Workshop Proceedings of the 10th International Conference on Intelligent Environments
J.C. Augusto and T. Zhang (Eds.)

doi:10.3233/978-1-61499-411-4-80

Affordable Development of Animated Avatars for Conveying Emotion in Intelligent Environments

J. Alfredo SÁNCHEZ,[a,1] Oleg STAROSTENKO,[a] Paula MEDINA,[a] Ofelia CERVANTES,[a] and Wanggen WAN[b]
[a] *Laboratory of Interactive and Cooperative Technologies, Universidad de las Américas Puebla*
[b] *Institute of Smart City, Shanghai University*

Abstract. This paper presents an approach to the development of emotional avatars that is based on the use of widely available technologies, such as 3D modeling and animation software. We demonstrate the application of a simplified version of the Facial Action Coding System (FACS) to represent basic emotions (anger, disgust, fear, happiness, sadness and surprise) and build an experimental platform to explore how animation and various avatar traits may impact emotion representation and recognition. We experimented with subjects in Mexico and China. Our results show remarkable recognition levels for some prototypical emotions, particularly happiness, sadness and surprise. This is relevant for some intelligent environments in which detecting whether a user experiences a positive or negative emotion is helpful. Our results also show that animation is an important factor in conveying emotion. Finally, from observing subjects with very different cultural backgrounds (Mexico and China), we confirmed the universality of prototypical emotions and the wide applicability of our approach for emotion representation.

Keywords. Emotion representation, emotion recognition, avatars, FACS, action units.

Introduction

In intelligent environments, technology is expected to support activities carried out by people in adaptive and unobtrusive manners. In this scenario, *affect* plays a key role, as users' moods and emotions are critical in determining the extent to which their needs and preferences are fulfilled, and also in conveying important information when communicating with other users. Research in affective computing has been addressing three major challenges: (1) *affect recognition*, so computers become aware of people's emotions; (2) *affect synthesis*, so computers convey emotions through their behavior and user interfaces; and (3) *affect mediation*, so technology does not hinder (and could potentially enhance) emotion transmission when people communicate with each other using computers and other devices, which has become commonplace in practically all contexts.

[1] Corresponding author. Email: alfredo.sanchez@udlap.mx

There has been significant progress in affect recognition techniques. Salient approaches have been based on analyzing physiological signals [14], facial cues [3], speech variations [18], body postures [10; 20], brain activity [19], or a combination of multiple signals [13]. Our own work has produced a platform for assisting developers of affect-aware applications in the integration of multiple approaches for emotion recognition [16].

Affect synthesis and mediation have not received as much attention from researchers as affect recognition has. Progress in affect synthesis has relied on modeling emotional states to provide computer systems with means to reason about generating emotions and related behavioral variations [11]. Mediation of affect has relied, for example, on text analysis in instant messaging systems to infer affective states and make them more explicit [12; 8], or graphical representations of models from psychology, as in the case of Russell's circumplex model [17].

Both in affect synthesis and affect mediation, 2D and 3D anthropomorphic avatars have been applied to provide users with graphical representations of emotions. In the context of intelligent environments, avatars can help in supplying users with a natural conversational counterpart for purposes that can range from learning to entertainment to assisted search, customer service, or decision making. However, designing avatars so they offer an engaging and natural interface for users, has not been straightforward. Realistic representations, which require significant technical capabilities, are not sufficient to develop natural conversational agents. One key characteristic in making avatars appear natural is the portrayal of emotions. Whether an avatar is a personification of a computer component or a graphical representation of a user, emotional reactions enhance the avatar's credibility and the accuracy of the affective state to be conveyed, respectively.

This paper describes the development of a set of emotional avatars, intended to investigate the application of widely available tools and techniques to produce recognizable affective states. Our goal has been to demonstrate that it is possible to impart emotional traits to avatars by using a simplified version of an emotion coding system while maintaining the capability to convey affective states. We refer to our approach as affordable in the sense that it integrates existing components and relatively simple representations. We conducted experiments that provide insight into the potential and limitations of various representation mechanisms as well as on similarities and differences in emotion recognition in diverse cultural settings.

The remainder of the paper is organized as follows: Section 1 briefly discusses current work in the area of emotional avatars. Section 2 provides details on the design and implementation of our set of avatars, whereas our experiments and observations are discussed in Section 3. Finally, implications and conclusions derived from our work thus far are presented in Section 4.

1. Related Work

The importance and impact on users of embodied agents that convey affective states have been studied from multiple perspectives. Beale and Creed [1] review work on emotional avatars in four major application areas: Education, video games, collaboration, and assistive technologies for behavior change. They also identify the need to work with users in order to understand how emotion representations impact people and to be able to produce more expressive avatars. Our work with diverse communities is aimed in this direction.

Sensing and communicating user emotions in real time in virtual environments has been the goal of the work reported by Di Fiore et al. [4]. They extract facial features from a video stream and select pre-defined poses to animate 2D avatars that display emotions that correspond to the user's expressions. We envision our 3D avatars set as a component of systems such as this, but we aim to endow avatars with more realistic and expressive features.

Fabri and Moore [7] designed an avatar that uses the Facial Action Coding System (FACS) to represent basic emotions. Although their avatar is not realistic and their work is conducted in the specific context of educational collaborative virtual environments, they made progress in the use of limited facial features to represent emotions and experimenting with users to assess expressivity and recognition. They compare the recognition of emotions using an avatar versus using photographs of human faces. Our work goes further by introducing a more realistic representation of avatars, modeling variations of some of the prototypical emotions, and studying the impact of animation and cultural background.

Approaches to provide avatars with emotion expressivity are varied. Gralewski et al. [9] reported an interesting approach, consisting of extracting textures from video sequences. They define an "expression space" based on those textures and allow users to browse and select emotions to generate new expression sequences. Compared to our work, theirs is based on data collected from actual faces, which makes it realistic but specific, whereas ours is based on general principles for emotion expression, but our avatars are fictitious.

The importance and appeal of avatars have motivated the availability of commercial applications that offer assistance in the use or construction of anthropomorphic avatars for various purposes. For example, Micro Expression Training Tool (http://www.paulekman.com/products/) uses photographs and videos to train people in the use of FACS. SitePal (www.sitepal.com) is a service that allows clients to create animated speaking characters with some emotional features intended for enhancing e-commerce websites. FaceGen (www.facegen.com) is a popular software package used to generate models of realistic human faces that can be exported for use in various applications.

Roesch et al. [15] have used FaceGen models as the basis for FACSGen, a tool for creating realistic characters with emotional expressions based on FACS. FACSGen allows users to manipulate FACS parameters so as to generate a wide range of facial expressions. Experimentation with FACSGen has involved asking users (both FACS experts and novices) about how close each resulting expression represents the intended emotion, and has reported generally satisfactory recognition levels. Our work differs in goals, as we intend to demonstrate the viability of conveying emotion with simplified representations, and also in methodology for validation, as we have relied only on lay users and have asked them to identify represented emotions. Results, however, are consistent with this and other existing research.

2. Modeling Emotion in Animated Avatars

In order to experiment with emotion representation and recognition using a simplified emotion model and widely available technologies, we designed a set of four avatars with diverse features intended to provide alternatives for observation and experimentation in emotion recognition. We present our design decisions in this section, with emphasis on how we model emotions and general facial modeling and animation.

2.1. Representing Emotion

Various techniques have been used to represent emotions in avatars. We have based our avatar design on the use of the Facial Action Coding System (FACS), which suggests the notion of Action Units (AUs), the relaxation or contraction of one or more facial muscles, associated with the expression of emotions [5]. Thus, for instance, *happiness* is associated with the activation of the cheek raiser (*orbicularis oculi, pars orbitalis*) and the lip corner puller (*zygomaticus major*) muscles, whereas *disgust* involves the nose wrinkler (*levator labii superioris alaeque nasi*), the lip corner depressor (*depressor anguli oris*, also known as *triangularis*), and the lower lip depressor (*depressor labii inferioris*) muscles. Specific dictionaries, such as FACSAID [6], provide full listings derived by specialists and codes that make it easier for researchers from other disciplines to apply FACS for various purposes. AUs are assigned numerical codes (from 1 to 46) as well as an intensity scale (from A to E, where A is just a trace and E is the maximum intensity), so an emotion such as intense happiness might be expressed as 6C+12E, whereas slight disgust could be 9A+15A+16A.

FACS has been used widely for emotion recognition purposes, both by humans and computer programs. Workshops are offered to train professionals in detecting AUs and inferring corresponding emotions, whereas images and video recordings are analyzed automatically to detect combination of AUs and their associated emotions.

Table 1. Action units (AUs) selected to represent prototypical emotions.

Emotion	AUs	Muscles involved
Anger 1	4 + 7 + 23	Brow lowerer, upper lid raiser, lid tightener, lip pressor
Anger 2	4 + 7 + 25	Brow lowerer, upper lid raiser, lid tightener, lips part
Disgust 1	4 + 10 + 17	Brow lowerer, upper lip raiser, chin raiser
Disgust 2	4 + 25 + 26	Brow lowerer, lips part, jaw drop
Happiness	12	Lip corner puller
Sadness	1 + 15	Inner brow raiser, lip corner depressor
Surprise	1 + 5 + 26	Inner brow raiser, upper lid raiser, jaw drop
Fear 1	4 + 5 + 25	Brow lowerer, upper lid raiser, lips part
Fear 2	1 + 5 + 25	Inner brow raiser, upper lid raiser, lips part
Contempt	R12+62	Right lip corner puller, eyes turn right

Following on our previous work [2], we opted for using FACS for emotion synthesis and focused on representing prototypical emotions (anger, disgust, fear, happiness, sadness and surprise). We also modeled contempt, which has been considered a universal emotion, though it has also been regarded as a secondary emotion that results from the mix of disgust and anger. In order to keep our design simple, we did not include provisions for varying emotion intensity in our current work.

Naturally, given the complexity and variability of human expressions, there is not a unique combination of AUs associated with each emotion. Since our goal has been to provide a simplified representation of emotions, we implemented only some of the possible combinations. In our avatar set, we modeled anger, disgust and fear with two distinct combinations of AUs for each emotion, whereas only one combination was used in each case for happiness, sadness, surprise and contempt. The AUs used in each case are summarized in Table 1. It should be noticed that our selection of action units is based on an ontology developed by Contreras [2], which delimits the AUs that are necessary to recognize prototypical emotions as well as some graphical characteristics (feature points and animation parameters) that have been taken into account for the visual representation discussed next.

2.2. 3D Modeling and Animation

Although an important component of a realistic avatar is its three-dimensional model, for the purposes of our work this is only the basis upon which emotional features are built. Our 3D models were constructed using the *3D Studio Max* software, which provides means to create and modify virtual characters. Polygon meshes were defined for four characters, codenamed *Blanca, Jenny, Carlos* and *Jamal*. Two of these models are illustrated in Figure 1.

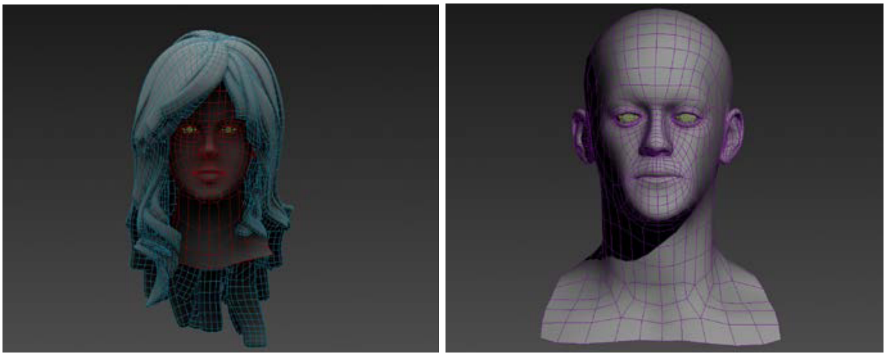

Figure 1. Polygon meshes for two avatars.

According to the MPEG-4 standard for face and body animation (MPEG-4 FBA), 84 points known as Facial Definition Parameters (FDPs) can be used to characterize a human face, whereas 68 Face Animation Parameters (FAPs) represent a complete set of facial actions that allow for the representation of natural facial expressions. Using our basic 3D models, rigs were constructed and were given a skeletal structure to match some FDPs and FAPs that make animation possible and natural. This structure and control point matching is illustrated in Figure 2, where FDPs are indicated by dots (red in the original images). Texture was added using *Photoshop* and the resulting models were exported to *Unity3D*, which was used to implement the animation needed to change from a neutral face to each of the prototypical emotions, as well as to present an interface for users to select a desired expression. Animations last actually five seconds to change an expression from neutral to a selected emotion. A snapshot of the resulting interface, which has been named “Chopiface,” is presented in Figure 3. In this figure, the avatar named “Jenny” has been selected and the button corresponding to “Fear 2” has been pressed. A working prototype is publicly available at http://chopis.com.

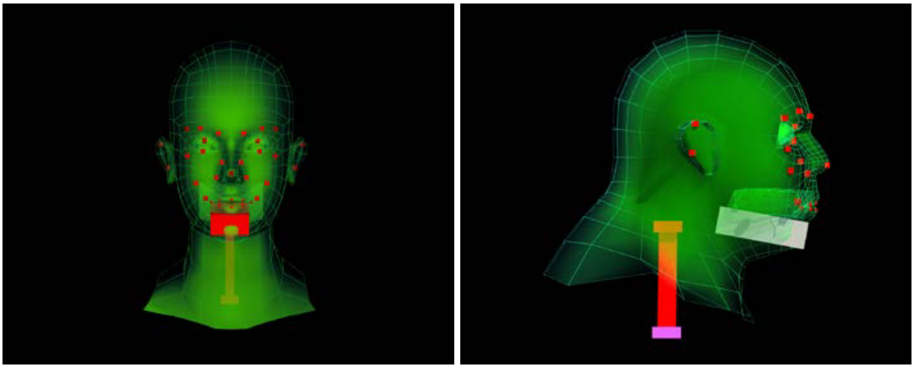

Figure 2. Bone structure for animation of avatars.

Figure 3. User interface after selecting an avatar and a specific emotion (fear).

Our current implementation is intended for observation and experimentation, hence the various buttons in the interface that allow users to select emotions to be acted by the avatar. This kind of interface could be useful also in a training scenario, in which users (e.g., students or employees) could learn about micro-expressions associated with specific emotions. However, we envision this prototype as a component that can be part of more complete affect-aware environments. For example, in an immersive environment, avatars can represent users and their affective states by matching FDPs as detected in the actual image of the users with those in their avatars. Remote users can thus convey their emotions and become aware of other users’ emotions through their avatars in real time. Another possibility is to include avatars in e-learning environments, in such a way that avatars express emotions in response to student progress or performance.

Figure 4. Modified interfaces for experimentation.

Figure 5. Scenario for experimentation.

3. Evaluating Representation and Recognition

We have conducted three experiments with users in order to evaluate the capabilities of our avatars to convey the intended emotions. In each case, we modified the interface so as to conceal button labels that indicate the name of the intended emotion. Instead, buttons were labeled with single capital letters, from A to J, as shown in Figure 4, in which all four avatars are displayed. Users were presented with this interface as well as with a printed list of the seven emotions, as illustrated in Figure 5. After pressing each button, they were asked to select the emotion from the list that most closely was represented by the animated avatar. The results from each experiment are discussed next.

3.1. Emotion Recognition Using Animated Avatars

A first experiment was conducted to investigate whether users would recognize emotions intended by our avatar design. A total of 96 users participated, 65 male, 31 female, their ages ranging from 14 to 56 (mean was 29, standard deviation 9.7). Years of schooling also varied significantly: zero (2 subjects), six (19 subjects), 9 (25 subjects), twelve (24 subjects) and 17 (26 subjects). Subjects were presented with one of the four avatars (so each avatar was presented 26 times) and the system displayed an animation from a neutral expression to that of an emotion after each button (labeled A-J) was pressed. Figure 6 summarizes the data collected from these observations.

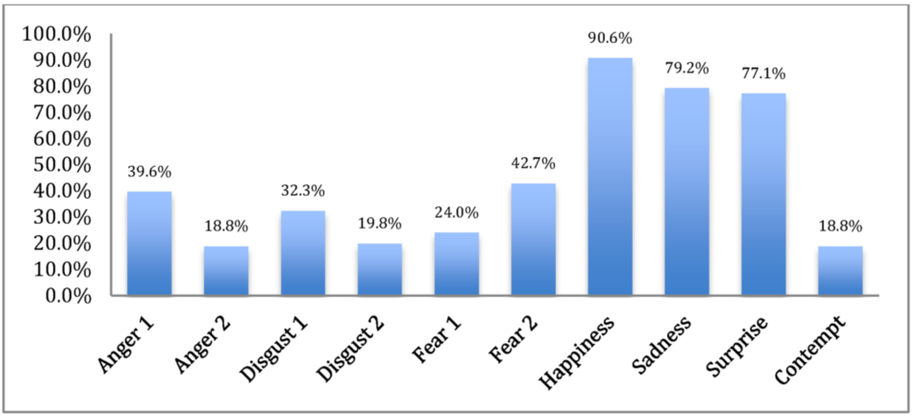

Figure 6. Recognition of emotions with animated avatars.

We can observe that happiness, sadness and surprise were consistently recognized by a great majority of the participants. One of the representations of fear and one of anger stand out as recognized by a good number of participants, but still tend to be confusing or ambiguous for most users. Only few participants identified correctly the alternative representation of fear, as well as those of disgust and contempt.

We probed into the data to investigate whether factors such as the subject's age or years of schooling, or the avatar's gender or race had any impact on emotion recognition, but we did not find indication of any such correlations.

3.2. Comparing emotion recognition with and without animation

We conducted a second experiment with a different sample of users in order to study the impact of animation in improving recognition. In this case, we observed 60 subjects (35 male, 25 female), ages averaging 32.7 (standard deviation 12.6). Subjects also were presented with one of the avatars but the expression changed from neutral to an intended emotion instantly, without any animation involved. Results from our observations are summarized in Figure 7. Recognition results were consistent with the previous experiment, but the figure also shows clearly that the use of animation does improve recognition consistently (only in one case, "*anger 2*", fewer subjects did not recognize the intended emotion when the avatar's expression was animated). What may be particularly interesting is that emotions that are typically hard to identify, such as contempt and “disgust 1”, became harder to recognize when animation was not used (18.8% recognized contempt with animation, barely 4% without animation).

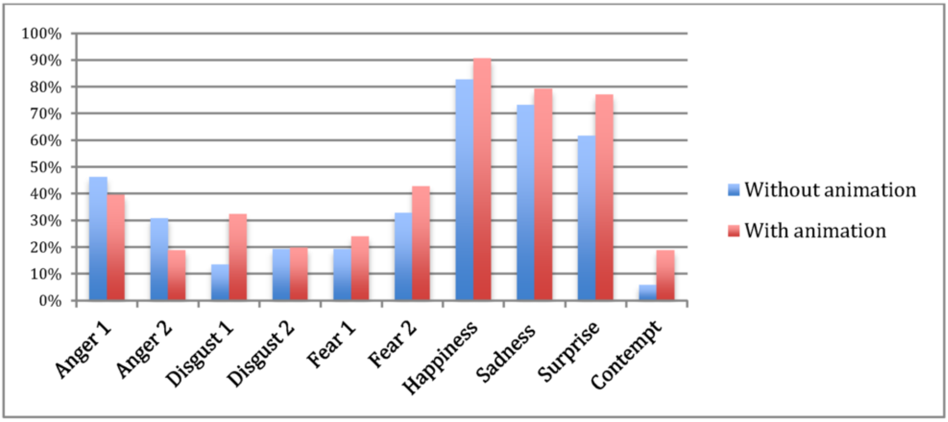

Figure 7. Assessing the impact of animation in emotion recognition

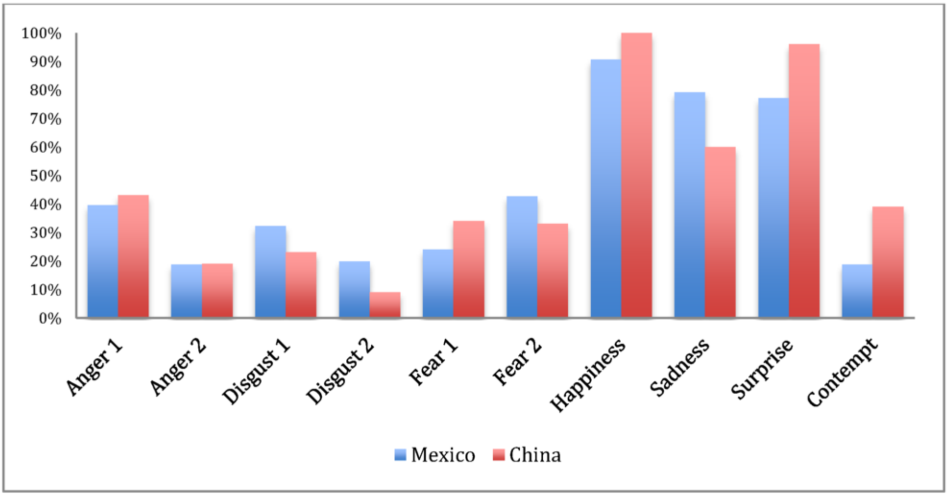

Figure 8. Comparing emotion recognition in diverse cultural settings.

3.3. Emotion Recognition in Diverse Cultural Settings

Both experiments 1 and 2 involved participants from Mexico. In our third experiment we considered 96 participants, just like in our first study, only this time all participants were Chinese (48 male, 48 female, all college students). Figure 8 presents results from our observations.

These results show that cultural background does not have a significant impact on the recognition of emotions as represented by our avatars. Though expected, given the universality of the so-called prototypical emotions, this is an important result as it validates positively the features we have selected for representing emotions in our avatars. Even the result with the greatest difference, again contempt (recognized only 19% of the times by Mexican subjects and 39% by Chinese subjects), is consistent with the controversy of whether this should be considered a prototypical emotion.

3.4. Some Implications

Avatars that are capable of conveying emotions can play a key role in intelligent environments. Our results show that it is possible to effectively convey *happiness*, *sadness* and *surprise* through 3D avatars by using only a reduced set of action units of the FACS system. We would like to emphasize the implications of these findings, especially for the developer of applications involving affect synthesis and affect mediation. From the affect synthesis perspective, it is helpful to know that representing these emotions in avatars requires only a few graphical features that can be enhanced by animation. From the affective mediation's standpoint, sensing these three prototypical emotions can be fundamental as a complementary input for intelligent environments to initiate appropriate actions when users engage in activities that involve communication through their avatars. When considering alternative representations for emotions using FACS, some combinations of AUs are consistently better recognized by users. Based upon our experimental results, we can suggest that the combinations used for "Fear 1" (4 + 5 + 25), "Disgust 2" (4 + 25 + 26) and "Anger 2" (4 + 7 + 25) would not be the most appropriate for implementation. Naturally, it is desirable to be able to detect all nuances of human affect in real time, and work is needed to improve their representation and recognition, but there are applications in which system behavior can be adapted meaningfully by considering only whether the user is exhibiting positive or negative reactions, such as happiness, sadness or surprise.

4. Conclusions

We have documented our experience with the design and development of 3D avatars that convey prototypical emotions by using widely available tools (3D Studio Max, Photoshop, Unity3D) as well as a simplified set of action units defined by the Facial Action Coding System (FACS). We also have conducted experiments to investigate the potential and limitations of our avatar representations with actual users in two very diverse cultural settings. Other developers and researchers can benefit from our results by considering the use of readily available tools as well as means to simplify the expression of emotions. Our experimental results show that animation plays an important role in emotion recognition, that cultural diversity does not have a significant impact on the way 3D avatars convey emotions, and that key emotions such as happiness, sadness and surprise can be modeled and implemented effectively using our approach.

Indeed, there is ample room for improvement and further research. Emotion intensity appears as a feature that could be modeled and that could improve recognition significantly. Also, the implementation of a service-oriented architecture would make it possible to incorporate our avatars into diverse applications, such as educational software, video games, and immersive intelligent environments.

Acknowledgments

Support for this research has been provided by the FP7 Program through the project Customised Advisory Sustainable Manufacturing Services (CASES), the Chinese Recruitment Program of High-end Foreign Experts, the National High Technology Research and Development Program of China (863 Program) under Grant No. 2013AA01A603, and the National Nature Science Foundation of China under Grant No. 61373084. We thankfully acknowledge the participation of a significant number of volunteers, especially of Yin Yanlu, Yang Chao and Xao Xu, who helped in conducting experiments and collecting data from Chinese subjects.

References

[1] Russell Beale and Chris Creed. Affective interaction: How emotional agents affect users. *Int. J. Hum.-Comput. Stud.* 67, 9 (September 2009), 755-776.

[2] R. Contreras, O. Starostenko, V. Alarcon-Aquino, and L. Flores-Pulido. Facial feature model for emotion recognition using fuzzy reasoning. In *Proceedings of the 2nd Mexican conference on Pattern recognition: Advances in pattern recognition* (MCPR'10), Jesús Ariel Carrasco-Ochoa, José Francisco Martínez-Trinidad, and Josef Kittler (Eds.). Springer-Verlag, Berlin, Heidelberg (2010), 11-21.

[3] Albert C. Cruz, Bir Bhanu, and Ninad Thakoor. 2012. Facial emotion recognition with expression energy. In *Proceedings of the 14th ACM international conference on Multimodal interaction* (ICMI '12). ACM (2012), 457-464.

[4] Fabian di Fiore, Peter Quax, Cedric Vanaken, Wim Lamotte, and Frank Reeth. Conveying Emotions through Facially Animated Avatars in Networked Virtual Environments. In *Motion in Games*, Arjan Egges, Arno Kamphuis, and Mark Overmars (Eds.). Lecture Notes In Computer Science, Vol. 5277. Springer-Verlag, Berlin, Heidelberg (2008), 222-233.

[5] P. Ekman and W. V. Friesen (1978). The facial action coding system: A technique for measurement of facial movement. Web document, available from http://face-and-emotion.com/dataface/ facs/description.jsp.

[6] P. Ekman, and E. L. Rosenberg (Eds.). *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)* (2nd ed.). Oxford University Press (2005), USA.

[7] M. Fabri, D. J. Moore, D. J. Hobbs. Mediating the Expression of Emotion in Educational CollaborativeVirtual Environments: An Experimental Study, in International Journal of Virtual Reality, Springer Verlag, London (2004). Available online at http://dx.doi.org/10.1007/s10055-003-0116-7.

[8] Gary Garcia-Molina, Tsvetomira Tsoneva, and Anton Nijholt. Emotional brain-computer interfaces. *Int. J. Auton. Adapt. Commun. Syst.* 6, 1 (December 2013), 9-25.

[9] Lisa Gralewski, Neill Campbell, Barry Thomas, Colin Dalton, and David Gibson. Statistical synthesis of facial expressions for the portrayal of emotion. In *Proceedings of the 2nd international conference on Computer graphics and interactive techniques in Australasia and South East Asia* (GRAPHITE '04), Stephen N. Spencer (Ed.). ACM, New York, NY, USA (2004), 190-198.

[10] Fernando López-Colino and José Colás. Synthesizing mood-affected signed messages: Modifications to the parametric synthesis. *Int. J. Hum.-Comput. Stud.* **70**, 4 (April 2012), 271-286. DOI=10.1016/j.ijhcs.2011.11.003 http://dx.doi.org/10.1016/j.ijhcs.2011.11.003.

[11] Kwang-Mo Jung, Sung-Hee Hong, Byoung-Ha Park, Kwang-Soon Choi, and Young-Choong Park. FPGA implementation of human emotion reasoning engine. In *Proceedings of the 2013 International Conference on Information Networking (ICOIN)* (ICOIN '13). IEEE Computer Society (2013), 157-162.

[12] Chunling Ma, Helmut Prendinger, Mitsuru Ishizuka. 2005. A Chat System Based on Emotion Estimation from Text and Embodied Conversational Messengers. Entertainment Computing – ICEC 2005. LNCS, Springer, 2005. Pp. 535-538.

[13] Imen Tayari Meftah, Nhan Le Thanh, and Chokri Ben Amar. Multimodal approach for emotion recognition using a formal computational model. *Int. J. Appl. Evol. Comput.* 4, 3 (July 2013), 11-25.

[14] Christian Peter, Randolf Schultz, Jörg Voskamp, Bodo Urban, Nadine Nowack, Hubert Janik, Karin Kraft, and Roland Göcke. EREC-II in use: studies on usability and suitability of a sensor system for affect detection and human performance monitoring. In *Proceedings of the 12th international*

conference on Human-computer interaction: intelligent multimodal interaction environments (HCI'07), Julie A. Jacko (Ed.). Springer-Verlag (2007), 465-474.

[15] E. B. Roesch, L. Tamarit, L. Reveret, D. Grandjean, D. Sander, and Scherer, K. R. (2011). FACSGen: a tool to synthesize emotional facial expressions through systematic manipulation of facial action units. *Journal of Nonverbal Behavior*, *35*(1), 1-16.

[16] J. Alfredo Sánchez, Ximena Cortés, Oleg Starostenko, Ofelia Cervantes, and Wanggen Wan. 2014. An extensible platform for seamless integration and management of applications for emotion sensing and interpretation. (Submitted).

[17] J. Alfredo Sánchez, Norma P. Hernández, Julio C. Penagos, and Yulia Ostróvskaya. 2006. Conveying mood and emotion in instant messaging by using a two-dimensional model for affective states. In *Proceedings of VII Brazilian symposium on Human factors in computing systems* (IHC '06). ACM (2006), 66-72. DOI=10.1145/1298023.1298033.

[18] Björn Schuller, Anton Batliner, Stefan Steidl, and Dino Seppi. Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge. *Speech Commun.* 53, 9-10 (November 2011), 1062-1087.

[19] Xiao-Wei Wang, Dan Nie, and Bao-Liang Lu. 2014. Emotional state classification from EEG data using machine learning approach. *Neurocomput.* 129 (April 2014), 94-106.

[20] Yisu Zhao. 2012. *Human Emotion Recognition from Body Language of the Head Using Soft Computing Techniques*. Ph.D. Dissertation. University of Ottawa, Ottawa, Ont., Canada, Canada. AAINR98175.