Bio-sensing for Emotional Characterization without Word Labels

Tessa Verhoef[1], Christine Lisetti[2], Armando Barreto[3], Francisco Ortega[2],
Tijn van der Zant[4], and Fokie Cnossen[4]

[1] University of Amsterdam
1012 VT Amsterdam, The Netherlands
`t.verhoef@uva.nl`
[2] School of Computing and Information Sciences
`lisetti@cis.fiu.edu, franciscoraulortega@gmail.com`
[3] Department of Bio-medical Engineering
Florida International University,
Miami, 33199 USA
`barretoa@fiu.edu`
[4] Artificial Intelligence Institute
University of Groningen
9747 AG Groningen, The Netherlands
`{C.M.van.der.Zant,f.cnossen}@ai.rug.nl`

Abstract. In this article, we address some of the issues concerning emotion recognition from processing physiological signals captured by bio-sensors. We discuss some of our preliminary results, and propose future directions for emotion recognition based on our lessons learned.

Keywords: Emotion Recognition, Affective Computing, Bio-sensing.

1 Introduction

In the past few years, a number of psychologists [1-3] have challenged the classical notion that emotion can be categorized by labels, using words such as 'anger', 'fear', 'happiness' [4], and have proposed to use dimensional representations of emotions for more realistic categorization of emotional states. Emotion labeling has also been found to be dangerously ethnocentric and misleading [5].

   Our current reported work is one of the first attempts to approach automatic emotion recognition with the novel method proposed by Peter and Herbon [6] that moves away from the notion of labeling emotional states with discrete categorical words. In the following section we describe which physiological modalities associated with emotions we chose to capture, the bio-sensors that we used, and the emotion elicitation method used with participants. The section after that explains how we created a data set suitable for training and testing emotion classifiers, and the processing of these bio-physiological signals for classification. Finally we discuss some of the lessons learnt from this experiment and propose future directions toward emotion recognition.

2   Emotion Recognition without Labeling Emotion with Words

Peter and Herbon [6] have proposed a method for avoiding the use of words for automatic emotion recognition and provided guidelines about how to structure emotions as a dimensional representation for the use in human-machine interaction. Labeling emotions can be problematic because the category borders are blurry and the word 'anger' for instance can describe many different emotional states. The method as described by [6] avoids these problems because it abandons labeling emotions with words.

The procedure to classify emotions for automatic emotion recognition proposed in [6] consists of the following four steps:

- Step 1: Elicit emotions while measuring physiological signals and ask test subjects to self-report in a way that can be translated into a dimensional structure.
- Step 2: Assign the physiological measurements to the related ratings.
- Step 3: Group emotions into clusters with similar physiology and place in dimensional structure.
- Step 4: Identify characteristic patterns in physiology for each cluster.

In this section we describe which sensor modalities we chose to capture and process the physiological signals that are associated with emotional states. We describe the emotion elicitation method based on psychological findings designed to collect data while eliciting emotions from the participants. With these experiments we completed step one and two of the procedure described by [6].

2.1   Bio-sensors Used to Collect Data

Our data set consisted of multimodal physiological evidence about the affective state of a user: galvanic skin response (GSR), and blood volume pressure (BVP). For a survey of the different modalities associated with emotional states and the recognition methods used to process these various modalities to date see [7].

Galvanic Skin Response (GSR): The GSR2 Thought Tech LTD device[1] shown in Figure 1.a was used to measure the Galvanic Skin Response (GSR). This method was introduced in the early 20th century and is based on the idea that conductance of an electric current is easier on moist skin. The autonomic nervous system, which consists of two subsystems – the parasympathetic and the sympathetic subsystems – has an influence on the control of sweat glands. In the case of higher sympathetic activity, the sweat glands get more hydrated and skin conductance increases. So, the sweat glands are used as resistors and the skin conductance can be measured with the GSR device by passing a small electric current across two electrodes that touch the skin.

Blood Volume Pressure (BVP): The Pulse Plethysmograph[2] shown in Figure 1.b. was used for measuring the Blood Volume Pulse (BVP), a signal from which information about the Heart Rate Variability (HRV) can be computed. HRV has been

---

[1] http://www.thoughttechnology.com/gsr.htm
[2] http://www.ufiservingscience.com/Pig1.html

linked to emotional processes and the autonomic nervous system [8]. In addition, information about vasoconstriction (constriction of the blood vessels) can be inferred by detecting a decrease in the amplitude of the BVP signal. Vasoconstriction is said to be related to emotional processing as well [9]. The sensing device shown in Figure 1.b. is a finger clip that uses an infrared emitter and receiver to measure the amount of light that is reflected back by the skin.

(a)   (b)

Fig. 1. (a).The GSR2 skin conductance device (Thought Tech LTD) (b). Pulse Plethysmograph.

2.2   Emotion Elicitation

We created an experimental set-up with the sensors described above in which test subjects were exposed to emotion eliciting stimuli, and data was captured from the sensing devices during that exposure. During the experiment, the physiological signals were measured with the non-invasive sensors while the participant was asked to keep the arm that was attached to the sensors as motionless as possible (to avoid generating noise data associated with body movements).

Stimulus design: The stimuli used for emotion elicitation consisted of movie fragments that were known to elicit a range of different emotions. Gross and Levenson [10] conducted a very thorough study to provide a selection of movie fragments best suited to elicit certain emotions. Using a large amount of test subjects and a wide selection of movie fragments they were eventually able to reduce it to a reliable set in terms of emotion discreteness and intensity. More recently, Nasoz et al. [11] performed another panel study in which they tested the movie selection again, and created a modified version of the movie selection which proved more appropriate. In order to allow for easy comparison of previous emotion recognition classification, our set of emotion eliciting movie clips was based on the selection of [11]. Some changes have been made though because during a pre-testing stage it appeared that people responded inappropriately to some of the movies. For example, The Shining, originally meant to elicit fear, is so well known and by now so old, that people often show a smile of recognition instead of fear. Similarly, Drop Dead Fred caused people to be annoyed more than amused. To find out whether these two movies should be

replaced, a small pilot study was conducted in which we showed the two movie clips as well as two alternatives (The Ring for fear and the Pixar short movie Boundin' for happiness) and asked people to rate the emotion that they experienced while watching the clip. They were asked to choose one of the following possibilities [Happy, Angry, Sad, Disgusted, Surprised, Afraid, Neutral or None of the above] and to rate the intensity of the felt emotion on a scale from 1 to 5. Fifteen test subjects participated in this pilot study among which 7 were female and 8 were male. Their ages varied from 22 to 57. The results are shown in Table 1. The difference in eliciting success (defined by the percentage of subjects that reported to have felt the intended emotion) and average reported intensity between the two movies for 'happy' is smaller than the difference for the 'fear' movies, but in both cases the alternative movie scores better, therefore we decided to replace them both. The final selection of movies for the main experiment therefore was: The Champ for sadness, Schindler's List for anger, The Ring for fear, Capricorn One for surprise, Bounding' for happiness and an episode of Fear Factor for disgust.

Table 1. Results of the pilot study about eliciting abilities of movie clips. Average eliciting success, average intensity rating and Standard Deviation of the intensity rating.

Movie clips `happy'	Eliciting success	Average intensity	SD
Drop Dead Fred	67 %	2.8	0.79
Boundin'	73 %	3.1	1.39
Movie clips `fear'	Eliciting success	Average intensity	SD
The Shining	87 %	2.6	1.14
The Ring	100 %	3.9	1.16

Procedure: During the main experiment, the user watched the six selected movie fragments which were separated with a reasonably long pause to make sure that the subject would be relaxed and in a neutral state again before the next movie started. Before the movies started there was such a pause as well, in which relaxing music was played and the participant was asked to breathe slowly and try to relax. After each fragment, the user was asked to self-report about which emotion he or she felt during the movie via a questionnaire. The written questionnaire was designed by adapting the concept of the Emotion Wheel [3] shown in Figure 2, which has been proposed to be an intuitive tool [3] for participants to (a) identify which emotions they are experiencing using labels from 16 different emotion families such as anger, fear, happiness (as most people are used to do when reporting on their emotional states) and (b) grade the intensity of their emotion within that family (e.g. the anger label can be fine-tuned to refer to rage or to annoyance, by respectively rating up or down the intensity of the experience).

There are ongoing discussions among psychologists about what is the minimum number of dimensions sufficient to differentiate the variety of emotions that humans can experience and about what these dimensions should be [1], [2], [3]. Although the Emotion Wheel proposes 16 dimensions as "spikes" around the wheel, it also incorporates a mapping from these labels to a continuous 2-dimensional structure of emotional states: the valence dimension indicates whether an emotional state is

associated with a positive or a negative experience, and the power dimension indicates the coping potential or how well the emotional situation can be handled. In addition, the Emotion Wheel allows for self-report of the intensity or how strongly the emotional state is experienced, with increasingly intense emotions radiating from the center of the wheel.
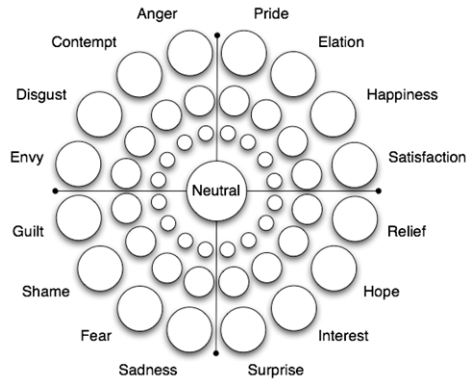

Fig. 2. The Emotion Wheel and Our Associated Questionnaire Notes, both adapted from [3]

Explanation: 16 different emotion families are arranged in a circular fashion. Please note that the word or label that represents each family can stand for a whole range of similar emotions. Thus, the Anger family also covers emotions such as rage, vexation, annoyance, indignation, fury, exasperation, or being cross or mad.

First identify approximately how you felt during the movie and choose the emotion family that best corresponds to the kind of feeling you experienced. Then determine with which intensity you experienced the respective emotion and check one of the circles in the "spike" corresponding to this emotion family -- the bigger the circle and the closer it is to the rim of the wheel, the stronger the emotional experience. If you felt no emotion at all, check the 'neutral' circle in the center.

The duration of the complete procedure was approximately 45 minutes. 25 test subjects participated in the experiment who varied in age from 21 to 41. The group consisted of 16 males and 9 females and the division of their ethnicities was as follows: 40 % Caucasian, 40 % Latin American, and 20 % Asian.

Some of the data had to be excluded from the data set. Whenever a participant self reported an emotion that did not match the intended emotion for a movie fragment, the data for that movie fragment was not used in the data set. Another reason to exclude data was unsuccessful recording of the signals. Sometimes participants moved the arm with the sensors too much which caused interruptions in the signals and made it impossible to compute the features in the signal. These recordings were therefore also excluded.

3   Physiological Data Classification toward Emotion Recognition

3.1   Feature Extraction

For each test subject the experiment resulted in two signals: the raw GSR signal and the raw BVP signal. To create the data set, we computed features from the recorded signals. The features we computed were the same as the ones that are assessed by Barreto et al. [12] with the only difference that we did not consider each movie as a complete segment from which each feature is computed over the whole segment, but we assessed the signals in intervals of 40 seconds, in order to have a sequence of feature values for each elicited emotion.

A typical GSR response consists of several temporary increases: the skin conductance responses (SCRs). Figure 3 shows an example of such a GSR response. An electrodermal response is often described with a few specific characteristics from these responses: amplitude, rise time and the half-recovery time [12]. The specific features that we computed from the GSR signal are: the number of GSR responses, Mean value of the GSR, average Amplitude of the GSR responses, average Rising time of the GSR responses and the average Energy of the responses (the total area under the half-recovery time). All these features were computed as described in [12].
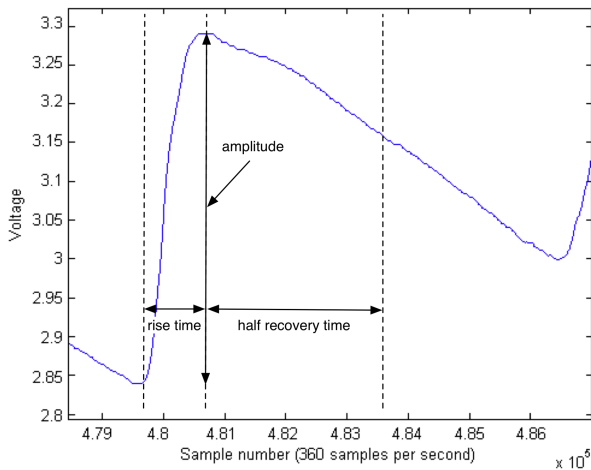

Fig. 3. Graphical annotated rendering of a Galvanic Skin Response

As mentioned earlier, the BVP signal can be used to compute features such as the Heart Rate Variability (HRV). A typical BVP beat is shown in Figure 4. As described in [12], the individual heart beats are first separated by finding the Inter Beat Intervals (or period) which is the time between two peaks in the signal. This series is usually analyzed by studying different frequency bands in which the Low Frequency (LF) (0.05-0.15Hz) band reflects sympathetic activity whereas the High Frequency (HF)

(0.16-0.40Hz) band reflects parasympathetic activity. The LF/HF ratio is computed as one feature, as well as the mean Inter Beat Interval, the standard deviation of the Inter Beat Interval and the mean amplitude of the individual beats that are detected in the segment.

Some problems were encountered with one of the six emotion categories because of the nature of the elicitation method. The duration of the surprise part in the movie clip that elicited surprise only lasted a few seconds. The way we compute the features from the signals requires the assessment of segments of at least 40 seconds because, for instance, a single GSR response can last that long. This made it unfeasible to use the signals that we recorded for surprise in the way the data is processed, and we decided to use only the data for the other 5 emotions and built a classifier for this 5-class problem.
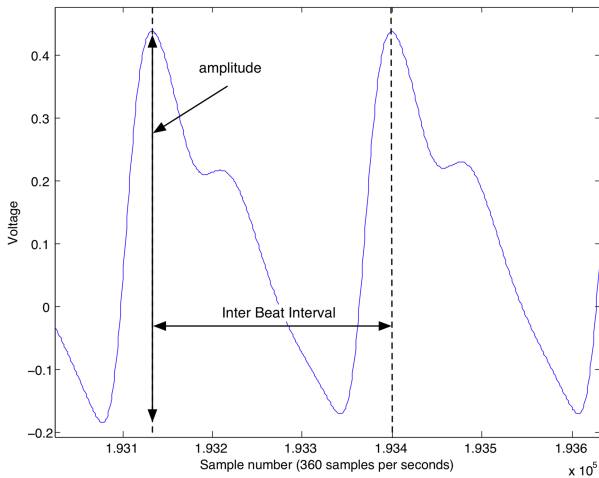

Fig. 4. Graphical annotated rendering of a Blood Volume Pressure beat

3.2 Normalization

The physiological response to emotion eliciting stimuli differs a lot from person to person. Therefore the data from the two biosensors was normalized so that it reflects the proportional difference in reaction for the different stimuli segments and also to re-scale the individual baselines. These normalization steps follow the example of those used by [12] with similar data.

The first normalization step uses a set of features that were computed during the relaxation interval that preceded the first movie clip of the experiment. This interval represents the baseline of the user's reaction. If X_e is one of the features for the segment eliciting emotion e, and X_r is the feature recorded during the relaxation interval, then Equation 1 computes the feature value of X_e after the first normalization step by dividing the original value by the one in relaxation.

$$X'_e = \frac{X_e}{X_r} \tag{1}$$

The second normalization step was also aimed at reducing the influence of individual differences and to make sure that the baselines and strength of the responses of the individuals are equalized. If X'_e is one feature value after the first normalization step, for the segment that elicited emotion e, the second normalization step follows Equation 2. The value is divided by the average individual response for that feature computed over all feature vectors n for all segments corresponding to the six emotions.

$$X''_e = \frac{X'_e}{\frac{1}{n}\sum_{i=1}^{n} X'_{ei}} \tag{2}$$

The last step normalized the features to a uniform range in order to eliminate differences in dynamic range and the chance that this could make some features dominate others. This min-max normalization step follows Equation 3 and maps all computed feature values to a value in the range from zero to one.

$$X_{norm} = \frac{X''_e - X''_{e\min}}{X''_{e\max} - X''_{e\min}} \tag{3}$$

3.3   Classification

Corresponding with step 3 in the process as described by Peter and Herbon [6], we tried to group the emotions by implementing the K-means clustering algorithm [13], which can find the centers of clusters that are naturally present in data. Then with this new class labeling (not associated with emotion words), we moved on to step 4 to indentify patterns in the physiological signals. We used a Static Bayesian Network approach to train and test a classifier with attribute selection. The performance that could be reached so far with this method was 60.0 % classified correctly.

4   Lessons Learned and Future Work

The investigation described in the previous sections of this paper was exploratory in nature and has been successful in providing direction for our future work. For example, this work reinforced our belief that attempts at describing emotional states through discrete emotion labels may be self-limiting. This view is also supported by the results from experiments by Barrett and Russell, which showed that subjects may experience several emotions at the same time [14]. Accordingly, we will approach future experiments using a methodology similar to that suggested by Peter and Herbon [6]. Thus, to specify emotional states we intend to use a multi-dimensional representation, such as, for example, the three dimensional space proposed by Russell and Mehrabian [2], in combination with personality data about the subject. The latter will be important to understand the differences between the subjects.

The work described here also highlighted the critical nature of the emotion elicitation component of the experimental protocol. In the future, we may still use elicitation techniques similar to those described in this paper. However, we will extend the length of time for elicitation to 60 seconds per movie clip. In addition, we will experiment with more realistic life experiences as suggested by Peter and Herbon [6]. Such type of elicitation may prove to be better suited for emotional state recognition.

To analyze the data resulting from the experiments we may explore other forms of cluster analysis (e.g., Kohonen Learning), which may provide a clearer picture when viewing an n-dimensional emotional state space. For classification techniques we will revisit methods used in this paper with additional techniques if needed.

Some specific recommendations for future work follow:

- More modalities: It might have been the case that the combination of only the two sensors that we used for recording the physiological signals did not contain enough discriminating ability to distinguish between five or more emotions. The combination of physiological signals with facial expression data and other modalities such as vocal intonation is likely to enhance accuracy.
- More participants: In the present research only 25 test subjects participated in the data collection. A lot of this data had to be excluded which resulted in a data set that was quite small. In a follow-up experiment, more participants should be invited so that the problem of not having enough data can be eliminated.
- Other elicitation methods: the way that we elicited the emotions in the experiment was one of the causes for the unfortunate composition of our data set. The stimuli should be of a longer duration so that also dynamic methods can seriously be applied and they should have the same duration for all emotions in order to be able to create a balanced data set, which would contain the same amount of examples for each category.

References

1. Russell, J.A.: A circumplex model of affect. Journal of Personality and Social Psychology 39(6), 1161–1178 (1980)
2. Russell, J.A., Mehrabian, A.: Evidence for a three-factor theory of emotions. Journal of Research in Personality 11(3), 273–294 (1977)
3. Scherer, K.R.: What are emotions? and how can they be measured? Social Science Information 44(4), 695–729 (2005)
4. Ekman, P.: An argument for basic emotions. Emotion: Themes in the Philosophy of the Mind (1992)
5. Wierzbicka, A.: Defining emotion concepts. Cognitive Science 16(4), 539–581 (1992)
6. Peter, C., Herbon, A.: Emotion representation and physiology assignments in digital systems. Interacting with Computers 18(2), 139–170 (2006)

7. Lisetti, C.L., Nasoz, F.: Using noninvasive wearable computers to recognize human emotions from physiological signals. EURASIP Journal on Applied Signal Processing 2004(11), 1672–1687 (2004)
8. Dishman, R.K., Nakamura, Y., Garcia, M.E., Thompson, R.W., Dunn, A.L., Blair, S.N.: Heart rate variability, trait anxiety, and perceived stress among physically fit men and women. International Journal of Psychophysiology 37(2), 121–133 (2000)
9. Hilton, S.M.: The defence-arousal system and its relevance for circulatory and respiratory control. J. Exp. Biol. 100(1), 159–174 (1982)
10. Gross, J.J., Levenson, R.W.: Emotion elicitation using films. Cognition & Emotion 9(1), 87–108 (1995)
11. Nasoz, F., Alvarez, K., Lisetti, C.L., Finkelstein, N.: Emotion recognition from physiological signals using wireless sensors for presence technologies. Cognition, Technology & Work 6(1), 4–14 (2004)
12. Barreto, A., Zhai, J., Adjouadi, M.: Non-intrusive Physiological Monitoring for Automated Stress Detection in Human-Computer Interaction. In: Lew, M., Sebe, N., Huang, T.S., Bakker, E.M. (eds.) HCI 2007. LNCS, vol. 4796, pp. 29–38. Springer, Heidelberg (2007)
13. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern classification and scene analysis. Wiley, New York (1973)
14. Barrett, F.L., Russell, J.A.: Independence and bipolarity in the structure of affect. Journal of Personality and Social Psychology 74(4), 967–984 (1998)