Published paper
Bamidis, Panagiotis D., Frantzidis, Christos A., Konstantinidis, Evdokimos I., Luneski, Andrej, Lithari, Chrysa, Klados, Manousos A., Bratsas, Charalambos, Papadelis , Christos and Pappas, Costas (2009) *An integrated approach to emotion recognition for advanced emotional intelligence.* In: Human-Computer Interaction. Ambient, Ubiquitous and Intelligent Interaction. 13th International Conference on Human-Computer Interaction, 19-24 July 2009 , San Diego, CA, USA. Lecture Notes in Computer Science (5612/2009). Springer , Berlin / Heidelberg, pp. 565-574.

http://dx.doi.org/10.1007/978-3-642-02580-8_62

An integrated approach to emotion recognition for advanced emotional intelligence

P.D. Bamidis[1], C. A. Frantzidis[1], E. I. Konstantinidis[1], A. Luneski[1], C. Lithari[1], M.A. Klados[1], C. Bratsas[1], C.L Papadelis[2], C. Pappas[1]

[1]Lab of Medical Informatics, Medical School, Aristotle University of Thessaloniki, Greece
[2]Center for Mind/Brain (CIMEC), University of Trento, Mattarello, Trentino, Italy
bamidis@med.auth.gr

Abstract. Emotion identification is beginning to be considered as an essential feature in human-computer interaction. However, most of the studies are mainly focused on facial expression classifications and speech recognition and not much attention has been paid until recently to physiological pattern recognition. In this paper, an integrative approach is proposed to emotional interaction by fusing multi-modal signals. Subjects are exposed to pictures selected from the International Affective Picture System (IAPS). A feature extraction procedure is used to discriminate between four affective states by means of a Mahalanobis distance classifier. The average classifications rate (74.11%) was encouraging. Thus, the induced affective state is mirrored through an avatar by changing its facial characteristics and generating a voice message sympathising with the user's mood. It is argued that multi-physiological patterning in combination with anthropomorphic avatars may contribute to the enhancement of affective multi-modal interfaces and the advancement of machine emotional intelligence..

Keywords: Emotion, Affective Computing, EEG, Skin Conductance, Avatar, Mahalanobis, classifier

1 Introduction

Emotions were until recently only rarely the discussion topic mentioned in human-computer interaction (HCI) studies of human intelligence. Lately, the arguments put alongside the significance of emotions gave birth to a new area of "emotional intelligence" within Ambient Intelligence (AmI).

Moreover, recent research on neuronal mechanisms involved in emotional processing provides some remarkable insights. There is now growing evidence that there exist two motivation systems - appetitive and aversive - activated according to the judgment of each situation as either pleasant or unpleasant. The reaction intensity is then modulated by the aforementioned systems and reflects the activation level [1]. The International Affective Picture System provides a set of normative emotionally evocative pictures that differ according to their arousal and valence dimensions and may be used for such experimental investigations [2].

Emotional intelligence has been demonstrated to be crucial during the performance of several cognitive functions [3]. Furthermore, the task of emotion recognition is very important during the interaction with other people. Humans communicate each other mainly due to their skill of emotional understanding. Theoretical research has demonstrated that the successful interaction of computers with humans will adopt basic principles required for the communication among human beings [4]. In order to achieve this goal it is essential to empower computers with emotion discrimination capabilities. Due to the fact that emotional processing modulates several aspects of human communication like voice and facial expressions, as well as, physiological reactions, emotional interactions could become more accurate, had they combined all or some of these features in concert.

Pattern recognition of emotional processing based on multi-physiological recordings has been growing recently as a field on its own within the HCI community [5]. Specific affective states, such as fear, melancholy, excitement, etc. have been demonstrated to show characteristic response patters regarding both the central and the autonomic nervous systems. This is further empowered by recent technological improvements enabling the use of wearable and miniaturized sensors [6], [7] which seem promising for a wide range of new health and medical applications by acquiring large sets of recorded data during daily realistic situations.

Previous studies have investigated the use of physiological pattern recognition during emotional processing. One of the first such works, was the one conducted by the MIT Media Lab in which a single subject intentionally expressed eight affective states over a period of more than a month [8]. During the experiment, features based on autonomic functions were extracted in order to discriminate between eight affective states by means of various techniques and classifiers. Anger was fully discriminated from the peaceful emotions (99%). Furthermore the eight emotions were separated into two classes according to their arousal dimension (80% for the high and 88% for the low arousal case), but the study faced difficulties trying to distinguish between pleasant and unpleasant emotions in a robust way (82% for the pleasant and 50% for the negative ones). A later work of the same team [9], improved the results achieving 81% recognition accuracy by seeding a Fisher Projection with the results obtained by means of the Sequential Floating Forward Search. The above classification rates regarded only a single subject. A more recent work [10] gathered physiological data from the autonomic nervous system from a single subject on different days and different times of the day. A great number of data segments (1000) was extracted and used for training a neural network classifier. Due to the large number of feature vectors (700 for training, 150 for testing and 150 for validation) it was feasible to robustly detect both arousal (96.58%) and valence (89.93%) dimensions of the emotions elicited by photos from the International Affective Picture System (IAPS) set [2]. Another study [11] used a long (45 min) show of slides and movie clips to elicit emotions in a fixed order. The emotion recognition task used non-invasive wearable sensors to gather data such as heart rate, skin temperature and phasic increases of the subject's electrodermal activity. Unlike previous studies, the sample included 31 participants. Recognition rates were 65.38% by means of k-Nearest Neighbors (kNN) and 69.28% when using an algorithm based on

Discriminant Function Analysis (DFA).

In previous work [12], we introduced a framework for the combination of multi-channel psycho-physiological recordings towards emotion aware computing. This aim of the current piece of work is to extend the emotion discrimination measured capacity obtained in a previous study through the use of a user-independent classifier [13], by presenting an integrated approach to emotion recognition through the fusion of signals obtained/derived from both the central (Electroencephalographic Response Potentials (ERPs), brain oscillatory activity) and the autonomic nervous system (skin conductance), and the utilization of machine learning techniques, such as the Mahalanobis classifier.

So, in the remaining of this paper, the experimental procedure, as well as, the description of the system architecture is provided in section II. The classification results achieved by the proposed approach are presented in section III and discussed in the last section.

2 Material & Methods

Healthy adult users (14 men and 14 females) were exposed to emotionally evocative-stimuli (pictures selected from IAPS) presented on a PC monitor. Each picture had a specific (L for Low, H for High) Valence-Arousal dimension (HVHA, LVLA, LVHA, HVLA). There were 40 repetitive trials from each one of the four affective space conditions (emotion categories). The sequence of the four conditions (or else blocks) was randomly selected for each subject. Each picture was presented for 1 second.

ERPs were recorded from nineteen sites with reference electrodes placed on the ear lobes at a sampling frequency of 500Hz. Skin conductance was recorded from medial phalanges of the non dominant hand. An off-line pre-processing step took place in order to remove artifacts from both signals. More specifically, EEG signals were filtered with a band-pass filter in the frequency range 0.5-40 Hz. Then, the Infomax Independent Component Analysis (ICA) technique was applied to remove artifacts caused by eye blinks. As for the electrodermal activity, the signal was digitally filtered by means of a low-pass short IIR with cut-off frequency at 2.5 Hz. Then, the data formed epochs time-locked to the stimulus onset. Finally, the average signal was computed for each stage.

The data from frontal (Fz), central (Cz) and parietal sites (Pz) distributed along the anterior-posterior midline of the brain, were analyzed and their main ERP components were extracted. Prominent peaks of the delta (0.5-4 Hz) and theta (4-8 Hz) oscillatory activity (Event Related Oscillations, EROs) from all electrode sites were used as features. The main features of a phasic skin conductance response (SCR) were also computed. These characteristics were the rise time, latency, amplitude and duration of the SCR response. The grand average signals from the aforementioned recordings are depicted in Figure 1.

Statistical analysis (Repeated Measures of ANOVA) with valence and arousal as within subject's factors and gender as between subject's factor was performed on the average values of the ERPs, EROs and on the SCR characteristics in order to estimate the discrimination capacity of each feature. Then, feature selection took place according to the p-values obtained from the selected features.

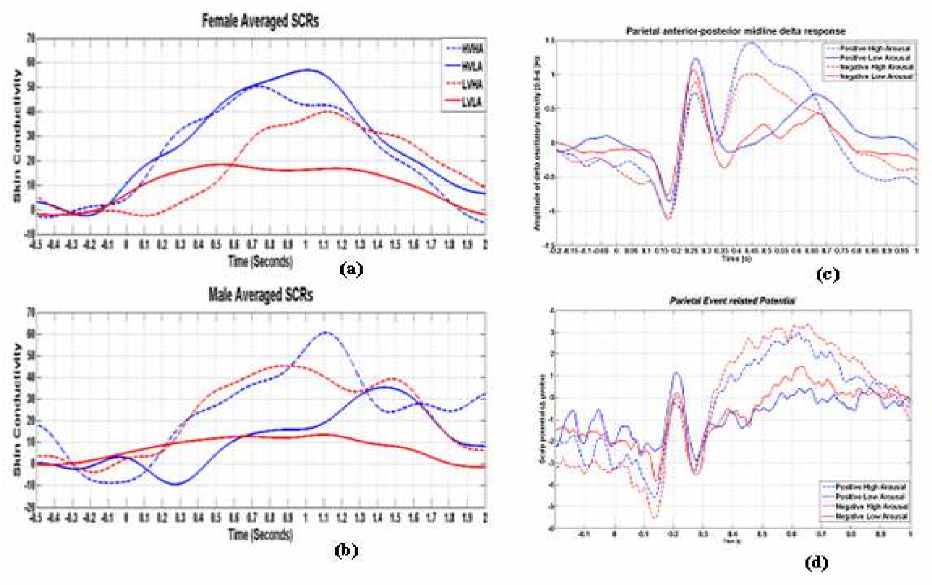

Fig. 1. Grand average waveforms of skin conductance responses are depicted for female subjects (a) and for male subjects (b). The Event-Related Potentials (ERPs) and Event-Related Oscillations (EROs) for the delta frequency band are represented in (c) and (d) respectively. In all subplots, "blue" indicates "High Valence", while "red" indicates "Low Valence"; "solid" curves represent "Low Arousal"; "dotted" curves represent "High Arousal".

2.1 The emotion Recognition Sub-System

The emotion recognition subsystem is comprised of the arousal and valence recognition sub-components. Firstly, the feature vector is classified according to its arousal status. The arousal differentiation is conducted using different features according to the subject's gender. Previous neuroscience studies have mentioned that males respond in a different way to the emotionally evocative stimuli than females [14]. Therefore, the gender effect is an important factor that should be taken into consideration during emotional interaction applications. The second part of the emotion recognition subsystem is intended to classify the feature vector according to its valence dimension. Once more, different feature vectors are used for the high and for the low arousing data segments.

The classifier used for the emotion recognition was based on the Mahalanobis distance and is computed as shown in equation (1):

$$D^2{}_M = (x - m_i)^T C_i^{-1} (x - m_i),\ \text{i=1, 2} \tag{1}$$

where C_i is the covariance matrix for the particular emotional category considered and T stands for the transposition operator. Let x be a feature vector being compared to a pattern class for which m_i is the class mean vector. The Mahalanobis distance is a minimum-distance classifier since a small value indicates a higher potential membership of the vector x in the emotional group under consideration.

The emotion recognition unit is the core component of the proposed application. According to the derived result, an XML file is created for the emotion description. This type of annotation is used in order to provide a standard format of the record which describes the classified emotion. Furthermore, it contains both quantitative (statistical) and qualitative information of both the signals and the derived features [15]. A block diagram of the emotion recognition subsystem is given in Figure 2.

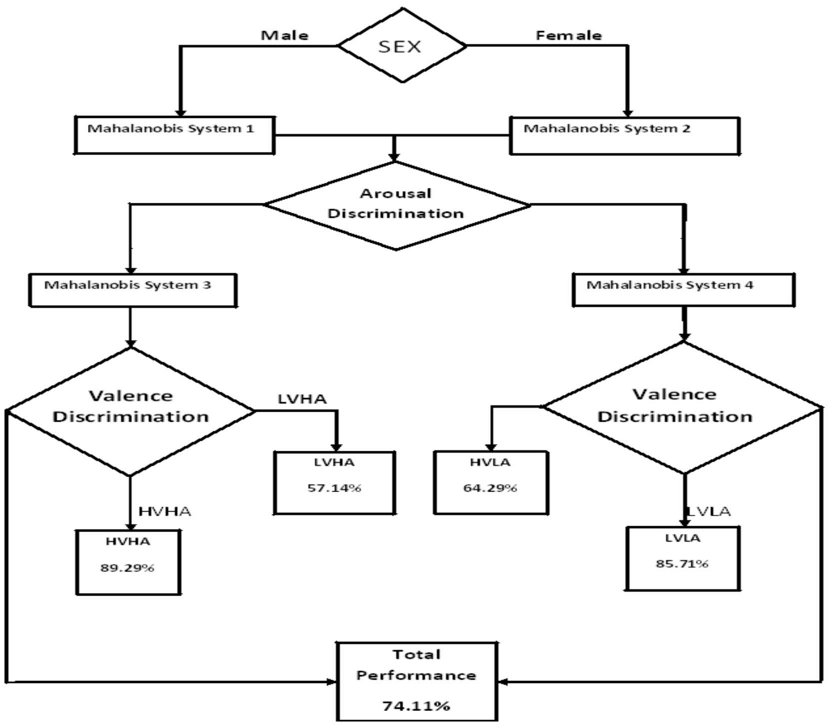

Fig. 2. Block diagram of the Emotion Recognition Sub-System and visualization of the classification ratings

2.2 Avatar Behavior Generator Sub-System

Attached to the emotion classification subsystem, an intermediate sub-system has been developed in order to transform the emotional data into Haptek HyperText commands [16] which act as the interface between the programming language and an anthropomorphic avatar. These commands change the avatar's characteristics (e.g. mouth, eyes, energy, etc.) which form suitable facial expressions according to the emotion identified by the classification sub-system. The output is a file with a predefined structure. The usage of the proposed intermediate sub-system offers platform independency, since the avatar implementation details remain hidden. So, it is easier for a new avatar to be added to the system in case of future web-based applications. Moreover, the same sub-system except from changing the avatar's facial expressions causes the avatar to respond with a voice message enhanced with some basic emotional features such as laugh, excitement, etc. [17]. This message is currently predefined according to the evaluated feature, with the sole purpose of attempting to counteract/neutralize the subject's emotional mood in case of experiencing a negative feeling or to enhance the subject's positive mood. The overall system architecture is visualized in Figure 3.

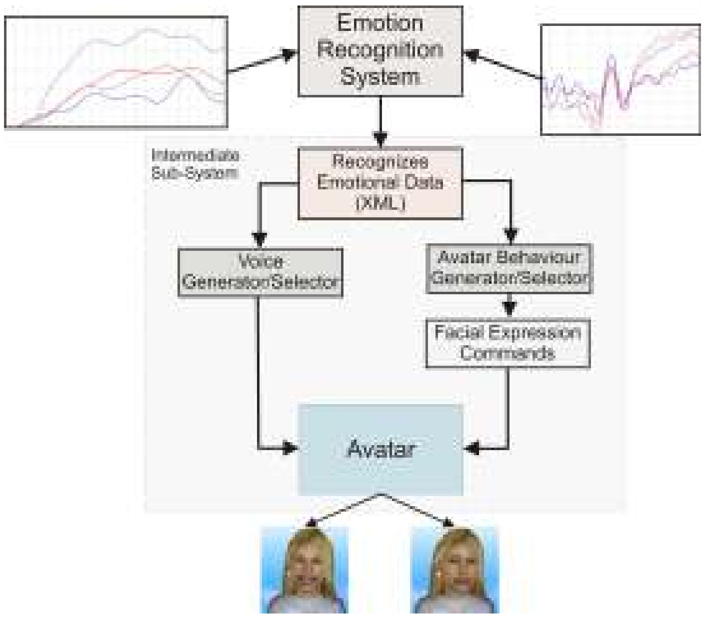

Fig. 3. Overall architecture of the system of integrated emotion recognition

As it can be seen, the significance of the proposed methodology lies mainly with the use of an affective protocol capable of eliciting and extracting the neuro-physiological signatures for a variety of discrete human emotions. The emotion recognition subsystem then adopts machine learning techniques, such as the Mahalanobis classifier in order to recognize the elicited emotion. The outcome is provided to an avatar that through expressions and verbal contents provides emotional feedback to the subject/user. The connectivity between the various subsystems is achieved by means of XML specifications since it provides platform independency [15]. The use of the avatar guarantees a multi-modal interaction, alongside a kind of emotional interaction by eliciting certain expressions (video) and speech (voice), thereby achieving human embodiment inside the computer. Consequently, the described approach aims to the enhancement of the human-computer interaction by combining features of the cerebral and the sympathetic activity in a structured way, and therefore attempts to establish a complete loop of affective HCI, which is presumably closer to the human-human interaction.

3 Results

The classification results indicate that pleasant and highly arousing stimuli, as well as, unpleasant and low arousing stimuli are classified in a very robust way, 89.29% and 85.71%, respectively. However, the classification accuracy of unpleasant highly arousing stimuli (57.14%) and pleasant low arousing stimuli (64.29%) is much lower. The overall/average performance obtained for the whole set of emotion categories reached 74.11%, as shown in Table 1.

Emotional Category	Performance Rate
HVHA	89.29%
HVLA	64.29%
LVHA	57.14%
LVLA	85.71%
Average Performance	**74.11%**

Table 1. Classification results for emotionally evocative stimuli

To investigate the inherent properties of the classification subsystem the obtained average performance was studied as a function of the feature number used. As shown in Figure 4, the classification performance is generally increased as more features are considered into the Mahalanobis distance computation.

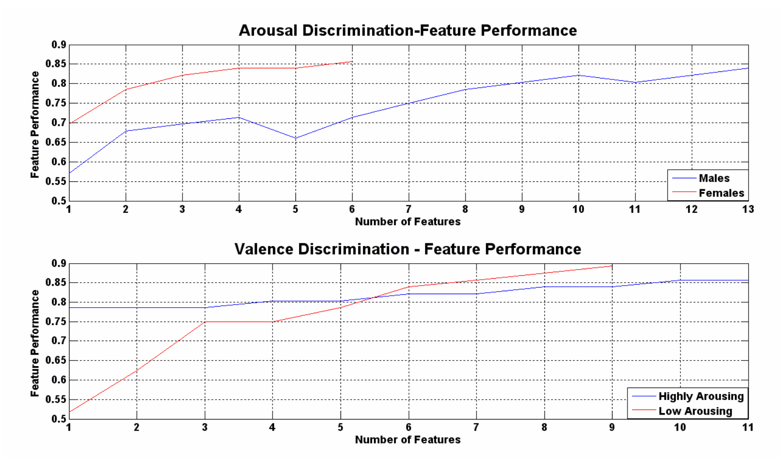

Fig. 4. Classification performance of the two components of the Emotion Recognition Sub-System in association with the number of features used by the classifier

The first subplot in Figure 4 depicts the accuracy rates for the arousal sub-component, while the second visualizes the valence discrimination task. According to the first sub-plot, the arousal discrimination task is simpler, faster and more accurate for females than males. Regarding the valence dimension, the low arousing stimuli can be differentiated much easier and in a more precise way than the highly arousing ones.

4 Discussion

The results indicate that the use of a multitude of features improves the overall classification capacity of the system. More specifically, there are certain ERPs or EROs features that are very good arousal indicators. On the other hand, some features derived from the electrodermal activity (SCRs) can discriminate between pleasant and unpleasant pictures. The fusion, of all these components accounts for the overall emotional discrimination result.

The use of the Mahalanobis distance as an emotion metric eliminates several drawbacks posed by linear classifiers such as the Euclidean distance. More specifically, the scaling of the coordinate axis as well as the correlations between the derived features is taken into consideration [18]. On the other hand, the computational cost is higher and the memory requirements grow quadratically as the number of the derived features increases. However, in the present case none of the four Mahalanobis sub-components employed more than thirteen features. The results obtained in the proposed study suggest that the use of Mahalanobis metrics is capable of classifying physiological recordings during emotional interaction paradigms.

As mentioned earlier, previous emotional interaction studies have employed various protocols, which generally can be divided in two main categories. Most of them, especially the earlier ones, used only a single subject which intentionally expressed affective states several times. These studies, have reported mainly better results ranging from 66% for valence discrimination [8] to 93.5% total performance [10]. In comparison to the aforementioned studies, the performance of the Mahalanobis classifier is much lower than the usage of neural network classifier (93.5%) [10], almost the same when using a quadratic classifier for arousal discrimination (84%) [8], a bit lower when combining linear and quadratic algorithms [9] with a Sequential Floating Forward Search technique (81.25%) and better in the case of valence discrimination by means of a quadratic classifier (66%) [8]. However, it should be clarified that the development of a user-independent classifier highly differs from extracting the neurophysiological pattern of a single subject during emotional processing. Therefore, any straightforward result comparison may lead to misinterpretations. In case of user-independent classifiers based on the k-Nearest Neighbor (kNN) method and Discriminant Function Analysis (DFA), the Mahalanobis classifier seems to perform better (74.11% versus 65.38% and 69.28%, respectively) [11].

The use of anthropomorphic avatars, which adopt its facial characteristics according to the user's affective state, facilitates the emotional interaction, since facial expressions are universally expressed and recognized by humans. Therefore, the proposed application may be employed in various applications like telemedicine, virtual or special education [19], monitoring of dangerous situations, entertainment, etc. Furthermore, the avatar is capable of taking action by generating a voice message in order to neutralize the user's negative emotions, such as fear, anger, disgust, melancholy, etc. Its beneficial use has already been reported by several studies which demonstrated that including an avatar as part of an emotional interaction interface helps to increase human performance [20]. Consequently, our study proposes the use of anthropomorphic avatars in order to mirror the user's affective state. In a future evolution of the system presented herein, we envisage the creation of emotional paradigms in terms of virtual user presence (users modeled as avatars, virtually interacting with each other). This notion may be employed in several applications such as virtual games, psychotherapy groups and specific web folkosonomies [21].

Last but not least, the whole approach is accompanied by a C# based system integrating the above in an internet downloadable form (to be demonstrated). The whole endeavor is taken in accordance with previous efforts to fuse the HCI domain with ideas from neurophysiology and medical informatics [12], [22], [23] so as to enrich the multimodal affective arsenal with more robust emotion identification tools.

5 References

1. Bradley, M.M., Codispoti, M., Cuthbert, B.N and Lang, P.J.: Emotion and motivation I: Defensive and appetitive reactions in picture processing. Emotion, vol. 1, no. 3, 276-298 (2001)
2. Lang, P.J., Bradley, M.M., & Cuthbert, B.N: International Affective Picture System (IAPS): Technical Manual and Affective Ratings, NIMH Center for the Study of Emotion and Attention (1997)
3. Picard, R.W. et al.: Affective learning – A Manifesto. BT Technology Journal, vol. 22, no. 4, 253-269 (2004)
4. Reeves, B., Nass, C.: The Media Equation: How People Treat Computers. Cambridge Univ. Press (1996)
5. Cockton, G.: Editorial:From doing to being:bringing emotion into interaction. Interacting with Computers, vol. 14(2), pp. 89-92, February 2002
6. Picard, R.W and Healey, J.: Affective Wearables. Personal and Ubiquitous Computing, vol.1, no. 4, pp. 231-240 (1997)
7. Konstantinidis, E.I, Bamidis, P.D, Koufogiannis, D.: Development Of A Generic And Flexible Human Body Wireless Sensor Network. In: 6th European Symposium On Biomedical Engineering (ESBME 2008), Chania, Crete Island, Greece
8. Healey, J., Picard, R.W: Digital processing of affective signals. In: International Conference on Acoustics, Speech and Signal Processing, vol. 6, pp. 3749-3752, Seattle, USA (1998)
9. Picard, R.W, Vyzas, E. and Healey, J.: Toward Machine Emotional Intelligence: Analysis of Affective Physiological State. IEEE, Transactions on Pattern Analysis and Machine Intelligence. vol. 23, no. 10, pp. 1175-1191, October 2001
10. Haag, A., Goronzy, S., Schaich, P. and Williams, J.: Emotion recognition Using Bio-Sensors: First Steps Towards an Automatic System. In: Affective Dialogue Systems. LNCS, vol. 3068, pp. 36-48. Springer, Heidelberg (2004)

11. Nasoz, F., Lisetti, C. L., Alvarez K., Finkelstein, N.: Emotion Recognition from Physiological Signals for User Modeling of Affect. In: 3rd Workshop on Affective and Attitude User Modeling. USA (2003)
12. Bamidis, P.D., Luneski, A., Vivas, A., Papadelis, C., Maglaveras, N., Pappas, C.: Multi-channel physiological sensing of human emotion: insights into Emotion-Aware Computing using Affective Protocols, Avatars and Emotion. Stud Health Technol Inform. 2007;129(Pt 2):1068-72.
13. Frantzidis, C., Lithari, C., Vivas, A., Papadelis, C., Pappas, C., Bamidis, P.D : Towards Emotion Aware Computing: a Study of Arousal Modulation with Multichannel Event Related Potentials, Delta Oscillatory Activity and Skin Conductivity Responses. In: 8th IEEE International Conference on BioInformatics and BioEngineering (BIBE 2008), pp. 1-6, Athens, Greece (2008)
14. Brody, L.R.: Gender differences in emotional development: A review of theories and research. Journal of Personality. vol. 53:2, pp. 102-149, April 2006
15. Luneski, A. and Bamidis, P. D.: Towards an Emotion Specification Method: Representing Emotional Physiological Signals. In: Proc. Of the 20th IEEE International Symposium on Computer-Based Medical Systems (CBMS'07), pp. 313-320, Maribor, Slovenia (2007)
16. The free Haptek Guide, http://haptek.com/, last access 24/2/2009
17. Loquendo: global supplier of speech recognition and speech synthesis technology and solutions, http://www.loquendo.com/, last access 24/2/2009
18. Cincotti, F. et al.: The use of EEG modifications due to motor imagery for brain-computer interfaces. IEEE Transactions on Neural Systems and Rehabilitation Engineering. vol. 11:2, pp. 131-133, June 2003
19. Luneski, A., Konstantinidis, E.I, Hitoglou-Antoniadou, M., Bamidis, P. D.: Affective Computer-Aided Learning for Autistic Children, In: 1st Workshop on Child, Computer and Interaction,ICMI 2008 Chania, Crete, Greece, 20-23, October 2008
20. Lisetti, C.L, Nasoz, F.: Using Noninvasive Wearable Computers to Recognize Human Emotions from Physiological Signals. EURASIP Journal on Applied Signal Processing. Vol. 2004(11), pp. 1672-1687 (2004)
21. Boulos K., Maged, N., Wheeler, S.: The emerging Web 2.0 social software: an enabling suite of sociable technologies in health and health care education. Health Information and Libraries Journal, vol. 24, no.1, pp. 2-23, March 2007
22. Vilon, O., Lisetti C.L.: Toward Recognizing Individual's Subjective Emotion from Physiological Signals in Practical Application, In: Proc. Of 20th IEEE International Symposium on Computer-Based Medical Systems (CBMS'07), pp. 357-362, Maribor, Slovenia (2007).
23. Luneski A., Bamidis P.D., Hitoglou-Antoniadou M. : Affective computing and medical informatics: state of the art in emotion-aware medical applications. Stud Health Technol Inform. 2008;136:517-22.