ISSN 1870-4069

Physiological and Behavioral Dataset Captured in a Learning Environment

Yesenia Nohemí González-Meneses, Blanca Estela Pedroza-Méndez, José Federico Ramírez-Cruz, José Crispín Hernández-Hernández, Edmundo Bonilla-Huerta

Tecnológico Nacional de México,
Tlaxcala,
Mexico

{yesenia.gm, blanca.pm, federico.rc, crispin.hh, edmundo.bh} @apizaco.tecnm.mx

Abstract. In this article, we describe the experiment for the capture of physiological and behavioral signals of college students while doing a learning activity. This is a fundamental task for the recognition of emotions in specific contexts; it is part of the data processing stage in the methodology of the research project in which it is intended to automatically recognize emotions in educational environments using machine-learning algorithms and acquisition technologies of physiological and behavioral signals to identify the connections between emotions and learning. Data capture is guided by a formal procedure designed for a specific educational context. In this, the characteristics of the learning environment (of the physical place and of the software), of the participants and of the capture devices used were considered. 82 college students in engineering careers participated in the data capture. The learning activity was through interaction with a basic algebra MOOC with which they learned basic concepts by watching some videos containing the theory and practical exercises, in the end, they answered an evaluation of the subjects studied. We use 2 physiological signal capture devices and 2 video capture devices.

Keyword: Physiological and behavioral signals dataset, emotion recognition, learning-centered emotions, educational innovation, learning environments.

1 Introduction

In the learning process of a student, emotions have a determining influence on the acquisition of knowledge. In people, emotions awaken curiosity, interest, and, therefore, the focus of attention.

The learning that is associated with emotions generated in everyday life or in a classroom - whether positive (such as joy or interest) or negative (such as fear or sadness) - will remain in our memory. The identification of student emotions during learning processes helps to select teaching strategies more suited to the student's model.

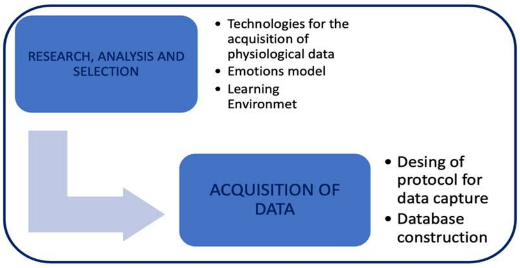

Fig. 1. Methodological stages.

This model is usually contained in educational software such as intelligent tutorial Systems (ITS). These types of tutorials consider students' emotions, their ways of learning, their intelligence, cognitive abilities, and other information that form the student's model [1].

The process of emotional recognition is less invasive when done automatically using devices that give us information about the emotional state of the student. The devices using for this work are thermal camera, video camera, heart rate sensor and Kinect 360 for Windows.

The problems of automatic recognition of emotions have been actively investigated in recent years. Despite the amount of research and attention, a clear solution which could be utilized by most people is still in the distant future. Several drawbacks have hindered the construction of an appropriate solution from a computational point of view. A big challenge to the effective performance of software and hardware that recognize emotions in real contexts is the difficulty in generating databases with spontaneous emotions.

Other aspects to consider are the areas of application, which have a significant effect on the accuracy of emotion recognition, and the degree of intrusion of the tools used [2]. The main objective in this work is to provide an adequate dataset for after processing them to build a database for the study of learning-centered emotions.

This is a fundamental task for the recognition of emotions in specific contexts. To justify the creation of the dataset, we did a review of state of the art on automatic recognition of emotions in learning environments and on databases for the identification of emotions. Considering the revision of the state of the art we identify the problems in the creation of a robust dataset for the automatic recognition of emotions.

Later, we set out the stages of procedure design for the capture of physiological and behavioral signals of students while they did a learning activity, and we present statistics of the dataset created.

After the analysis of recent related works, we can identify that in learning environments, there is a lack of adequate methodologies to recognize the emotional state of students during learning processes through interactions with computers. These methodologies preferably would need to integrate diversified characteristics obtained even from the fusion of data coming from the use of different technologies that acquire physiological and behavioral signals.

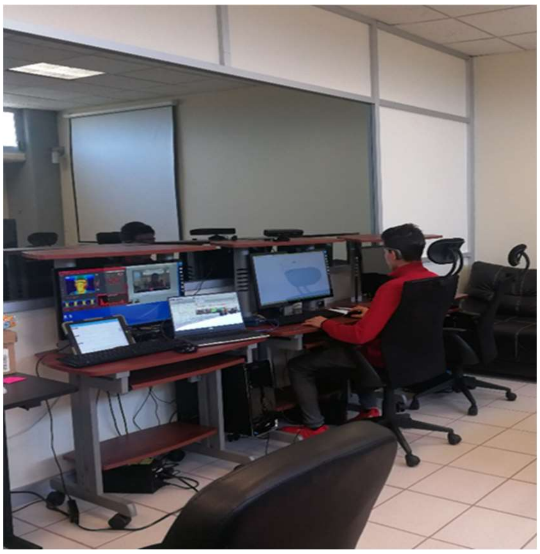

Fig. 2 User experience laboratory of the UPPue.

They should be based on a model of emotions that is realistic and allows becoming closer, overall, to the cognitive learning process that contributes to the analysis of the emotion-learning relation (i.e., the connections between emotions and learning).

We also identify that the main factor that affects the performance of emotion-recognizers in real contexts is the difficulty in generating databases of spontaneous emotions.

Generally, works are done with acted databases, which provide portraits of emotions to represent prototypical and intense emotions that facilitate the search of correlations and the subsequent automatic classifications.

The problems that are usually faced in the capture of physiological and behavioral signals data in real contexts are: Use of invasive technologies that cause discomfort in the capture of spontaneous expressions, synchronization in the different capture devices, the environmental conditions where the capture takes place, the correct sensing of the devices, the previous emotional state of the participants, the availability of the participants, the physical appearance of the face (amount of hair or artifacts on the face), the necessary capture time.

Due to the above mentioned, the formal design of a procedure for the capture of data (either physiological, behavioral or both) is necessary. The procedure will guide the execution of a controlled experiment with the objective of avoiding errors in the acquisition of data, describing the technologies to be used, the capture process, the environmental conditions, and the preparation of the participants.

To address the lack of methodologies owed to the weaknesses of existing data bases, we made a priority the creation of a more robust dataset whose construction is guided by the steps of a formal procedure for data capture (either physiological, behavioral, or both). This priority is important because data are the basis for the implementation of

Table 1. Sample characteristics statistics.

Total Students	Average Age (years)	Men	Women	Engineering	Master	PhD	High School
82	21	50	32	69	8	4	1

a model of emotion identification with an acceptable level of accuracy that is more in line with reality in an educational context.

2 State of the Art

To address the state of the art, we first reviewed the research on automatic recognition of learning-centered emotions. After this analysis, we identified the lack of databases with emotional expressions captured in real learning environments as the main problem.

From this, we made an analysis of the databases constructed in various research works and captured both in generic environments and in specific educational contexts. In this way we identify that the creation of a physiological and behavioral database as an initial stage in the recognition of learning-centered emotions and the problems that must be addressed in the construction of the dataset is necessarily.

A database of emotional expressions is a collection of images or video clips, speech, and physiological signals related to a wide range of emotions. Its content corresponds to emotional expressions related to the context in which they were captured and then tagged; this is essential for the training, testing, and validation of algorithms to develop expression-recognition systems.

The labeling of emotions can be done on a discrete or continuous scale. Many of the databases are usually based on the theory of emotions [3] that assumes that there are six basic emotions and an approximate variety of 22 secondary emotions on a discrete scale. However, in some databases, emotions are labeled on the continuous arousal-valence scale [4]. And other databases include the AUs (Action Units) based on the FACS (Facial Action Coding System) [5].

The databases of expressions of emotions are mostly formed only by facial expressions and these are classified into posed and spontaneous. In posed expression databases, the participants are asked to display different emotional expressions, while in spontaneous expression database, the expressions are natural. Spontaneous expressions differ from posed ones remarkably in terms of intensity, configuration, and duration. In most cases, the posed expressions are exaggerated, while the spontaneous ones are subtle and differ in appearance.

Apart from this, synthesis of some AUs are barely achievable without undergoing the associated emotional state so it is not possible to capture physiological data since these cannot be controlled by people and therefore do not correspond to the emotion that is acted out.

For our research we first made a compilation of databases of facial expressions (approximately 18 databases of works from 2010 until June 2021), although our interest lies in hybrid databases that contain facial images and physiological data, so far, we have only found one database of this type publicly available. The DEAP database [6]

Table 2. Data obtained from each of the experiments.

# Experiment	**1**	**Pulse**	**YES, range: 513-553, Mean: 546**
Date	10/10/2018	% Pulse Data	83%
Name	Student´s Name	Total Captures	21672
Sex	M	Sample Range	9.505263158
Age	20	Captures Per Minute	571
Engineering	Computation Science	Files CSV/JPG	2074 / 0
Set Time (min)	38	Time AVI (min)	03:48
Set Time (seg)	2280	Video (min)	39.00
Start Time	08:40	Kinect (in)	39
Ending Time (real)	09:19	Observations	none
Real Time	39		

Table 3. Statistics of raw data files.

Pulse Sensor "txt"	**Thermal Camera "csv/jpg/avi"**	**Webcam "wmv"**	**Kinect "xed"**	**Files Per Experiment**	**Total Files (82 Experiments)**
1	4801	1	1	4804	393,928

contains physiological recordings (of EEG) and facial video of an experiment where 32 volunteers watched a subset of 40 music videos.

In this experiment participants are also asked to rate each video according to the emotion it caused in them. These databases correspond to facial expressions of basic emotions, which have been captured while people perform different activities.

Most databases are comprised of viewers watching media content (i.e. ads, movie trailers, television shows, animated gifs and online viral campaigns) and of posed facial expressions like the database DEAP [6], CK++ [7] and Affective (the most robust database for the amount for data stored) [8].

Subsequently, we also compiled databases of facial expressions that correspond to learning-centered emotions (approximately 13 databases) like the database of [4, 9] that have been captured while students perform some learning activity in a wild setting with spontaneous facial expressions, so it is possible to capture physiological data to complement them.

These criteria are what we consider for the selection of the databases analyzed. These correspond to learning-centered emotions that have been captured while students perform some learning activity in a wild setting.

Some of them contain physiological data but unfortunately, they do not make them available to the public or provide information about the characteristics of these data, let alone details of the processing they do with them. Of these, four databases include physiological data. The most complete is the one mentioned in [10] in which 67 students participate and use three physiological data sensors and a video camera.

In general, they have been created with data from very few participants. From this analysis, we have identified as an area of opportunity the creation of a physiological and behavioral database supported by a formal procedure for data collection. The definition of the procedure allows us to run a controlled experiment in a natural environment and specific that can be replicated the number of times necessary to create a robust dataset.

3 Methodology

In this section, we describe the methodological stages, which can be seen in Figure 1. First, we will describe the activities of research, analysis, and selection of critical elements for the creation of a spontaneous dataset captured in real educational environments through the execution of a controlled experiment.

This description includes the physiological and behavioral signals to be captured, the data-acquisition technologies used to capture the signals, the emotions model on which the identification of emotions is based, and the activity in which the students participate while their data is captured.

The next stage includes the creation of the dataset. We will create a dataset of student features obtained in real time while doing a learning activity. Such a database will be formed from two physiological characteristics (the temperature of the face and the heart rate) and two behavioral characteristics (the facial image and the movements of the head and of the superior limbs).

3.1 Investigation, Analysis and Selection

After the corresponding investigation and analysis, the first three elements of this stage have been selected. The physiological and behavioral data acquisition technologies that were chosen are the following:

1. Logitech Webcam®: to capture continuous video during the 35 minutes that the experiment lasts in "MP4" format (1920X1080, 28 f / s).

2. ICI 9320P thermal camera: to capture arrays of face temperatures, in "CSV" files (240X320), generating 1 file per second. An "AVI" video (600 images per minute) of approximately 3.5 minutes is also generated.

3. Cardiac pulse sensor implemented with Arduino: a "TXT" file of the sensor signal power readings is generated with a range of approximately 9.5 captures per second (10Hz sampling rate).

4. Kinect for Windows®: continuous video is recorded during the minutes that the experiment lasts in an "XED" file.

For the emotion model, for now both the discrete and the continuous model will be considered. The types of learning-centered emotions that are tried to recognize are: interest (pleasure), boredom, confusion, and frustration.

The educational environment with which students will interact for the teaching-learning process will be a MOOC of basic Coursera algebra working only with the first subject (approximate duration of 36 min.). The young people of the sample will be college students of careers in the engineering area, whose ages range between 18 and 24 years.

Table 4. Number of experiments with complete files per device used.

Thermal Camera	Pulse Sensor	Video	Kinect	Complete Experiments	Thermal Camera, Video and Pulse
64	48	79	73	44	34

3.2 Procedure for the Capture of Physiological and Behavioral Data

The procedure guides the execution of a controlled experiment with the objective of avoiding errors in the acquisition of data, describing the technologies to be used, the capture process, the environmental conditions, and the preparation of the participants.

In the first part of the procedure, the stages of research, analysis, and selection of key elements for the creation of a spontaneous dataset captured in real educational environments are described. These were already described in the previous section.

The steps to follow to carry out the data capture experiment are listed below:

1. Upon arrival the student is cordially greeted.
2. You explain the experiment.
3. You explain the definitions and characteristics of the emotions.
4. You ask him/her if he/she has any questions.
5. You are given the *Informed Consent Format* to sign the authorization to record your data.
6. You ask to sit in front of the computer.
7. The use of the tutorial and the self-assessment is explained.
8. The heart rate sensor is placed on the left ring finger.
9. The Kinect, the video camera, the thermal camera, and the heart rate sensor are initialized.
10. Data recording starts.
11. Once the MOOC session is over, device recordings stop.
12. Captured data are saved in storage units.
13. Student is asked how he feels and is grateful for his participation by offering some candy, cookies, or a bottle of water.

In the second part, we explain legal, ethical, and environmental (physical place) aspects that should be considered in the design of a formal procedure. Regarding legal and ethical aspects, in Mexico, we will be respectful of the Federal Law on the Protection of Personal Data in Possession of Individuals DOF: 05/07/2010, published in the Official Gazette of the Federation on July 5th, 2010, this protects the privacy of individuals with respect to the treatment we give their personal information.

Its provisions are applicable to all individuals or corporations, public and private sector, both at the federal and state level, which carry out the processing of personal data in the exercise of their activities, therefore, companies such as banks, insurers, hospitals, schools, telecommunications companies, religious associations, and professionals such as lawyers, doctors, among others, are obliged to comply with what is established by this law.

Also, part of the procedure is the design of an informed consent letter with which the student authorizes their participation and the use of their data for research purposes. In this letter they are asked for their personal data, and they are explained what the experiment consists of.

In addition, they are informed that there is no consequence of deciding whether to participate in the experiment. They are informed that full confidentiality of their data will be kept and that they will only be used for research purposes. Finally, they are asked to authorize their participation by signing the letter.

4 Implementation of the Experiment

For the execution of the experiment, the voluntary participation of students in a learning activity while their behavior and physiological signals are being recorded through different data acquisition devices is necessary. In this section, we will describe the details about the execution of the data capture experiment based on the design of the procedure described above.

Information such as characteristics of the selected sample, place for data capture, devices used and their location, the way in which the data was collected, and creation of files with the captured data and their storage will be described. Finally, we will show a summary of the results obtained, the dataset created, and the statistical analysis of the information collected.

The following captures were carried out at the *Universidad Politécnica de Puebla* (UPPue) in its User Experience Laboratory, with the participation of engineering, masters, and doctoral students. In figure 2 we can observe the physical characteristics of the place.

At UPPue, 82 students from different engineering (information technologies, biotechnology, computer science and financial sciences) were captured. Statistics of the characteristics of the total experiments carried out so far as number of participants, age, sex, and level of studies, are shown in Table 1. Of the 82 participants, 50 are men and 32 women with an average age of 21 years.

The information shown in Table 2 was collected from each experiment. The data from experiment 1 is taken as an example. The first part is general data such as experiment number, date of capture, start time, end time, participant data. The second part corresponds to the emotions test responses, statistics of each of the sensors used and the video cameras.

At the end there is an observation section to describe some special situation that arose during the capture and needs to be documented. So far, the dataset contains the raw data sensed by each device used, stored in their corresponding files. Pulse sensor data is stored in 1 data file, *txt*. For thermal camera data the number of files varies depending on the scheduled capture time, if the programmed time is 40 minutes, 2400 *csv* files, 2400 *jpg* files and 1 *avi* file (4801 files) are stored.

Kinect video is stored in 1 *xed* file. And the webcam video is stored in 1 *wmv* file. Thus, by experiment, they are generated with a total of 4804 files, stored on an external 4TB hard drive. The summary of stored raw data statistics is shown in Table 3.

For the statistical analysis of the thermal camera temperature matrices, 360 more files are generated per minute to be analyzed, since the complete files of the face per minute are divided into 6 zones (front, nose, mouth, chin, right cheek, and left cheek) thus generating 60 new files per minute (for each area of the face) per 6 zones, in total 360 files per minute. Thus, the number of new files generated is increased in proportion to the number of minutes to be analyzed.

In the case of the analysis of the heart pulse files, 2 new files are generated for each minute to be analyzed, one to store the beats per minute and the other to store the signal power readings of the minute to be analyzed, if we analyze 6 minutes it gives us a total of 12 more files.

A summary of the number of experiments that have complete data files for each of the devices used is shown in Table 4. In this we can also see the number of experiments considered as complete, these are those experiments that are counted on with information on at least 4 of the devices used.

Thus, we have 78% of experiments with complete thermal camera files, 59% with useful pulse sensor data, 96% with full video and 89% with Kinect video. This represents 54% of experiments with complete data files from at least 4 devices.

5 Conclusion

The dataset contains the raw data counted by each device used and stored in their corresponding files. All these files make up raw data dataset. We also performed a summary of the number of experiments that have complete data files for each of the devices used.

The experiments considered complete are those experiments with which there is information on at least 4 of the devices used. So, we have 78% of experiments with complete files of the thermal camera, 59% with useful data from the pulse sensor, 96% with full video, 89% with video from the Kinect. This represents 54% of experiments with complete data files from at least 4 devices.

Constructing a dataset with physiological and behavioral signals of humans performing activities in real environments, both controlled and uncontrolled, is a complex, laborious and time-consuming task, especially when the challenge is to integrate data from various devices to capture different signs. Regarding the scope, a dataset was developed with captures from 2 physiological signal acquisition devices, heart rate and temperature.

Behavioral data of 82 students were obtained through video recordings. This fulfilled the objective of creating a dataset of physiological and behavioral information obtained from students learning in real online environments.

Regarding the contributions to the state of the art of automatic recognition of learning-centered emotions, they can be identified first because of the scarcity of dataset of physiological and behavioral signals of individuals performing learning activities, the creation of a formal procedure for the capture of data associated with learning-centered emotions.

This can be generalized for application in different virtual learning environments. In addition, a dataset with physiological signals of heart rate and facial temperature data is created, with behavioral images obtained from video recordings of the face of students performing learning activities. This type of data was not jointly identified in any of the works analyzed in the state of the art.

Most of the dataset found in the state of the art are created with data from a few students and consist only of video images, with a maximum of 9068 images. From this dataset, different research proposals may arise that complement the results obtained.

References

1. González, C., Mora, A., Toledo, P.: Gamification in intelligent tutoring systems. In: Proceedings of the second international conference on technological ecosystems for enhancing multiculturality, pp. 221–225 (2014) doi: 10.1145/2669711.2669903
2. Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., Taylor, J. G.: Emotion recognition in human computer interaction. IEEE Signal processing magazine, vol. 18, no 1, pp. 32–80 (2001) doi: 10.1109/79.911197
3. Ekman, P.: Emotions revealed: recognizing faces and feelings to improve communication and emotional life. New York: Times books (2003)
4. D'Mello, S., Graesser, A.: Dynamics of affective states during complex learning. Learning and Instruction, vol. 22, no. 2, pp. 145–157 (2012)
5. Ekman, P., Friesen, W. V., Hager, J. C.: Facial action coding system. The manual, Salt Lake City, Network Information Research Corporation (2002)
6. Koelstra, S., Muhl, C., Soleymani, M., Lee, J. S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A. Patras, I.: Deap: A database for emotion analysis; using physiological signals. IEEE transactions on affective computing, vol. 3, no. 1, pp. 18–31 (2012) doi: 10.1109/ TAFFC.2011.15
7. Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., Matthews, I.: The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression. In: IEEE computer society conference on computer vision and pattern recognition-workshops, pp. 94–101 (2010) doi: 10.1109/CVPRW.2010.5543262
8. El Kaliouby, R., Picard, R. W.: Affectiva Database. MIT Media Laboratory (2019) https://www.affectiva.com
9. González-Hernández, F., Zatarain-Cabada, R., Barrón-Estrada, M. L., Rodríguez-Rangel, H.: Recognition of learning-centered emotions using a convolutional neural network, Journal of Intelligent & Fuzzy Systems, vol. 34, no. 5, pp 3325–33636 (2018)
10. Botelho, A. F., Baker, R. S., Heffernan, N. T.: Improving sensor-free affect detection using deep learning. In: Artificial Intelligence in Education: 18th International Conference, pp. 40–51 (2017) doi: 10.1007/978-3-319-61425-0_4