

Multimodal biometric recording architecture for the exploitation of applications in the context of affective computing

Enrique P. Calot and Jorge S. Ierache

Laboratorio de Sistemas de Información Avanzados,
Facultad de Ingeniería, Universidad de Buenos Aires,
Buenos Aires, Argentina

`ecalot@lsia.fi.uba.ar`, `jierache@yahoo.com.ar`

Abstract. In affective computing, it is important to design techniques that allow devices to acquire emotional states. To create and test these techniques it is necessary to have datasets that have several modalities namely, keystroke dynamics, electroencephalography, facial expressions, voice tone, heart rate, among others. This article presents a multimodal dataset that allowed us to detect the subjectivity that subsists in certain modalities —as are the surveys— and that is often overlooked, against objective modalities such as keystroke dynamics and electroencephalography. This article presents the creation of an environment in order to acquire a multimodal dataset. Work has also been done on the analysis of brain waves and their correspondence with other modalities.

Keywords: biosignal devices, affective computing, multimodal acquisition

1 Introduction

Affective computing is the study and development of systems and devices that can recognize, interpret, process and stimulate human emotions. Rosalind Picard defines that affective computing is “information technology that relates to emotions, not only those considered more important, such as joy or sadness, but also interest, boredom or frustration, which are those that are given in relation to computers” [28]. The emotional education of an individual in a multimodal context allows the simultaneous registration of different biometric information that allows deducing the emotional state of the individual. Affective computing and the development of applications in this area require emotion acquisition and learning. Future systems will interact with humans empathically with a diverse repertory in multimodal biometric contexts as well as the stimulation techniques.

Our goal is to create an empathic machine proficient in interpreting the human emotions and acting in consequence [29]. Controlling emotions of users shall increase productivity, minimize human errors and even prevent major disasters.

In order to study correlations between different acquisition modalities we created a dataset with 12 voluntary participants and recorded audio, video, heart

rate, keystroke dynamics and brain activity. To record different emotions we generate different emotions using visual and sonorous stimuli. Knowing how each modality corresponds to another will help predict emotions with only a keyboard or built-in microphone avoiding intrusive devices such as electroencephalographers (EEG).

2 Emotion modelization

In psychology, to model emotional states, there are different approaches, namely a dimensional and evaluation-based approach[11] and a categorical one. The former states that there is a small number of basic emotions, strongly connected in our brain and universally recognized[5]. This theory about the universality and interpretation of non-verbal affective expressions in terms of basic categories has been the most commonly adopted approach in research on automatic measurements of human affectivity. However, some researchers have shown that in everyday interactions, people have non-basic, subtle and quite complex affective states such as thinking, shaming or depression. Such states can be expressed through dozens of facial and body expressions, audio or anatomically possible physiological cues. Therefore, a single denomination (or any small number of discrete classes) may not reflect the complexity of the affective state conveyed by those rich sources of information[30]. For this reason, a group of researchers advocate the use of the dimensional description of human affectivity, where affective states are not independent of each other.

They are related to each other in a systematic way[8,30,21,32]. The most widely used dimensional model is a circular configuration called *Circumplex of Affectivity* developed by [30] (see Fig. 1(a)). This model is centered on the hypothesis that each basic emotion represents a bipolar entity that is part of the same emotional continuum. The proposed axes or poles are arousal (relaxed *vs.* excited) and valence (pleasure *vs.* dislike).

In this work we are using the dimensional approach employing the circumplex.

3 Stimulation

Each session took approximately one hour, although the time depended on the participant. The first step was to bring the subjects to a similar initial state considered neutral and free of any stimulus. This step consisted of a half minute on average of data entry, one minute relaxing with silence, 4 minutes and a half with a game and 3 minutes to obtain responses from his state (considered neutral or initial). The game consisted of using the mouse to grab certain colored dots on the screen while being played a relaxing background music. Then the four phases started in the order: happiness, calm, sadness, and Anger lasting approximately 13 minutes each. The experiment sweeps the four quadrants of the arousal–valence model[2] (see Fig. 1(b)) in a clockwise path moving over each phase in which the same stimulatory sequence is applied. They combined five

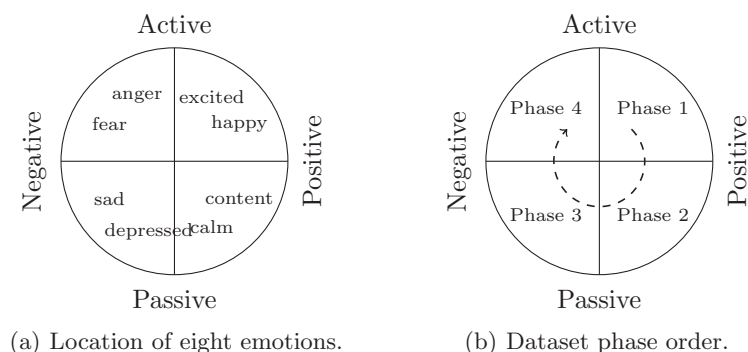


Fig. 1: Arousal-Valence Model.

minutes of visual stimulation (images), five minutes of audiovisual (videos and audio) and around three other for the responses to the current state (self-report, voice, and typing).

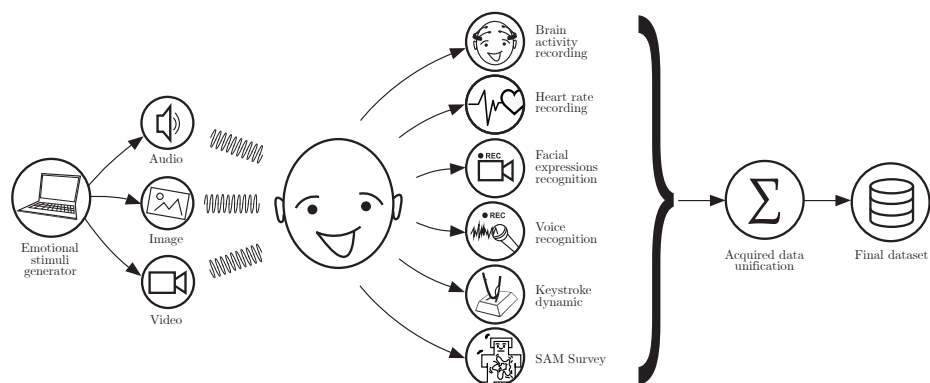


Fig. 2: Multimodal acquisition process.

To perform the visual stimulation, we requested a set of data compiled by [18] named IAPS. The set contains an image database used in various scientific studies related to emotions and is considered a de facto standard. We have also downloaded four videos with a high emotional content. Firstly, a subset of images taken from IAPS (100 images, 25 per phase) is shown, and then a video or audio, whose theme is related to the transient affective phase. For each phase, upon completion of image, video and audio stimulation, the subject is asked to complete a self-report using the Self-Assessment Manikin (SAM) survey proposed by Lang[17] (see Sect. 4.6). Then, we proceed to capture voice and keystroke dynamics. See Fig. 2.

4 Data acquisition and modalities

Several input interfaces had been used to create this dataset. The following sections summarize the main features of each modality.

4.1 Encephalography (EEG)

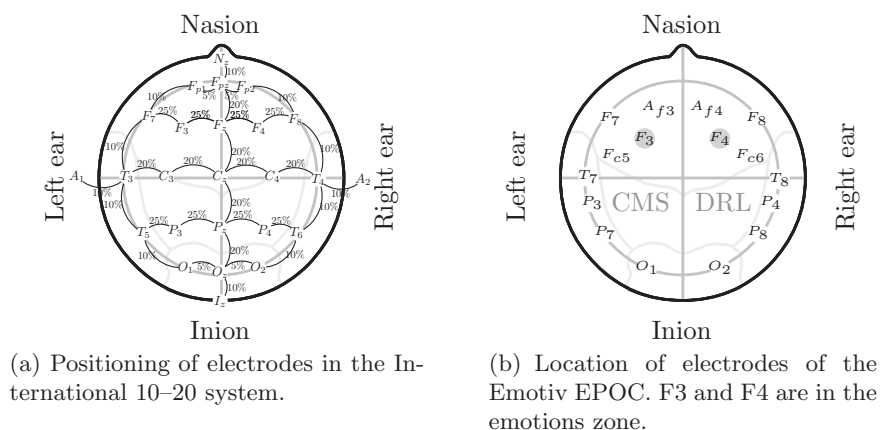


Fig. 3: Location of EEG electrodes.

A Brain-Machine Interface (BMI) facilitates communication between mental or cognitive functions created from a person's brain, by capturing electrical signals, to be processed, classified and communicated with specific applications or devices[15,13,14]. This dataset used the Emotiv EPOC headset for the acquisition of encephalographic signals by means of 14 electrodes, which were initially considered those located in the prefrontal cortex (F3, F4, AF3, AF4, according to the International 10-20 system[16], see Fig. 5). OpenViBE was employed to receive and store the raw samples and then apply filters to the electrodes and perform post-processing. This can be used to calculate the position of the emotional state in the bidimensional arousal-valence model[6,25].

4.2 Heart rate

The heart rate was acquired through a wristband placed on the participant (Xiaomi Mi Band 1s), which records the heart rate continuously over time. It is a polycarbonate band with an optical heart rate sensor that weights 14.5g.

4.3 Video

Video capture is performed continuously at all times. To analyze emotional gestures it is possible to use the tool `clmtrackr`[26], which detects faces and then

interprets emotions utilizing a training from a database of mapped points. The implementation uses constrained local models fitted by regularized landmark mean-shift developed by Audun Mathias Øygaard and based on [31]. To detect the emotional states, the MUCT database[22] was adopted, supplementing some extra images of Øygaard. The emotional model provides it with 18 coefficients[3]. The video can also be analyzed with any other tool.

4.4 Audio

All audio tracks are preserved in stereo for further analysis. The participant is required to read a text that is a different shuffle of the corresponding sentences employed for keystroke dynamics.

4.5 Keystroke dynamics

Keystroke dynamics are acquired after the stimulation, requesting the participant to enter a list of sentences of a neutral emotional load. Those passages had been randomly shuffled through the four phases. Both, the flight time and the hold time, are stored in order to reconstruct all keyboard events. In this regard, important advances were made in the laboratory, with a new method for classifying patterns and a replication of one of the most commonly used method[7]. A framework in Python has been coded in order to provide functions to analyze keystroke data. It includes a library compatible with standalone Python and Jupyter notebooks with a Python kernel.

4.6 Self-report (SAM)

A last-minute survey is conducted asking the participant about his current emotional state. It makes use of the widely adopted methodology to capture emotions named Self-Assessment Manikin (SAM) proposed by Lang[17]. As described by Bradley and Lang[2], SAM is an easy-to-administer non-verbal method for quickly assessing the pleasure, arousal, and dominance associated with a person's emotional reaction to an event.

It consists of 5 drawings that characterize if the person is happy or sad (valence) and how excited (arousal). For each domain, there are 9 points, one per figure and an intermediate point between them.

The Fig. 4 presents the paper and pencil version of SAM, illustrating its non-verbal representation, with figures of various points along each of the two dimensions. SAM extends from a smiling and happy figure to a frowned and unhappy figure when the valence dimension is represented and extends from an excited figure with eyes open to a relaxed sleepy figure for the arousal dimension[2]. We have omitted a dominance dimension to simplify the experiment.

SAM has been used effectively to measure emotional responses in a variety of situations, including reactions to photos [10,19], images[23], sounds[1], advertisements[24], painful stimuli[20], among others. In addition, SAM has been applied to children[9], patients with anxiety disorders[4], people with similar phobias[12], psychopaths[27], and other clinical conditions[2].

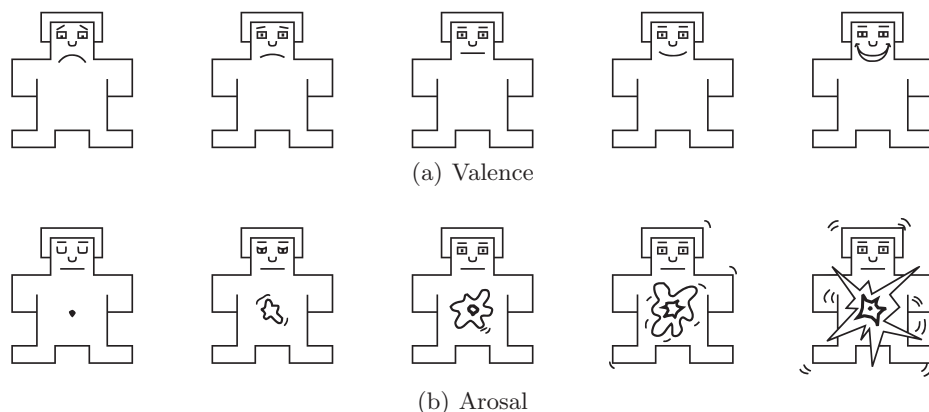


Fig. 4: SAM and its different states along the dimension axis

5 Environment description

In total 12 subjects performed all the four phases plus two researchers who performed an initial capturing test. The captures were made in December 2016 during the period of two weeks. The construction of the space, the programming of the testing interface and the devices setup took roughly six months of preparation.

To avoid environmental distractions, the acquisition was conducted in a conditioned room within the facilities of the Laboratory of Advanced Information Systems at the Faculty of Engineering of the UBA. This space was named as Emotional Immersion Chamber (EIC), consisting of approximately 4m², it is communicated with an adjacent control room where researchers enduringly supervise connectivity and captures. It has also a waiting room where the participants stay while the researchers prepare the equipment and begin the experiment. The room is isolated in both, sound and visibility, the windows are covered with green panels to facilitate the extraction of images from the video. Sound proofing had been performed to measure the level of noise and distraction in the space.

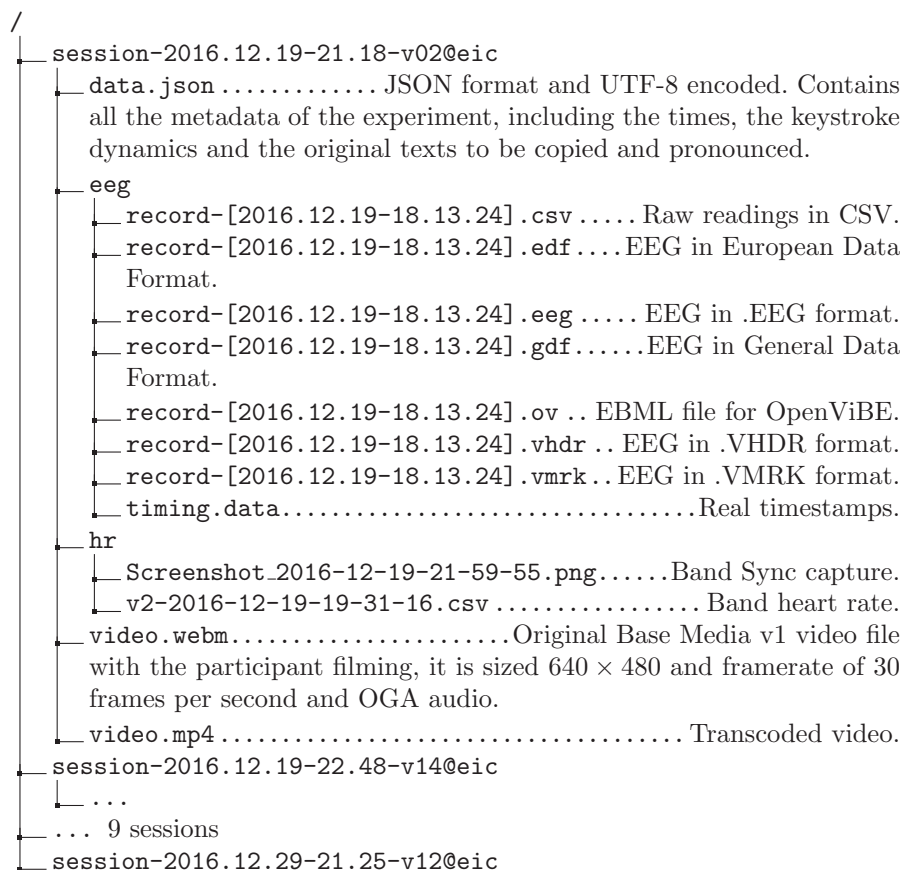
6 Data analysis

The resultant dataset consists of several files organized in directories as shown in Table 1.

There is an infinite number of opportunities to exploit the acquired data. Not only the analysis of the correlations between the different modalities but also it is possible to obtain emotional patterns that are uniquely evident in some of them. In the specific case of EEG interpretation, as a preliminary research, Fig. 5 displays a prediction of the dimensions of arousal and valence of subject

7. The first white band corresponds to the relaxation time, the four next bands represent each phase in the stimulation —dark gray for visual and light gray for audiovisual— and response —white— parts respectively. Dashed lines indicate phase transitions. It is projected to develop this line by replicating other available methods.

Table 1: Dataset structure



The raw information related to emotional states was successfully captured inside the EIC. Although some problems with the real time synchronization were detected at the beginning they were solved and a major dataset had been produced. The EIC is now available for further research on the topic. We created several Jupyter notebooks[33] to analyze all the raw information.

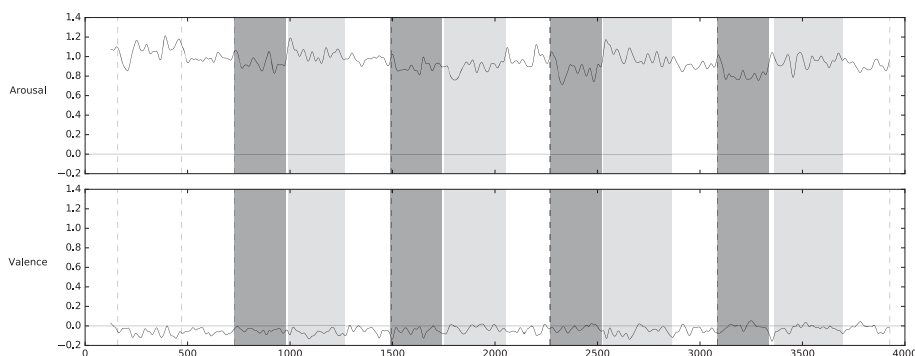


Fig. 5: Predicted arousal and valence from subject 7. Time is measured in seconds.

6.1 Future lines of research

From the collected data set, several research lines are drawn, which are intertwined in a single main objective: the generation of redundancy from the inference of emotional states through the use of multivariate biosignals.

In that sense, at first, we will seek to establish an emotional pattern for each individual through the interpretation of their respective brain activities. In addition, the generalization of this inference will be explored in order to determine the existence of a common framework that allows the emotional models of individuals to be independent.

To complete the studies on the inference of the emotional profile of an individual it will be used the heart rate as an indicator of the level of excitement that, together with keystroke dynamics, brain waves and facial recognition, will allow the construction of a first prototype of an empathic machine.

Acknowledgments. This paper acknowledges support from the University of Buenos Aires with the project under UBACyT 2014-2017 GEF localized in the Laboratorio de Sistemas de Información Avanzados, Engineering Faculty, University of Buenos Aires. The project recognizes the collaboration from the ISIER (FICCTE UM) assisting only in BMI application concerns, under the PID 01-001/12/14.

References

1. Bradley, M.M.: Emotional memory: A dimensional analysis, pp. 97–134. Taylor & Francis, Hillsdale, NJ: Erlbaum (1994), <https://books.google.com.ar/books?id=H0TrAgAAQBAJ>
2. Bradley, M.M., Lang, P.J.: Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry* 25(1), 49–59 (1994)

3. Calot, E.P., Rossi, F.M., Gonzalez, N., Hasperué, W., Ierache, J.S.: Avances en educación de dinámica de tecleo y el contexto emocional de un individuo aplicando interfaz cerebro computadora. pp. 872–876 (jun 2016), <http://hdl.handle.net/10915/53247>
4. Cook, III, E.W., Melamed, B.G., Cuthbert, B.N., McNeil, D.W., Lang, P.J.: Emotional imagery and the differential diagnosis of anxiety. *Journal of Consulting and Clinical Psychology* 56(5), 734–740 (1988), <http://dx.doi.org/10.1037/0022-006X.56.5.734>
5. Ekman, P., Friesen, W.V.: *Unmasking the Face: A Guide to Recognizing Emotions from Facial Clues*. Malor Books (2003), <https://books.google.com.ar/books?id=TukNoJDgMTUC>
6. Feldman, L.A.: Valence focus and arousal focus: Individual differences in the structure of affective experience. *Journal of personality and social psychology* 69(1), 153 (1995)
7. González, N., Calot, E.P., Ierache, J.S.: A replication of two free text keystroke dynamics experiments under harsher conditions. In: 2016 International Conference of the Biometrics Special Interest Group (BIOSIG). pp. 1–6 (Sep 2016), <http://ieeexplore.ieee.org/document/7736905/>
8. Grandjean, D., Sander, D., Scherer, K.R.: Conscious emotional experience emerges as a function of multilevel, appraisal-driven response synchronization. *Consciousness and cognition* 17(2), 484–495 (2008)
9. Greenbaum, P.E., Turner, C., Cook, III, E.W., Melamed, B.G.: Dentists' voice control: Effects on children's disruptive and affective behavior. *Health Psychology* 9(5), 546–558 (1990), <http://dx.doi.org/10.1037/0278-6133.9.5.546>
10. Greenwald, M.K., Cook, III, E.W., Lang, P.J.: Affective judgment and psychophysiological response: Dimensional covariation in the evaluation of pictorial stimuli. *Journal of psychophysiology* 3(1), 51–64 (1989)
11. Gunes, H., Schuller, B., Pantic, M., Cowie, R.: Emotion representation, analysis and synthesis in continuous space: A survey. In: *Automatic Face & Gesture Recognition and Workshops (FG 2011)*, 2011 IEEE International Conference on. pp. 827–834. IEEE (Mar 2011)
12. Hamm, A.O., Globisch, J., Cuthbert, B.N., Vaitl, D.: Startle reflex modulation in simple phobics and normals. *Psychophysiology* 28, 228 (1991)
13. Ierache, J.S., Pereira, G., Calot, E.P., Iribarren, J.: Framework for brain computer interface implemented to control devices in the context of home automation. In: *XIX Congreso Argentino de Ciencias de la Computación*. No. 5736 (2013), <http://lsia.fi.uba.ar/papers/ierache13b.pdf>
14. Ierache, J.S., Pereira, G., Nervo, F., Iribarren, J.: Emotional status focused on stimuli by applying brain-machine interface. In: Feierherd, G.E., Pesado, P., Sposito, O.M. (eds.) *Computer Science & Technology Series. XX Argentine Congress of Computer Science, Selected papers*, pp. 245–256. Editorial de la Universidad Nacional de La Plata (EDULP) (2015)
15. Ierache, J.S., Pereira, G., Sattolo, I., Iribarren, J.: Aplicación de interfases lectoras de biose textasciitilde nales en el contexto de la domótica. In: *XV Workshop de Investigadores en Ciencias de la Computación* (2013), <http://lsia.fi.uba.ar/papers/ierache13.pdf>
16. Jasper, H.H.: Report of the committee on methods of clinical examination in electroencephalography. *Electroencephalography and Clinical Neurophysiology* 10(2), 370–375 (2017/07/18 1958), [http://dx.doi.org/10.1016/0013-4694\(58\)90053-1](http://dx.doi.org/10.1016/0013-4694(58)90053-1)

17. Lang, P.J.: Behavioral treatment and bio-behavioral assessment: computer applications, pp. 119–137. Ablex, Norwood, NJ (1980)
18. Lang, P.J., Bradley, M.M., Cuthbert, B.N.: International affective picture system (iaps): Affective ratings of pictures and instruction manual. Technical report A-8 (2008)
19. Lang, P.J., Greenwald, M.K., Bradley, M.M., Hamm, A.O.: Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology* 30(3), 261–273 (1993)
20. McNeil, D.W., Brunetti, D.G.: Pain and fear: A bioinformational perspective on responsivity to imagery. *Behaviour research and therapy* 30(5), 513–520 (Sep 1992)
21. Mehrabian, A.: Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology* 14(4), 261–292 (1996)
22. Milborrow, S., Morkel, J., Nicolls, F.: The muct landmarked face database. Pattern Recognition Association of South Africa (0) (2010), <http://www.dip.ee.uct.ac.za/textasciitildenicolls/publish/sm10-prasa.pdf>, <http://www.milbo.org/muct>
23. Miller, G.A., Levin, D.N., Kozak, M.J., Cook, III, E.W., McLean Jr, A., Lang, P.J.: Individual differences in imagery and the psychophysiology of emotion. *Cognition and Emotion* 1(4), 367–390 (1987)
24. Morris, J.D., Bradley, M.M., Waine, C.A., Lang, J.B.: Assessing affective reactions to advertisements with the self-assessment manikin (sam). In: Southern Marketing Association Conference (1992)
25. Oude Bos, D.: Eeg-based emotion recognition-the influence of visual and auditory stimuli. *Capita Selecta (MSc course)* (2006)
26. Øygaard, A.M.: Emotion detection example. *clmtrackr*. (2015), http://auduno.github.io/clmtrackr/examples/clm_emotiondetection.html, accessed: 2017-05-23
27. Patrick, C.J., Bradley, M.M., Lang, P.J.: Emotion in the criminal psychopath: Startle reflex modulation. *Journal of Abnormal Psychology* 102(1), 82–92 (1993), <http://dx.doi.org/10.1037/0021-843X.102.1.82>
28. Picard, R.W.: *Affective Computing* (MIT Press). The MIT Press (2000)
29. Rossi, F.M., Calot, E.P., Ierache, J.S.: Educación emocional de un individuo en contextos multimodales en computación afectiva. In: XIX Workshop de Investigadores en Ciencias de la Computación (WICC 2017, Buenos Aires, Argentina) (2017)
30. Russell, J.A.: A circumplex model of affect. *Journal of Personality and Social Psychology* 39(6), 1161–1178 (1980), <http://dx.doi.org/10.1037/h0077714>
31. Saragih, J.M., Lucey, S., Cohn, J.F.: Deformable model fitting by regularized landmark mean-shift. *International Journal of Computer Vision* 91(2), 200–215 (Jan 2011), <http://dx.doi.org/10.1007/s11263-010-0380-4>
32. Scherer, K.R., Schorr, A., Johnstone, T.: *Appraisal Processes in Emotion: Theory, Methods, Research*. Series in Affective Science, Oxford University Press (2001), <https://books.google.com.ar/books?id=IWLnBwAAQBAJ>
33. Thomas, K., Benjamin, R.K., Fernando, P., Brian, G., Matthias, B., Jonathan, F., Kyle, K., Jessica, H., Jason, G., Sylvain, C., et al.: Jupyter notebooks – a publishing format for reproducible computational workflows. *Stand Alone 0(Positioning and Power in Academic Publishing: Players, Agents and Agendas)*, 87–90 (2016), <http://doi.org/10.3233/978-1-61499-649-1-87>