Multimodal Recognition of Emotions with Application to Mobile Learning

Maria Lucia Barron-Estrada, Ramon Zatarain-Cabada, Claudia G. Aispuro-Gallegos
Instituto Tecnológico de Culiacán
Culiacán, Sinaloa, México
rzatarain@itculiacan.edu.mx

***Abstract*— A great variety of emotional recognition systems have been implemented, but few have been applied in the real world due to the high cost of the necessary technology and due to the low percentage of recognition effectiveness, when one does not work with spontaneous emotions. This paper presents the initial implementation of a system of multimodal recognition of emotions using mobile devices and the creation of an affective database through a mobile application. The recognizer works into a mobile educational application to identify user's emotions as they interact with the device. The emotions that the system recognizes are engagement and boredom. The affective database was created with spontaneous emotions of students who interacted with an educational mobile application called Duolingo and a mobile information gathering application called EmoData. The developed system has a regular percentage of effectiveness for audio recognition and position recognition. The corpus used for training is still small. We believe that as the number of registrations in the affective database increases, the precision will improve.**

Keywords—Intelligent-tutoring-system; artificial-intelligence; Affective computing; Mobile learning

I. INTRODUCTION

In recent years, the advance in the recognition of emotions has been impressive, generating a great variety of recognition systems [1] [2], affective databases [3], techniques for machine classification, which they manage to improve certain aspects such as increasing recognition accuracy and reducing processing time, among others.

The importance of recognizing emotions in learning has also been confirmed, defining it as a key component in the development of tutoring systems, and intelligent learning environments capable of responding to the student's affective needs. When a student is learning something new or applying their knowledge to solve an exercise, it goes through several emotional states, which can be confusion, frustration, boredom, commitment or other affective states [4].

Despite the advances in the recognition of emotions, some problems persist in such systems, such as requiring invasive and/or expensive sensors, devices, and databases in an adequate and realistic way for machine learning models. This produces problems related to generalization through contexts, time and individual differences [5]. On the other hand, it is necessary to collect spontaneous emotions on a large scale for which a valuable alternative to this problem may be the capture of data through mobile devices.

The main contribution of this work is the implementation of an emotional recognition system using mobile devices, which is easily integrated into an educational mobile application, to incorporate the ability to identify the affective state of the student and perform actions aimed at improving their learning. As part of the contribution of this work, we created an affective database through EmoData, the developed mobile application. Our work is innovative because we focused on detecting emotions related to learning through the natural use of mobile devices, which is why its field of applications is enormous.

The paper is organized as follows: Section 2 presents the related work; section 3 describes the Mobile Application EmoData; section 4 explain and discuss experiments and results and finally, conclusions are presented in section 5.

II. RELATED WORKS

The mobile recognition system presented by Wu et al. [6] captures images of the face, performs facial recognition based on distances between facial features, captures voice signals, and processes them based on Mel Frequency Cepstral Coeffi cients (MFCCs), Delta MFCC and Delta-Delta MFCC. This system uses a support vector to classify and recognize emotions neutral, happiness, anger and sadness. The percentage of average effectiveness achieved is 87%.

MoodScope [7] is a mobile application developed for the iOS operating system, which deduces the mood of the user based on the use of his cell phone, based on a statistical model that uses information from SMS, emails, phone calls, search engine Web and the location of the device. This application is transparent to the user, records the user's interaction with their cell phone, and stores the data on a server through a data connection or Wi-Fi.

EmotionSense [8] is a mobile application to detect the emotion of the user based on the habits of using the device. This application records the physical location where the device is located and how it is moving, the noise of the environment and monitors calls and messages, with the aim of developing a pattern of user habits that allow you to deduce your emotion. The components used are GPS, accelerometer and microphone.

Samurai fruit [9] is a variation of a game for mobile devices for iPhone that recognizes the emotion of the player based on the behavior of their tactile gestures on the screen of the device. This game consists of squeezing and cutting fruits through tactile movements; while the user is playing, the application captures and records the coordinates of each point of a stroke, the contact area of the finger at each point and the duration of the time of a stroke. The contact area is used to measure the pressure given by the players. The emotional states they identified are excited, Relaxed, Frustrated and Bored.

III. EMODATA IMPLEMENTATION

EmoData is a mobile application developed for the Android operating system, which is responsible for monitoring and collecting information on cell components to relate them to the affective state and the body posture of the user. The cellular components with which the application works are the microphone, accelerometer and gyroscope. Figure 1 shows the operation of the EmoData application used for the collection of device signals used to determine the user's emotion.

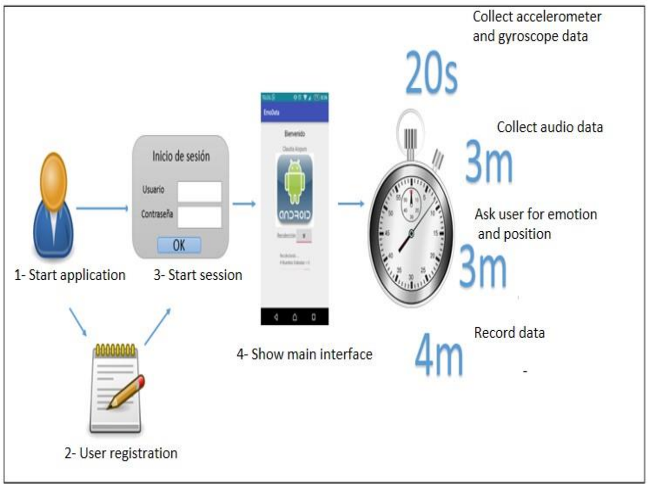

Figure 1. EmoData Operation.

The user starts the application and the system presents the initial interface of the system, where the user can register or log in. Once the user logs in, the application presents its main interface shown in Figure 2. In the main interface, the user indicates whether the collection task will activate or not. During the collection process, the application gathers every 20 seconds the values of the accelerometer and gyro axes. To collect the information from the accelerometer and gyroscope, the mobile application communicates with the sensor system and from an instance of the class *Sensor Manager* obtains access to these sensors to capture the values of the X, Y and Z axes at that moment. We collected and stored files with a duration of 3 minutes per period from the audio received by the microphone.

Every three minutes, the application asks the user to report their affective status and posture through the interfaces shown in c) and d) of Figure 2. We stored the information every four minutes, which consists of sending the information to a Web server in case of having an Internet connection, and if it is not possible, then we recorded the information locally to send it again later.

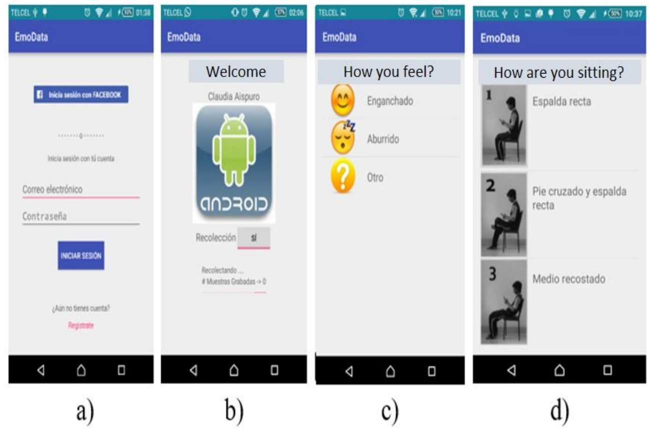

Figure 2. Main interfaces of EmoData.

IV. EXPERIMENTS, RESULTS, AND DISCUSSIONS

The experiment consisted of using the educational application Duolingo [10] and EmoData to generate the information that will be stored in the database. The test subjects were ten students of the Culiacán Institute of Technology of which seven were men and three women of different ages, ranging from 18 to 30 years. We created an affective database through a sequence of steps, described below:

Stimulus: Because we needed a tool with which we could capture the voices and emotions of a student speaking on a cell phone, we used the Duolingo educational application. DuoLingo presents different types of exercises that can be translations from English to Spanish and vice versa, pronunciation exercises, multiple-choice exercises, among others.

Signal collection: While the student was practicing with DuoLingo, the EmoData application recorded the audio and information signals from the accelerometer and gyroscope sensors.

Feature extraction: We implemented an application developed in Java that uses the MusicG [11] tool to extract the main features of tone and intensity audio recordings and a new feature extraction method for mean, standard deviation of the values of each of the axes of the accelerometer and gyroscope, and the coefficient correlation of the X and Y-axes.

Classification: We obtained the emotion classification by only considering the position and emotion reported by the student.

The emotional recognition system receives the user's affective signals and analyzes them to return the user's emotion as a result. We present in figure 3 the multimodal recognition system where it shows that the system receives as input the three affective signals: the audio, the accelerometer data and the gyroscope data. The method of fusion that is used to combine the modalities is at the decision level; This means that emotion recognition is first performed unimodally, identifying three emotions, one for each type of recognition made and then merging by means of a fuzzy system of rules.

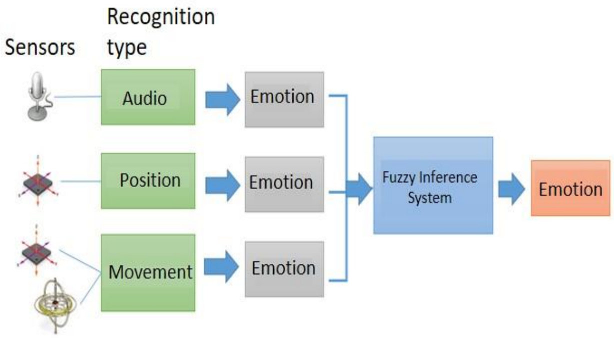

Figure 3. Multimodal recognition system.

To perform the fusion, we used the three emotions identified in the unimodal recognizers as input to a fuzzy inference engine by means of the jFuzzyLogic library [12]. The fuzzy system executes a set of fuzzy rules, to determine the final emotion that can be Engagement, Boredom or None. An example of a fuzzy rule is:

If eAudio is boring and ePosition is engagement and eMovement is engagement THEN emotion is engagement;

We obtained different experiment results performed for the accelerometer and gyroscope sensors to recognize the posture position. In the first case, we obtained a regular result of 73% and in the second case a bad result of 27%. In the case of audio recognition, we obtained values of 50%. Initial results are not very good but as the corpus of emotion registers increase, the accuracy of the recognition will gradually improve. We believe that the main challenge of this work, which is our important contribution to the research, was to implement the EmoData application.

V. Conclusions

The developed system is able to perform the recognition of emotions using the information generated by the sensors of a mobile device. The recognition system is implemented in such a way that it can be used by various types of mobile applications including educational applications such as tutoring systems, games or any other application that needs to incorporate the recognition of emotions continuously, non-invasively and in the natural environment for the user.

The recognizer system is available for use through a web service, which receives the input data (the audio in serialized wav format, the string in JSON format with the sequence of values of the axes of the accelerometer and gyroscope), and returns the emotion identified as output.

As future work, we are going to increase the number of corpus registers (this initial version was created with only 10 users), which will improve the accuracy of the recognition.

References

[1] K. D. Kory, Jacqueline Sidney, “A Review and Meta-Analysis of Multimodal Affect Detection,” vol. xx, no. x, 2015.

[2] S. Zhang and P. Hui, “A Survey on Mobile Affective Computing,” no. 1, 2014.

[3] O. Langner, R. Dotsch, G. Bijlstra, D. H. J. Wigboldus, S. T. Hawk, and A. van Knippenberg, “Presentation and validation of the Radboud Faces Database,” *Cogn. Emot.*, vol. 24, no. 8, pp. 1377–1388, 2010.

[4] S. D’Mello, “A selective meta-analysis on the relative incidence of discrete affective states during learning with technology.,” *J. Educ. Psychol.*, vol. 105, no. 4, pp. 1082–1099, 2013.

[5] R. S. J. D. Baker, S. K. D’Mello, M. M. T. Rodrigo, and A. C. Graesser, “Better to be frustrated than bored: The incidence, persistence, and impact of learners’ cognitive-affective states during interactions with three different computer-based learning environments,” *Int. J. Hum. Comput. Stud.*, vol. 68, no. 4, pp. 223–241, 2010.

[6] Y. Wu, S. Lin, and D. Yang, “A Mobile Emotion Recognition System Based on Speech Signals and Facial Images,” pp. 212–217, 2013.

[7] R. Likamwa, Y. Liu, N. D. Lane, and L. Zhong, “MoodScope: Building a Mood Sensor from Smartphone Usage Patterns,” *MobiSys ’13 Proceeding 11th Annu. Int. Conf. Mob. Syst. Appl. Serv.*, pp. 389–402, 2013.

[8] K. K. Rachuri, M. Musolesi, C. Mascolo, P. J. Rentfrow, C. Longworth, and A. Aucinas, “EmotionSense : A Mobile Phones based Adaptive Platform for Experimental Social Psychology Research,” Int. Stud., pp. 281–290, 2010.

[9] Y. Gao, N. Bianchi-Berthouze, and H. Meng, “What Does Touch Tell Us about Emotions in Touchscreen-Based Gameplay?,” ACM Trans. Comput. Interact., vol. 19, no. 4, pp. 1–30, 2012.

[10] https://es.duolingo.com/ retrieved January 30, 2018.

[11] https://code.google.com/archive/p/musicg/ retrieved January 30, 2018.

[12] P. Cingolani and J. Alcalá-Fdez, “JFuzzyLogic: A robust and flexible Fuzzy-Logic inference system language implementation,” in IEEE International Conference on Fuzzy Systems, 2012.