Emotion Recognition Using Bio-sensors: First Steps towards an Automatic System

Andreas Haag, Silke Goronzy, Peter Schaich, and Jason Williams

Sony Corporate Laboratories Europe, Hedelfinger Str. 61
D-70327 Stuttgart, Germany
{Haag,Goronzy,Schaich,Williams}@Sony.de

Abstract. The detection of emotion is becoming an increasingly important field for human-computer interaction as the advantages emotion recognition offer become more apparent and realisable. Emotion recognition can be achieved by a number of methods, one of which is through the use of bio-sensors. Bio-sensors possess a number of advantages against other emotion recognition methods as they can be made both inobtrusive and robust against a number of environmental conditions which other forms of emotion recognition have difficulty to overcome. In this paper, we describe a procedure to train computers to recognise emotions using multiple signals from many different bio-sensors. In particular, we describe the procedure we adopted to elicit emotions and to train our system to recognise them. We also present a set of preliminary results which indicate that our neural net classifier is able to obtain accuracy rates of 96.6% and 89.9% for recognition of emotion arousal and valence respectively.

1 Introduction

The use of emotion in computers is a field which is becoming increasingly en vogue at the current time. In many ways emotions are one of the last and least explored frontiers of intuitive human-computer interaction. This can perhaps be explained by the fact that computers are traditionally viewed as logical and rational tools, something which is incompatible with the often irrational and seeming illogical nature of emotions [6]. It is also apparent that we as humans, while extremely good at feeling and expressing emotions, still cannot agree on how they should best be defined [20]. These reasons are then topped by the fact that emotion recognition is itself a technically challenging field.

Yet there are a number of good reasons why emotion is a fitting topic for Human-Computer Interaction research. Humans being emotional creatures should theoretically be able to interact more effectively with computers which can account for these emotions. This can take place in at least three ways:

First of all, computers which understand the emotions of their users would have the potential to take appropriate courses of action, which otherwise would not be realised. For example, reducing music volume in times of extreme stress, or suggesting

E. André et al. (Eds.): ADS 2004, LNAI 3068, pp. 36–48, 2004.

suitable films according to the mood of the viewer. The computer could also learn to recognise the particular likes and dislikes of its user and anticipate appropriate actions in an unobtrusive manner [8]. Understanding that its user likes soothing music on returning from work, the computer could automatically ensure that such music is playing as the user enters the house. In fact, the computer might even be more informed in this respect than the user who may plausibly not appreciate this fact themselves. As such emotion-recognising computers would become more like personal companions than the mere tools that they are today.

Secondly, it has been demonstrated that interaction between human and computer is largely social [7]. As emotions are a key component of human social interaction, there is a need for computers to understand this component in order to be perceived by users as truly effective in communication. This is under-pinned by the fact that as Nass points out, the various components of communication are generally combined together by humans. Nass also points out that the expectations of users with regard to the communicative abilities of computers is set to increase as interfaces appear more sophisticated. A highly realistic avatar with stunningly realistic speech synthesis will therefore be expected to have similarly sophisticated powers of speech recognition and understanding. As emotion is an essential aspect of human-human communication it will inevitably be expected that such computers will be able to recognise the content of spoken communication in its entirety, inclusive of the emotional component.

Thirdly, emotions can also be used as an additional interaction modality to improve communication effectiveness in the same way as the additional use of speech or gestures. Just as the use of gesture can help dis-ambiguate the meaning of a particular utterance (e.g. “put that there”), so too can emotion be used to disambiguate the meaning of verbal utterances such as for example when sarcasm is used (e.g. “Great, this is all I need!”).

Being inherently multi-modal, there are a number of ways in which emotions can be recognised. This may be via speech, facial expression, gesture and / or a variety of other physical and physiological cues. This spread of modalities across which emotion is expressed leaves the field open for many different potential recognition methods. Two of the more researched methods include speech [12,14] and face recognition [15]. In this paper we focus on an alternative and lesser-used method, namely the use of bio-sensors. Although some research has been conducted in this field e.g. by Picard, [6] Healy [9] or Christie [5], the areas of speech and face recognition are far more explored.

The use of bio-sensors as a means of recognising emotion has a number of advantages. The size of such sensors is rapidly decreasing to the extent that it should soon be a routine matter to incorporate them into jewellery (e.g. ring or bracelet) or clothing. Indeed this is already possible in many cases. [19] This method of measuring bio-signals and thus emotions might be anticipated as being less disturbing than being "watched" by a camera as is the case with facial expression recognition. Facial expression recognition is also difficult when light levels are low and when the

user is moving. Additional difficulties exist when recognising emotion from speech, for example in applications where users are listening to or watching music or movies. In such cases it cannot be expected that they would talk while watching, and even if this was the case, severe problems currently exist in recognising and isolating speech from ambient background noise [17].

There are therefore strong arguments for developing systems which can recognise emotions from bio-signals, and which can ultimately use a combination of methods (including speech and expression recognition) to capitalise on the advantages of each method. In this paper we will focus on the issues surrounding the recognition of emotion from bio-sensors and then describe a methodology we have adopted to both elicit and recognise emotions such that they can be recognised in future.

2 The Bio-signals

When we are frightened our heart races, our breathing becomes rapid, our mouth becomes dry, our muscles tense, our palms become sweaty, and we may want to run. These bodily changes are mediated by the autonomic nervous system, which controls heart muscle, smooth muscle, and exocrine glands [10]. The autonomic nervous system itself can be divided into sympathetic and parasympathetic divisions. Both operate in conjunction with each other and with the somatic motor system to regulate most types of behaviour, whether in normal or emergency situations. Although several visceral functions are controlled predominantly by one or the other division, and although both the sympathetic and parasympathetic divisions often exert opposing effects on innervated target tissues, it is the balance of activity between the two that helps maintain an internal stable environment in the face of changing external conditions, [2]. Certain emotions can affect this balance and can result in a wide variety of bodily reactions comparable to the ones described above.

These bodily reactions can be monitored and measured. These signals are then referred to as bio-signals. What we can observe from the outside are the bodily reactions only. Our goal is to use these reactions and by means of special bio-sensors, deduce the emotional state of the user.

We use the following set of bio-signals:

- **Electromyography** (EMG) refers to the muscle activity or frequency of muscle tension of a certain muscle. This signal was chosen because high muscle tension often occurs under stress. The absolute level of the muscle tension however strongly depends on the muscle where it is measured.
- **Electrodermal activity** -also referred to as skin conductivity (SC)- basically measures the conductivity of the skin, which increases if the skin is sweaty. This signal was found to be a good and sensitive indicator of stress as well as other stimuli and also helps to differentiate between conflict-no conflict situations or between anger and fear. The problem with this signal is that it is also influenced by external factors such as outside temperature. It therefore needs reference measurements and calibration.

- **Skin temperature** simply describes the temperature as measured on the surface of the skin. Since muscles are tense under strain, the blood vessels will be contracted and therefore the temperature will decrease. Comparable to the SC, the skin temperature also depends on external factors. Furthermore it is a relatively slow indicator of changes in emotional state.
- **Blood volume pulse** (BVP) is a measure to determine the amount of blood currently running though the vessels, e.g. in the finger of a test subject. A photoplethysmograph (PPG) which consists of a light source and photo sensor are attached to the skin and the amount of reflected light, which depends on the amount of blood, is measured. BVP can be used to measure vasoconstriction and the heart rate.
- **Electrocardiogram** (ECG). The ECG signal is the manifestation of contractile activity of the heart. This can be recorded either directly on the surface of the chest or alternatively on the limbs. Recording ECG from the limbs is less inconvenient but more vulnerable to artifacts. For the initial system we decided to use the ECG from the chest to prevent the system being influenced by artifacts. It can be used to measure heart rate (HR) and inter-beat intervals (IBI) to determine the heart rate variability (HRV). A low HRV can indicate a state of relaxation, whereas an increased HRV can indicate a potential state of mental stress or frustration.
- **Respiration** sensors measure how deep and fast a person is breathing. This is measured by applying a rubber band around the chest. Fast and deep breathing can indicate excitement such as anger or fear but sometimes also joy. Rapid shallow breathing can indicate tense anticipation including panic, fear or concentration. Slow and deep breathing indicates a relaxed resting state while slow and shallow breathing can indicate states of withdrawal, passive like depression or calm happiness.

We used a combination of these signals, to derive a set of features that can be used to train a neural network classifier. This is then used to automatically detect the emotional state of a user in terms of arousal and valence values, based on the bio-signals recorded.

3 The Bio-sensors

For measuring the different body functions described in the previous section, we used the sensor set ProComp+ [1]. Their application is depicted in the following Figures.

Figure [1] shows the sensors for measuring EMG. We chose to measure the muscle activity using the masseter muscle, since the muscle movement has been described as reliable in that location [9]. Figure [2] shows a standard ECG sensor. In Figure [3] the respiration sensor is shown applied to the chest while in Figure [4] the skin conductivity, BVP and temperature sensor can be seen applied to the fingers of the left hand (the non-dominant hand should be used for measuring).

Fig. 1. EMG Sensor applied to the jaw

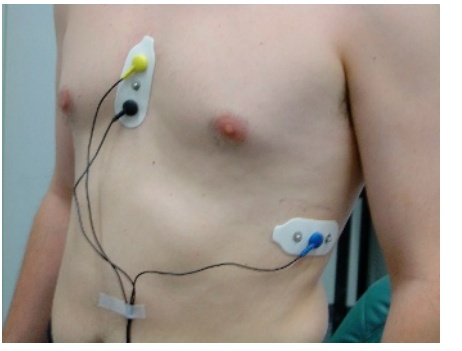

Fig. 2. ECG Sensor applied to chest

Fig. 3. Respiration sensor

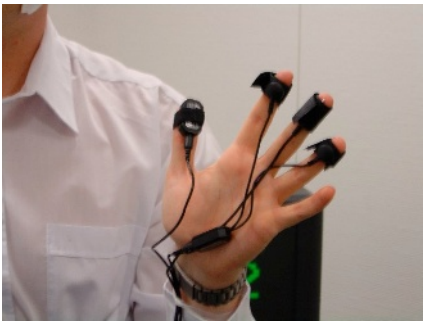

Fig. 4. Skin conductivity, temperature and BVP sensor

We are well aware that the current form and method of application of the bio-sensors is anything but intuitive and natural. However considering the current trend towards wearable computing, it can be expected that the bio-sensors will sooner or later be tiny enough to be inconspicuously incorporated into clothing, jewelery or the like. In this case the users, although aware that their bio-signals are being monitored, will not be physically aware of them. However, for research purposes we chose the rather bulky, cabled sensors, because they allow a certain flexibility e.g. in terms of placement of the sensors. This flexibility is important given the fact that many aspects of sensor usage are still not completely clear, e.g. which muscle is most appropriate for measuring the EMG signal.

Once however, it is clear, exactly where and how the bio-signal can be optimally measured, special dedicated sensors can be used. It might also be argued that measuring some signals such as the ECG is always disturbing no matter how tiny the sensors are because users would have to undress to apply them to their chest. An important component of our research however is to find redundancies in the set of bio-signals which may allow for less complicated sensor arrangements to be developed. It may therefore for example be conceivable to derive information about the HR from the BVP instead of the ECG and its more inconvenient sensor arrangement.

4 Data Collection Experiments

Having established a set of signals which may be of use for recognising emotion, it is then necessary to define a methodology in order to enable the system to translate the signals coming from these sensors into specific emotions. A number of issues arise in such an endeavour. Of chief importance is the fact that unlike emotion recognition from speech or facial expression [13], comparatively little is known about which bio-signals accompany which emotions. Before being able to train the system to recognise certain signals therefore, it is first necessary to elicit emotions and then record the various bio-signals that accompany them.

An initial data collection process therefore had to be designed to allow this to be acheived. Designing such a process gives rise to two broad issues which need to be considered. The first concerns the question of how to actually elicit emotions for signal recording purposes. The second issue concerns the definition of a procedure which ensures that emotions are elicited in as valid and reliable a way as possible.

Dealing first with the issue of how to elicit emotions: a number of options are possible each possessing different moral and practical aspects. Procedures with the most validity rely on creating a strong sense of realism in the subject. While this may largely be acceptable for more positive emotions, the elicitation of negative emotions would most likely need to rely on procedures such as causing pain or humiliation which are ethically questionable. In addition to the ethical aspect, such procedures are extremely time consuming and require a large amount of care and effort both in planning and execution. At the other extreme, procedures which require comparatively little planning such as asking subjects to act out certain emotions [9, 11] are likely to be insufficiently real for bio-sensors. The use of bio-sensors as lie detectors is testament to this.

It was therefore essential that the procedure to be adopted elicited real emotions, but at the same time minimised the ethical and practical dis-advantages of methods which create the most extreme sense of reality. An ideal solution as a starting point proved to be the IAPS photo set [2,3]. This takes the form of over 800 photographs which have been classified by a large number of participants in terms of arousal and valence. Arousal refers to how strong the picture content is. Valence (sometimes known as ‘pleasure’) refers to how positive or negative the content is considered to be. The IAPS photoset therefore possessed a number of advantages from the outset in that it was validated, and classified in terms of two simple but relevant dimensions. It therefore represented an appropriate first step for our classifier. The photoset was also amenable to a laboratory setting which was necessary in order to screen out any confounding variables such as physical exertion, visual or audio distractions which might influence the sensors.

Having decided on the elicitation materials, it was then necessary to decide upon a procedure which enabled the elicitation of emotions in such a way that they were neither biased by the original mood of the participant nor by emotions elicited early on in the experiment. To achieve this, we adopted a procedure whereby participants would rest for 15 minutes before beginning the experiment. Each participant was then shown a set of 10 low arousal and neutral valence photographs for 15 seconds each in order to try and normalise their mood. After that participants were shown a set of 5 low arousal positive and a set of 5 low arousal negative photos, followed by

medium arousal positive and negative arousal photos, followed by high arousal positive and negative photos. In between each set of 5 photos where either the arousal or valence changed, participants were shown a set of low arousal, neutral valence photos in order to normalise their mood once more.

We adopted a progressive increase in arousal level in order to minimise the disturbing effect which the content of any one photograph might have on subsequent photographs. We also ensured that the high arousal, negative valence photographs were shown last as pilot studies had shown that the content of the photographs (e.g. mutilations and body parts) was so disturbing that it was difficult for subjects to return to a normal resting state after seeing them. Participants also subsequently anticipated seeing a disturbing photo every time a new photograph was shown which affected the sensor readings. The procedure therefore represented a necessary compromise but proved to deliver useful data from which it was possible to derive a preliminary set of bio-signal features. The manner in which features were extracted from these signals is now reported on in the next section.

5 Feature Extraction

In order to derive features from the various bio-signals, we use a common set of feature values which are processed and used as an additional input or as a substitute to the raw signal for the classification. These common feature values are:

Running mean: The running mean computes a vector of mean values over time for a specific input signal by using a large rectangular window that is shifted across the feature vector. The size of the rectangular window depends on the input signal. This makes it possible to distinguish between phasic, fast changes and tonic, slow moving components in the analysed Signals.

Running standard deviation: This procedure calculates the local standard deviation of the signal and results in a vector that could be described as the activity and changes in the current signal.

Slope: The slope is just a simple approximation of the first derivative and therefore indicates fast changes in the sensory data. Filtering is often necessary before using this procedure as input signal noise may disturb the results.

Having described these common feature values, the precise manner in which each of the bio-signals are treated will now be described:

5.1 ECG

In order to process the ECG features we first subtract the global mean value of the raw signal that might be shifted up or downwards by the recording device. The signal

is filtered with a low pass filter with f_c=90Hz to remove high frequency noise, a very sharp high pass f_c=0.5Hz to eliminate low frequency drifts that occur when the subject is moving and then a notch filter at f_n=50Hz that removes influences of the power line. We detected QRS complexes by the following Algorithm (1). This is a derivative based method for detecting QRS complexes of the ECG signal [1] consisting of a weighted sum of the smoothed first derivative and the approximated second derivative.

$$y(n) = (\alpha \cdot |x(n) - x(n-2)|) + (\beta \cdot |x(n) - 2x(n-2) + x(n-4)|) \quad (1)$$

where x(n) is the input signal, α and β are the weight parameters for the balance between the smoothed three point first derivative and the approximated second derivative and y(n) is the output of this stage. α and β are set to 1.3 and 1.1 according to [1]. A further smoothing of the result is obtained by the introduction of an M-point moving average filter (2) .

$$y_{filt}(n) = \frac{1}{M} \sum_{j=0}^{M-1} y(n-j) \quad (2)$$

The resulting signal is scanned with an adaptive threshold that is defined by (3)

$$thres(n) = \frac{\max[n-k, n+k] - \min[n-k, n+k]}{\alpha} \quad (3)$$

The result is a list of samples each indicating the onset of a single QRS complex. This list is used to calculate a set of different features like heart rate (HR), heart rate variability (HRV) and the inter beat interval (IBI) between consecutive heart beats. The HRV for example is influenced by the sympathetic and parasympathetic vagus nerve and therefore a good indicator for the temporary dominace of one of them.

The HR is calculated simply by using the difference between two consecutive detected QRS complexes (t_{HB})

$$HR(n) = \frac{1}{60} \cdot \frac{t_{HB}(n) - t_{HB}(n-1)}{f_s} \quad (4)$$

However in order to make it more stable against artefacts and irregularities of the heart function we introduce a logic that calculates the mean duration of the previous IBIs and checks if the next heart beat occurs within a range of +/- 15 %.

If this constraint is violated, then we assume an out of range increase of the heart rate and correct the signal by introducing a beat at the expected position. This prevents missed QRS complexes or anomalies from influencing the HR. The HR itself is used as a feature as well as the previously mentioned common features derived from it.

5.2 BVP

This BVP waveform is potentially very useful in extracting a set of useful features for emotion recognition. It is possible to extract the HR by a procedure that is similar to the extraction of the HR from the ECG signal. This is done by subtracting the global mean, the differentiation and the following:

$$Y(n) = (X(n) + \min(X))^2 \tag{5}$$

$$thres(n) = \frac{\max[n-k, n+k] - \min[n-k, n+k]}{\alpha} \tag{6}$$

After that we also scan the signal if it exceeds the computed threshold.

The process which follows is identical to that used in the ECG feature processing. As you might guess the HR is the same as that derived by the ECG, however it is not always possible to use the ECG, on the other hand the BVP Sensor sometimes delivers too many artefacts. It might therefore be better to have a redundant or complimentary system.

Another value which can be calculated from the BVP is the vascular dilatation in the limbs. This Feature which is also known as the 'pinch' of the BVP is the complete amplitude of the raw signal. It is calculated by shifting a rectangular window over the signal. The window is analysed according to the maximum and minimum values. The difference between the maximum and the minimum value vector is the resulting pinch vector.

$$P(n) = \max[n-k, n+k] - \min[n-k, n+k] \tag{7}$$

The shape of the BVP signal is subject to change according to the workload of the heart [4]. For example the dichotic notch signal in a resting person can clearly be recognized and suppressed during exercise. We have also implemented an algorithm that extracts a section of 16 detected heart beats and normalises it according to the min/max values. We also computed a histogram with 100 bins and analysed the ratio between the sum of the interval [40,70] and [70,100].

5.3 Skin Conductivity

The skin conductivity consists of two separate components. There is a slow moving tonic component that indicates a general activity of the perspiratory glands from temperature or other influences and a faster phasic component that is influenced by emotions and the level of arousal. We process three features from this signal: Firstly, a raw file is normalised and filtered with a low pass filter with $f_c = 4Hz$ and the common features like slope mean and standard deviation. Two features are then taken from [4]

$$GSC_n = \frac{X - \min(X)}{\max(X) - \min(X)} \tag{8}$$

where a time varying local sample mean is obtained by windowing and subtracting the resulting mean from the original signal.

$$g_\mu(n) = g(n) - \frac{1}{N}\sum_{k=0}^{N-1} g(n-k) \tag{9}$$

The second measure is a time varying estimate of the local variance of the signal.

$$g_v(n) = \frac{1}{N-1}\sum_{k=0}^{N-1}\left(g(n-k) - \frac{1}{N}\sum_{l=0}^{N-1} g(n-l)\right) \tag{10}$$

5.4 EMG

The EMG signal is smoothed by a moving average filter with a length of 128 points and then processed using the common features in a similar manner as with the other signals already described.

5.5 Respiration

Respiration was recorded using a rubber belt around the chest. The processing was performed in the same way as with the basic features of the BVP. In this case however the amplitude ratio was left out.

5.6 Temperature

Temperature was recorded and the common features processed directly from the raw signal.

6 Classification Experiments

After having extracted the features as described in the previous section, we then trained a statistical classifier, with the goal of learning the corresponding emotion for a set of features with which it is presented. There are different options for building such a classifier. Fernandez [16] for example used HMMs, while Healy [9] used Fisher linear discriminant projection We chose to use a neural network that was trained and tested as described in the next section. Prior to training and testing all

features were normalised into the range between [0,1] before being converted into pattern files for the neural network simulator [18] :

$$X_n = \frac{X - \min(X)}{\max(X) - \min(X)} \tag{11}$$

6.1 Neural Network Classifier

The data for the experiment was gathered from a single subject on different days and different times of the day. Each pattern vector represents a frame of 2 seconds length. The target values that represent valence and arousal were extracted from the IAPS files [2,3], normalised to values between [0,1] and appended to the patterns that were recorded during the presentation of the corresponding picture. The dataset was then divided into three different sets for training, validation and testing. For the current experiments we extracted 1000 patterns with 700 patterns for training, 150 patterns for testing and 150 patterns for validation.

We trained two separate networks for valence and arousal each with an input vector of length 13, a hidden layer with 10 neurons and a single output node that gives an estimation of the corresponding valence or arousal value. For a first approach we decided to use a limited set of input features (features set) only to get a basic feeling for the features and their influence and importance to the quality of the estimator.

This feature set included the following features. Heart rate (std), BVP heart rate (std), BVP amplitude (std), EMG amplitude, skin conductivity (std), respiration amplitude (std), respiration rate (std). The logistic function was selected as the activation function for the hidden and output layer and the training was done with resilient propagation as the learning function.

6.2 Results

The independent test patterns were passed through the trained network and the results were analysed with the band function that counts a pattern as correct if the estimated and the target value lie within a range of a specified distance that we will refer to as bandwidth. This is shown in the table below:

Table 1. Classification results

	Arousal		Valence	
Bandwidth	0.1	0.2	0.1	0.2
Correct	89.73%	96.58%	63.76%	89.93%
Wrong	10.27%	3.42%	36.24%	10.07%

The results show that the estimation of the valence value is a much harder task than the estimation of arousal. If we only allow a bandwidth of 10% for the output to be counted as correct, we achieve 89.7% correct classification for arousal and 63.8% for

valence, indicating that valence is more difficult to detect from the signals. However, by increasing the allowed bandwidth to 20%, we can achieve 96.6% correct classification for arousal and 89.9% for valence, which is a highly acceptable result. This shows that it is possible to provide a good estimation of the mood which a participant is feeling (in terms of the likely section of the valence and arousal space) for a specific test person. Even though valence seems to be much harder to classify, 89.9% is still a good result and echoes the fact that emotion recognition from speech also works relatively well for arousal but not so well for valence [12].

7 Conclusion

This paper described a method of measuring emotions in humans which is still relatively unexplored, namely using bio-sensors. We adopted this approach because we hope that it could be a supplement for other emotion-related modalities as well as a good alternative in situations where they are not applicable. We first described how we designed experiments to elicit emotions before describing how we measured and processed them. Finally we described our neural net classifier and the recognition rates for arousal and valence that we achieved on data from a specific test subject.

There are clearly more steps to take in this research area. First of all, we aim to improve the accuracy of the network. This could be achieved either by adding some meaningful features (i.e. the full feature set) that are computed from the existing sensor signals or by using different bio-signals such as EEG data. There is also much scope to expand our system to incorporate other means of emotion recognition. Currently we are working on a facial expression system which can be integrated with bio-signal features. In addition we are working on a system to detect the eye-blinking frequency as this might be a good indicator of the amount of attention test subjects pay to their current tasks.

Our current system is set up to recognise arousal and valence values. While this goes part of the way towards achieving emotion recognition, it is still some distance from being able to recognise distinct emotions, especially given the fact that different emotions may have similar or identical valence or arousal values.

The emotion recognition results which we have described are also taken from one subject only. An important requirement of emotion recognising systems is that they do not require high levels of training for each user, as this would simply not be practical in the real world. It is essential therefore that subject-independent classifiers can be developed. This is a key goal of our research.

As a final word, we would emphasise that while the use of bio-signals is a valuable means of recognising emotion, it is currently questionable as to whether it is sufficient on its own. In our view the combination of different sources on top of bio-sensor signals, such as video analysis, motion detection or emotion recognition from speech is a necessary step to avoid the limitations of single modality systems.

References

1. Rangaraj M. Rangayyan; Biomedical Signal Analysis – A Case-Study Approach; IEEE Press 2002
2. Lang, P.J.,Bradley, M.M., & Cuthbert, B.N. (2001). International affective picture system (IAPS): Instruction Manual and Affective Ratings. Technical Report A-5. Gainesville, FL. The Center for Research in Psychophysiology, University of Florida.
3. Center for the Study of Emotion and Attention [CSEA-NIMH] (2001). The international affective picture system: Digitized photographs. Gainesville, FL: The Center for Research in Psychphysiology, Univerity of Florida
4. Barreto A., Heimer M., and Garcia M., "Characterization of Photoplehtysmographic Blood Volume Pulse Waveforms for Exercise Evalution," Proceedings 14th Southern Biomedical Engineering Conference, Shreveport, Louisiana, April, 1995, pp. 220-223
5. Chrisite, "Multivariate Discrimination of Emotion-specific Autonomic Nervous System Activity", Master Thesis in Science of Psychology, Virginia
6. Picard R.W., "Affective Computing", MIT Press, 2000
7. B. Reeves and C. Nass, *The Media Equation*, Center for the Study of Language and Information, Stanford University (1996).
8. Picard R.W.; Toward computers that recognize and respond to user emotion; IBM Systems Journal; Vol 39, Nos 3&4, 2000
9. Healy J. and Picard R., "Digital processing of Affective Signals", ICASSP 98
10. Principles of neural science, 4th Edition, E. R. Kandel, J. H. Schwartz, T. J. Jessell, McGraw-Hill, 2000
11. Vyzas E., Picard R., "Offline and Online Emotion Recognition from physiological data", EBA 1999
12. Tato, R., Santos, R., Kompe, R., Pardo,J.M., "Emotional Space Improves Emotion Recognition", ICSLP-2002, Denver, 2029-2032.
13. Ekman P. & Friesen W.E., "Facial Action Coding system" Consulting Pychologist Press, Palo Alto (California), 1977
14. Cowie R., "Describing the emotional states expressed in speech", ISCA workshop on speech and emotion, Belfast 2000
15. Turk M. and Pentland A., Eigen Faces for Recognition. Journal of Cognitive Neuroscience, 3(1), 1991.
16. Fernandez R., Picard R., Signal Processing for Recognition of Human Frustration; Media Laboratory, Perceptual Computing TR 447, MIT, Cambridge, MA, 1997
17. Juang B.H & Soong F.K., Hands-free Telecommunications; HSC 2001, pp.5-10; Kyoto, Japan
18. SNNS; Stuttgart Neural Network Simulator; http://www-ra.informatik.uni-tuebingen.de/SNNS
19. Pentland A.; Perceptual Intelligence; Communications of the ACM; Volume 43, Number 3 (2000), Pages 35-44
20. Richins, Marsha L. (1997), "Measuring Emotions in the Consumption Experience," Journal of Consumer Research, 24 (September), 127-146.