Embedded Emotion Recognition: Autonomous Multimodal Affective Internet of Things

Jose A. Miranda, Manuel F. Canabal, Marta Portela García, Celia Lopez-Ongil

Departamento de Tecnología Electrónica Universidad Carlos III, Madrid, Spain
Email: jmiranda@ing.uc3m.es

Abstract. – The term Internet of Things (IoT) is spreading out in the industry and in the academic world, specifically in those parts focused on making a better-connected world. On top of IoT, trying to gather the best user experience with an interconnected world, the Affective Internet of Things (AIoT) is being used. AIoT uses sensing technology empowered with the capability of detecting or predicting the emotional or affective state of the person. This new IoT branch can be used not only to provide a better user experience, in which the machine or device knows what the user likes, but also to solve real and current sociological problems by detecting those situations based on the user's emotion, such as sexual aggressions. In this paper, Bindi, a new autonomous multimodal system based on AIoT for sexual aggression detection, is proposed. Within this context, Commercial off-the-shell (COTS) sensors together with a light simplified embedded machine learning approach for emotion recognition have been implemented within a low power, low resource, wireless and wearable Cyber-Physical System (CPS) [1].

1. Introduction

Using technology for solving real and current sociological problems is a breakthrough challenge. Problems such as bullying, gender violence or domestic violence, require a deep sociological education adjustment, which is a long-term process. Meanwhile, tools to prevent these situations are needed to create a safer society. For example, in sexual aggression situations, when trying to prevent those from a technological and sociological point of view, a safe, trustable, and inconspicuous tool can provide a crucial help to the victims. Thus, this tool needs to be aware of the affective state of the user, i.e. to recognize specific emotional states of the user. Within the AIoT, a device using affective or emotion recognition can provide an early intervention help, which could interconnect responders' circles, emergency services, and others, to help the victim before or during sexual aggressions and can even gather the different evidences for further use. Thus, a wearable device including all these features can be the solution.

Research on emotion recognition in humans started decades ago [1], [2]. In most of the cases, the proposed emotion recognition systems are based on a unimodal approach. For example, in [3], the authors propose an emotion detection application in driving fatigue, which involves analysis of the face image acquired by a camera. In [4], a survey on speech emotion recognition systems is presented, in which the voices from the different users are used as data input. There is a vast body of literature on the automatic

[1] The work described in this paper is patent-pending.

emotion recognition based on unimodal frameworks. One of the disadvantages of these systems is that the information used for extracting or detecting the emotion only comes from one source. This could lead to a loss of information, as the variations observed into the acquired variable could not be enough to detect the affective state with a high accuracy. On the other hand, there is literature [5], [6], that proposes multimodal systems for emotion recognition. Multimodal systems provide a clear advantage w.r.t. unimodal ones from a statistical point of view, making the decision process more robust and, at the end, more accurate. However, in most of the literature, these systems are conceived under the concept of a general emotion recognition system, i.e. a system used for detecting any emotional state over the external applied stimulus, based on different emotional metrics. Moreover, they are mostly used in laboratory facilities, where the complexity, size and other parameters of the equipment, rather than functionalities or capabilities, are not a concern. They even use expensive clinical equipment, such as ProComp devices [7], to acquire all the physiological variables under analysis. Therefore, the applicability of these multimodal systems to a usable wearable solution is not clear, as a correct integration of measures from various sensors as well as low power consumption and inconspicuous is not achieved easily.

On this basis and going towards the tool to prevent sexual aggression situations, this work proposes to build an autonomous multimodal wearable system, Bindi[2], based on physiological variables, environmental and user audio. To this end, and trying to achieve the maximum simplicity, COTS sensors, a light and simplified embedded intelligent system following an approximate computing approach, and wireless capabilities, are used to provide a system ready to work in a real application, having in mind not only the technological challenges such as the power consumption, security of the communications, etc., but also the sociological issues such as the inconspicuous or stigmatization character of the device. Through a deep sensor data analytics and a light embedded machine learning approach [8], authors have come up with a new framework for sexual aggression detection, all integrated into a wearable CPS.

The paper is organized as follows. Section 2 describes the current solutions in emotion detection for gender violence using wearable devices. Section 3 describes the proposed wearable and autonomous multimodal emotion recognition system. Section 4 details the problems observed in the current system and how they are going to be addressed. Finally, Section 5 concludes the paper.

2. Emotion recognition on wearable devices

There are already commercial solutions claiming to detect affective states, some of them using wireless sensing technology and others using the smart phones embedded technology by means of a mobile application. Among these solutions, there are wearable devices. For example, FEEL [9] is a bracelet with four sensors to measure Galvanic Skin Resistance (GSR), PhotoPlethysmoGraphy (PPG), Skin Temperature (SK) and Inertial Motion (IMU – accelerometer). This device detects primary emotions such as joy, sadness, and happiness, but it does not realize further actions with this

---

[2] Bindi is an autonomous multimodal AIoT system designed in the Universidad Carlos III de Madrid by UC3M4Safety team. The authors of this paper, working in the Electronic Technology Department, have been focused on physiological variables sensing and processing.

information. Another device is Embrace [10], also a bracelet focused on early detection of epilepsy seizures through GSR data. Most of the commercial solutions just acquire real-time physiological data without even relate these variables with any emotion, such as E4 [11], another bracelet with real-time physiological monitoring capabilities. Specifically, when looking for technological solutions related with the prevention or detection of sexual aggressions, there are devices including panic buttons with communication and geolocation capabilities, such as SAFER PRO [12], which is a mobile-independent panic button with GSM and GPS enabled, or NIMB [13], which is another panic button with similar features. Up to the knowledge's author, there are no commercial devices that integrate physiological or another human variable tracking with panic button or geolocation capabilities. Within this context, Bindi integrates an autonomous multimodal framework together with current commercial devices capabilities, panic buttons and GPS. The next section details the proposed system.

3. Proposed system

Bindi, Fig. 1, has been designed and implemented to help in the struggle against sexual abuse, by providing a tool that could act as an autonomous system, which is essential in scenarios where the victim is not able to ask for help. The target of this system implies numerous aspects, from technological to sociological fields. The system must be a wearable solution to be carried daily by the user. In this line, there are others factors that need to be fulfilled such as safety, low power consumption, privacy and wireless communications. All these areas apply to all wearable systems but in this case, an inconspicuous need is strongly required, to avoid any removal from aggressors or the victims' stigmatization. Moreover, Bindi must work without user interaction, in an autonomous way, detecting blocking states of panic. The proposed system is composed of three devices. The first device is a bracelet that provides physiological variables monitoring by means of three different sensors: GSR, PPG and SK. It performs the first trigger w.r.t. machine learning algorithms of the system using the average raw data value of each sensor for a temporary window of ten seconds. Specifically, the machine learning algorithm used is a K-Nearest Neighbors (KNN), which has been implemented following an ad-hoc training (unipersonal training) leading up to 85% accuracy, as explained in [8]. The acquisition process and other relevant information related to the specific sensors and circuity is detailed in section 4. The second device is a pendant, which acquires the audio through a Microelectromechanical System (MEMS) microphone and performs specific pre-processing. Acquired audio signal is sent wirelessly for further processing; due to the limited bandwidth of the wireless communication, the audio is compressed. Finally, the third device is a smartphone, which acts as the central unit of the system. It connects to the two previous devices, makes all the data fusion and realizes further processing to provide a robust trigger based on the user's emotional state. Apart from the autonomous trigger, these three devices have a panic button. The different generated alarms are forwarded along with the GPS location to a net of contacts or the emergency services through smartphone internet connection, GSM/GPRS, and SMSs (for the cases of low connectivity).

One of the novelties of Bindi is the panic detection, which is performed by means of mapping changes on the acquired variables with specific emotions using the Pleasure-Arousal-Dominance-Familiarity (PADF) space [6], which is a four-dimensional space formed by the level of enjoyment, activation, control, and internalization over the external presented stimulus (i.e. *Valence, Arousal, Control* and *Dominance*). **Bindi** has been successfully prototyped and tested, using an in-house developed software tool for training and monitoring volunteers, as explained in [8]. Further research and development should be done in terms of signal acquisition and processing refinement, emotion inference and statistical tests on larger set of volunteers, to propose a usable solution to prevent sexual aggressions on women. Some of this work is detailed in this document.

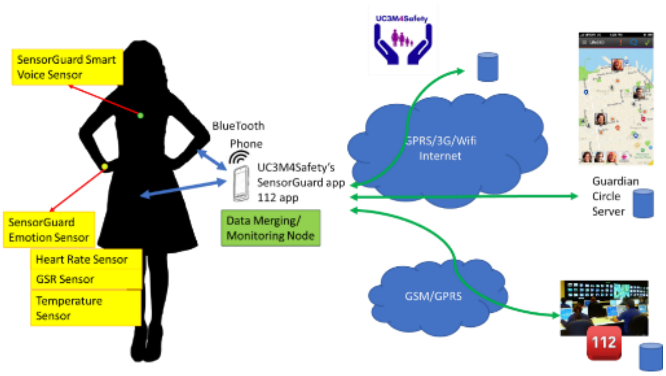

Fig. 1. Bindi system level architecture

4. Problems Encountered with Physiological Variables Acquisition

In this section, specific problems related to the current development status are detailed and different solutions are proposed.

4.1. Heart Rate: Motion Artifact Removal

PPG sensors provide the Blood Volume Pulse (BVP) raw signal. Applying a post-processing algorithm to this signal (obtaining the frequency of systolic points), Heart Rate (HR) can be obtained. PPG signals are highly susceptible to noise and, there are existing solutions proposed in the literature to solve it. For example, in [14] an accelerometer is implemented to correct the noise produced by the movement. But, there are more than one source of noise in this signal and the accelerometer may not detect all of them. In fact, the PPG signal can be expressed by the following expression:

$$S_{PPG} = S_{AMB} + S_{VAS} + S_{MECH} + S_{ELEC}$$

The first term is referred to the ambient light changes detected by the photodetector of the PPG sensor. The second term is related to the volumetric veins or arterial changes under the skin. The third term is the effect of sensor movement in placement and orientation relative to the skin. And the last one is the electrical noise due to the hardware implementation. In Bindi, MAX30101 by Maxim Integrated™ is used. This is a high sensitivity pulse oximeter and heart rate sensor for wearable health. It includes all the necessary front-end circuitry to ease the design-in process. The data acquisition is done through Inter-Integrated Circuit (I2C) communication. For the sake of the system simplicity and prize, Bindi does not have an accelerometer intended to suppress the noise due to movement. For the first prototype of Bindi, a band pass filter (0.5 – 4 Hz, i.e. 30 – 240 bpm) and a four second moving average of the BVP data have been applied to get rid of the noise. For example, when having no movement activity, as it can be observed in the first plot of Fig. 2, the BVP is recovered successfully.

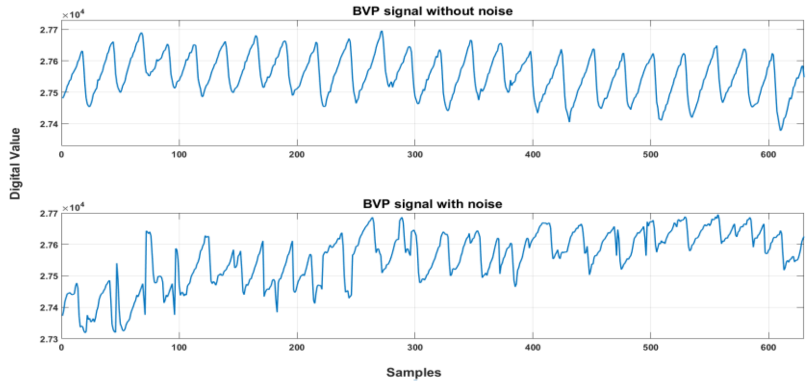

Fig. 2. BVP signal without and with strong movement activity

However, when having movement activity and the same beats per minute, the obtained results are not as accurate as desired, second plot in Fig. 2. If a peak-detection algorithm is run through this wave, it will detect different inter-beat intervals of heart rates. Therefore, the HR estimations are affected as the BVP data differ in the two cases referred, due to the different noises. For this reason, three different proposals, found in the literature, are under research to solve this problem. First, usage of the same or similar preprocessing stage and adding independent component analysis together with adaptive filters, as stated in [15]. Second, application of a preprocessing stage using variance characterization series, empirical mode decomposition and singular value decomposition. The obtained output is afterwards filtered by a novel 2-D filtration strategy based on Harr wavelet transform, as stated in [16]. And, third, implementation of a differential measurement using two different photodetectors and study the common-mode rejection between these two measurements, as stated in [17]. Moreover, a complementary study is planned, including two LED sensors, one of them being in contact with the skin and the other not, to detect only motion artifact, as in [18].

4.2. GSR: Multi-component and body location

GSR is one of the most important signals to be acquired when detecting any possible human emotion [19], as this variable is directly related with the level of *Arousal* or *Activation* over the external stimulus. In Bindi, the front-end circuitry using two silver-silver chloride electrodes (Ag-AgCl) and two amplification stages have been implemented. An Analog to Digital Converter (ADC) is used to acquire the measured voltage, which is directly related with the impedance of the skin. A four second moving average of this raw data is used. However, GSR can be decomposed into different components, rather than just the raw signal: a tonic component (slow changes – basal skin conductance level) and a phasic component (rapid changes – specific and non-specific event related levels). In the next prototypes of the system, these components will be extracted. This could lead to a more accurate system, as Event-Related Skin Conductance Responses (ER-SCR) which can provide more information related to the external stimulus. Moreover, no extra hardware acquisition is needed to obtain these components. A deep study on to the relevance of these features for the specific Bindi implemented machine learning, that is data analytics, is currently under process.

The body location of the GSR electrodes is also a crucial parameter when designing the system. Marieke van Dooren et al. in [20] study the responsiveness and similarity of GSR measurements on 16 different skin conductance measurements locations. In that work, *Arousal* and *Valence* measurement experiments to see the quantifiable differences in terms of emotion recognition have been conducted under different humidity and temperature conditions. To study the location dependability for the GSR sensor in Bindi, different body locations are being tested for future prototypes.

4.3. Temperature: Feature extraction and body location

In Bindi, MAX30205 by Maxim Integrated™ is used. This device converts the body temperature measurements to a digital data using a high-resolution, ΣΔ analog-to-digital converter. The communication is done through I2C. Afterwards, a four second moving average of the acquired raw temperature data is used. Different temporal and frequency features to be extracted from the raw data are under consideration. The extraction of new features from this variable would not suppose any extra hardware.

Based on different studies, such as [21], the body temperature turns out to provide relevant information related with the affective or emotional states. In the literature, different studies or experiments can be found, based on the emotion detection through the body temperature. For example, in [22], the authors propose using the fingertip temperature; in [23], the authors propose using the facial temperature; thus, there is no standard location identified for this. However, the relationship of the variable with the identification of different emotional situations is clear and has been confirmed, Kataoka et al. [24] investigated and found a direct relationship between stressful tasks and the skin temperature. Moreover, the reason for choosing a place near to the hand is because the sympathetic innervations of the arteriovenous anastomoses are densely distributed on the palms zone [25]. To study the location dependability for the temperature sensor in Bindi, different body locations is being tested for future prototypes.

5. Conclusions and future work

In this paper, Bindi, a new autonomous multimodal system based on AIoT for sexual aggression detection, is presented. Within this context, Commercial off-the-shell (COTS) sensors together with a light simplified embedded machine learning approach for emotion recognition have been implemented within a low power, low resource, wireless and wearable Cyber-Physical System (CPS). A first prototype of Bindi has been designed, developed and implemented using simplified hardware and software techniques. The obtained results have been satisfactory in terms of accuracy and reliability for detecting fear or panic situations based on the offline training performed, obtaining an 85% of accuracy (panic detection accuracy) [8]. Different technical problems have been identified and actions for addressing those are being under research and development.

References

1. R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, J.G. Taylor “Emotion recognition in human-computer interaction”. IEEE Signal Processing Magazine, 2001. pp 32-80. DOI: 10.1109/79.911197.
2. J. Behav. “Measuring emotion: the self-assessment manikin and the semantic differential.” Ther. Exp. Psychiatry, vol. 25, no. 1, pp. 49–59, M. M. Bradley and P. J. Lang, 1994.
3. A. Tischler, Martin & Peter, Christian & Wimmer, Matthias & Voskamp, Jörg. “Application of emotion recognition methods in automotive research”. The Open Automation and Control Systems Journal, 2015, 7, 655-660.
4. Moataz El Ayadi, Mohamed S.kamel. “Survey on speech emotion recognition: Features, classification schemes, and databases”. Patter Recognition 44, 2011, 572-587.
5. R.W. Picard “Toward Machine Emotional Intelligence: Analysis of Affective Physiological State”. IEEE T. on Pattern Analysis and Machine Intelligence. Vol. 23, No. 10, October 2001.
6. Sander Koelstra “DEAP: A Database for Emotion Analysis using Physiological Signals” IEEE T. Affective Computing, 2012.
7. Thought Technology Ltd. “ProComp2™ Quick Start and User’s Guide”. Manual number SA7420, rev3.0.
8. Jose Angel Miranda Calero, Rodrigo Marino, et. al. “Embedded Emotion Recognition within Cyber-physical systems using physiological signals”. Design of Circuits and Integrated Systems Conference, November 2018. Accepted.
9. G. Eleftheriou, P. Fatouros, and C. Tsirmpas, “Unobtrusive emotion recognition system,” Jan. 18, 2018, uS Patent App. 15/648,730.
10. Simone TOGNETI, Ivan CENCI, Daniele RESNATI, Maurizio Garbarino. “Apparatus for electrodermal activity measurement with current compensation”. Sep. 25, 2016. uS Patent App. 61/802,519.
11. Mladen Barbaric, Minkyu Choi, Kihyun Kim. “Wearable Biosensing Device”. Jul. 28, 2014. uS Patent App. 29/509,637.
12. LEAF WEARABLES. 2018. SAFERPRO. [ONLINE] Available at: https://www.leafwearables.com/products/safer-pro?variant=12367579873358. [Accessed: 28 July 2018].
13. NIMB. 2018. [ONLINE] Available at: http://nimb.com. [Accessed: 28 July 2018].
14. S. H. Kim, D. W. Ryoo, and C. Bae, “Adaptive noise cancellation using accelerometers for the PPG signal from forehead,” in Proc. Conf. IEEE Eng. Med. Biol. Soc., Aug. 2007, pp. 2564–2567.
15. Peng. F., Zhang Z., Gou Xiaoming. “Motion artifact removal from photoplethysmographic signals by combining temporally constraint independent component analysis and adaptative filter”. Biomedical Engineering OnLine 2014, 13:50.
16. Pang Bo, Ming Liu. “A novel approach framework based on statistics for reconstruction and heartrate estimation from PPG with heavy motion artifacts”. SCIENCE CHINA, February 2018, Vol. 61 022312:1-022312:12.

17. Yuta Kiboyama. “Motion Artifact Cancellation for Wearable Photoplethysmographic Sensor”. Thesis, MIT, Department of Electrical Engineering and Computer Science, July 9, 2010.
18. Shimazaki Takunori, Hara Shinsuke, et. al. “Cancelation of Motion Artifact Induced by Exercise for PPG-Based Heart Rate Sensing”. 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2014.
19. G. Valenza, A. Lanat`a, E. P. Scilingo, and D. De Rossi, “Towards a smart glove: Arousal recognition based on textile electrodermal response,” in Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE. IEEE, 2010, pp. 3598–3601.
20. Marieke van Dooren, et. al. “Emotional Sweating across the body: Comparing 16 different skin conductance measurement locations”. Physiology & behavior 106 (2012) 298-304.
21. Lauri Nummenmaa, Enrico Glerean, et. al. “Bodily maps of emotions”. PNAS, January 2014, vol.111, no.2., 646-651.
22. G.Shivakumar, P. A. Vijaya. “Emotion Recognition Using Finger Tip Temperature: First Step towards and Automatic System”. International Journal of Computer and Electrical Engineering, Vol. 4, No. 3, June 2012.
23. Liu Z., Wang S. (2011) Emotion Recognition Using Hidden Markov Models from Facial Temperature Sequence. In: D’Mello S., Graesser A., Schuller B., Martin JC. (eds) Affective Computing and Intelligent Interaction. ACII 2011. Lecture Notes in Computer Science, vol 6975. Springer, Berlin, Heidelberg.
24. Kataoka et. al. “Development of a skin temperature measuring system for non-contact stress evaluation”. Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, vol. 20, No. 2, 1998.
25. Walløe L. “Arterio-venous anastomoses in the human skin and their role in temperature control”. Temperature: Multidisciplinary Biomedical Journal. 2016;3(1):92-112