7	
EMG	58.8	67.3	78.5	67.1	73.5	87.0	61.0	70.2	78.6	65.2	72.2	78.7	70.0	69.9	76.4	71.6	70.7
EDA	55.5	67.8	86.6	73.2	79.6	87.0	68.7	79.4	86.3	61.5	77.0	84.3	72.9	79.7	79.6	75.9	
Feature fusion	64.5	71.4	84.6	69.5	81.6	89.5	65.9	74.2	82.7	59.5	74.6	82.7	68.0	69.9	78.0	74.4	
Decision fusion (mean)	66.5	69.8	84.2	74.8	82.0	88.3	69.9	75.0	86.3	64.8	78.6	83.9	73.3	72.8	81.2	76.8	72.7
Decision fusion (pinv)	56.3	66.1	69.2	72.0	72.2	68.8	63.8	60.5	71.0	56.3	71.0	67.9	63.2	71.1	72.0	66.8	
Mean	59.1	64.0	75.7	66.6	69.1	77.0	63.6	69.2	77.6	58.7	71.3	75.1	63.4	65.8	70.8	68.5	
Mean (category)		66.3			70.9			70.1			68.4			66.7			

B: Baseline (no pain) H: Heat (pain) E: Electrical (pain) Hx: Heat pain of intensity x (1 = lowest, 3 = highest) Ex: Electrical pain of ...

modalities and trained and tested RFs on the resulting 1,872-dimensional feature space. This approach is called *Feature Fusion*. Third, we applied two types of *Decision Fusion*, in which an individual RF is trained for each modality. In this approach, each RF not only yields the class that the majority of its trees predicted, but it provides a score for each possible class (the proportion of trees that predicted this class). There are many ways how to aggregate these classifier scores into a final decision. Here we used (1) a fixed mapping approach calculating the *mean* of all RF scores per class and selecting the class with the highest score, and (2) a trained mapping approach that learns class-score weights for each modality by calculating the pseudoinverse (pinv) matrix [33].

IV. EXPERIMENTS

Dataset: In this work we experimented with stimulus-aligned samples, which had been cut out from the continuous recording of the main stimulation phase of the X-ITE study (about 90 minutes per participant). See Sec. III-B for details about the time windows. The phasic (short) and tonic (long) stimuli were evaluated separately, because the number of repetitions is very different (one for tonic vs 30 for phasic). Combining both in one recognition framework is possible but challenging. Thus, we leave this for future work. Due to technical problems, some sensor modalities and stimuli are not available. We used the intersection set of the samples, i.e. we only included samples for which there was no failure for any of the used sensors (frontal RGB camera, audio, ECG, EMG, EDA). This way we were able to use data of 127 subjects. The tonic dataset contained 865 samples in total; the phasic dataset contained 26,308 samples. The classes were approximately balanced in both the tonic and the phasic dataset.

Validation: We applied leave-one-subject out cross validation to estimate the performance of various models on unseen subjects. The performance is reported with the accuracy measure[1] (in percent), which is intuitive and well-suited for the balanced datasets we consider here. In the following we report experimental results of several classification tasks: pairwise binary classification of phasic and tonic pain, 4-class intensity estimation of phasic and tonic pain, and the full 7-class pain recognition task. We considered the single sensor modalities and their fusion (three approaches), heat and electrical pain (which are two different stimulus modalities associated with

[1]The accuracy is defined as the quotient of the number of all correctly classified samples and the total number of tested samples.