Facial Affect Sensing for T-learning

Isabelle Hupont[1], David Abadía[1], Sandra Baldassarri[2], Eva Cerezo[2], and Rafael Del-Hoyo[1]
[1] Interactive Audiovisual Technologies Centre, Aragon Institute of Technology, Huesca, Spain.
{ihupont, dabadia, rdelhoyo}@ita.es
[2] Computer Science and Systems Engineering Department, University of Zaragoza, Zaragoza, Spain.
{sandra, ecerezo}@unizar.es

Abstract—**Interactive Digital TV has arisen new forms of interaction not traditionally associated with this medium. One of its applications is the so-called "t-learning" or TV-based interactive learning. To date, most existing t-learning applications confine themselves to edutainment more than to more complex and pedagogical forms of learning, due to the technological constraints imposed by set-top boxes. This paper overcomes those technological barriers to propose the first t-learning affective aware tutoring tool. The tool allows to capture by means of a camera the facial expressions of the student while performing evaluations (tests) through a broadcasted interactive t-learning application at home. It integrates a novel Artificial Intelligence method for facial affect recognition, where facial expressions are evaluated with a psychological 2-dimensional continuous affective approach. Thanks to it, the t-learning tool is able to automatically extract emotional information from the learner, which is further presented in a simple and efficient way to the distance tutor so that he can be aware of student's encountered difficulties and emotional progression during the learning process.**

Affective computing; emotional intelligence; facial expression classification; intelligent tutoring systems; interactive digital television.

I. INTRODUCTION

The growing success of IDTV (Interactive Digital TV) has arisen new services that have not traditionally been associated with this medium, such as commerce or learning. To date, the offer has been mostly based on the contents available through broadcast, but the increasing availability of broadband communications in digital interactive television receivers (set-top boxes), together with the fact that IDTV users are abandoning their passive habits, envisages a new range of highly interactive services. Moreover, being an enhancement to traditional TV sets, IDTV is easy to use and well known for everybody, meeting the socially important need to offer online services to people who cannot afford to buy a computer or lack the knowledge to use such technologies.

Interactive Digital TV is emerging as a potential important medium to create opportunities for learning at home. The term t-learning has been adopted [1] to denote TV-based interactive learning. Although the World Wide Web (WWW) based distance learning methods seem to still be the current dominating trend, the utility of television itself as a learning tool is also well recognized [2]. This is especially true for the socially disadvantaged communities where television has far more penetration than WWW interaction. As people are used to TV environment, and by taking into account the high impact in society due to its nature as mass media, interactive TV is a good opportunity to reduce the digital gap and to provide new ways of learning. Moreover, since virtually every household has access to a TV set, expensive classroom setups are not necessary providing new learning opportunities for those social groups that would hardly have access to traditional forms of education.

T-learning is not just an adaptation for IDTV of the e-learning techniques used in the Internet. It has its own distinctive characteristics, mostly related to the usability and technological constraints imposed by the television set and the set-top box, such as the fact of using a simple remote control to operate them -which reduces the possibilities of interaction with the student- or the fact that set-top boxes have lower computer power than a personal computer. For that reason, in most t-learning applications, learning via IDTV has been more edutainment than formal learning [3]. More engaged and intelligent t-learning interactive applications are needed to achieve a more complete and efficient learning, such as tutoring systems where the student and the tutor can interact as in traditional learning.

Emotions play an essential role in daily tasks, such as perception and learning [4]. The main difference between an expert human teacher and a distance learning tutoring tool is that the former recognizes and addresses the emotional state of learners to, based upon that observation, take some action that positively impacts learning (e.g. by providing support to a learner who is likely to otherwise quit). Giving those kinds of perceptual abilities to distance tutoring systems would considerably benefit the learning process. However, to our knowledge, there is any t-learning tutoring system in the literature with the ability of intelligently recognize affective cues from the student.

Facial expressions are the most powerful, natural and direct way used by humans to communicate affective states. Thus, making a t-learning tutoring system able to interpret facial expressions would allow it to be affective-aware and therefore more pedagogical.

Facial expressions are often evaluated by classifying face images into one of the six universal "basic" emotions or categories proposed by Ekman [5] which include "happiness", "sadness", "fear", "anger", "disgust" and "surprise" [6, 7]. There are a few tentative efforts to detect non-basic affective

states, such as “fatigue”, “interested”, “thinking”, “confused” or “frustrated” [8, 9]. In any case, this categorical approach, where emotions are a mere list of labels, fails to describe the wide range of emotions that occur in daily communication settings and ignores the intensity of emotions.

To overcome the problems cited above, some researchers, such as Whissell [10] and Plutchik [11], prefer to view affective states not independent of one another but rather related to one another in a systematic manner. They consider emotions as a continuous 2D space whose dimensions are evaluation and activation. The evaluation dimension measures how a human feels, from positive to negative. The activation dimension measures whether humans are more or less likely to take some action under the emotional state, from active to passive. Dimensional representations are attractive because they provide a way of describing a wide range of emotional states. They are much more able to deal with non-discrete emotions and variations in emotional states over time, since in such cases jumping from one universal emotion label to another would not make much sense in real life scenarios, such as learning scenarios. However, very few works have chosen a dimensional description level, and in the few that do the problem is simplified to a two-class (positive vs negative and active vs passive) [12] or a four class (quadrants of 2D space) [13] classification, thereby losing the descriptive potential of 2D space.

This paper proposes the first t-learning affective aware tutoring tool. Section II describes a method for continuous facial affect recognition, able to output the exact 2D location in the evaluation-activation space of a facial expression and thus to consider intermediate emotional states. This method is based on the combination of different Artificial Intelligence algorithms and is capable of analyzing any subject, male or female of any age and ethnicity, since it has been validated with an extensive universal database. In Section III the t-learning tutoring tool itself is explained in detail. The tool defies IDTV technical limitations by achieving to capture through a camera the facial expressions of the student while performing the evaluations (tests) proposed by a broadcasted interactive t-learning application at home. It integrates the facial affect recognition method to intelligently extract emotional information from the captured video and present it to the distance tutor in a simple and efficient way so that he can be aware of student’s encountered difficulties during the learning process. Finally, Section IV comprises the conclusions and future work.

II. A Novel Method for Continuous Facial Affect Recognition

This section describes a novel method for sensing facial emotions in a continuous 2D affective space. Section A explains the selection of the features serving as inputs to the system. Section B describes the facial expressions classification method itself. Finally, the results of emotional classification obtained in the 2D space are analyzed in detail in section C taking human assessment into account.

A. *Facial Inputs Selection*

Many studies suggest that all the necessary information for the recognition of facial expressions is contained in the deformation of a set of selected characteristics of the eyes, mouth and eyebrows [5]. Following that methodology, the initial inputs of our system were established in a set of distances and angles obtained from 20 characteristic facial points. In fact, the inputs are the variations of these angles and distances with respect to the “neutral” face. The chosen set of initial inputs compiles the distances and angles that have been proved to provide the best classification performance in other existing works [6, 7]. The points are obtained thanks to faceAPI [14], a commercial real-time facial feature tracking program that provides Cartesian facial 3D coordinates. It is able to track up to +/- 90 degrees of head rotation and is robust to occlusions, lighting conditions, presence of beard, glasses, etc. The initial set of parameters tested is shown in Fig. 1. In order to make the distances values consistent (independently of scale, distance to the camera, etc.) and independent of the expression, all the distances are normalized with respect to the distance between the eyes (“ESo”). The choice of angles provides a size invariant classification and saves the effort of normalization.

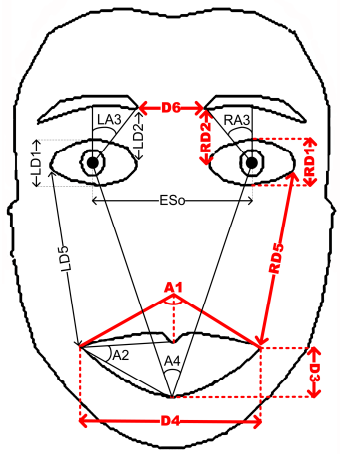

Figure 1. Facial parameters tested (in bold, the final selected parameters).

In order to determine the goodness of the parameters, a study of the correlation between them was carried out using the data (distance and angle values) obtained from a set of training images. For this purpose, two different facial emotion databases were used: the FGNET database [15] that provides video sequences of 19 different Caucasian people, and the MMI Facial Expression Database [16] that holds 1280 videos of 43 different subjects from different races (Caucasian, Asian, South American and Arabic). Both databases show Ekman’s six universal emotions plus the “neutral” one. A new database has been built for this work with a total of 1500 static frames selected from the apex of the video sequences from the FG-NET and MMI databases.

A correlation-based feature selection technique [17] was carried out in order to identify the most influential parameters in the variable to predict (emotion) as well as to detect redundant and/or irrelevant features. Subsets of parameters that are highly correlated with the class while having low

intercorrelation are preferred. In that way, from the initial set of parameters only the most significant ones were selected to work with: RD1, RD2, RD5, D3, D4, D6 and A1 (in bold in Fig. 1). This reduces the number of redundant and noisy inputs in the model and thus computational time, without losing relevant facial information.

B. Facial Expressions Classification in a Continuous 2D Affective Space

The facial expressions classification method starts with a classification mechanism in discrete emotional categories that intelligently combines different classifiers simultaneously to obtain a confidence value to each Ekman universal emotional category (section B.1). This output is subsequently expanded in order to be able to work in a continuous emotional space and thus to consider intermediate emotional states (section B.2).

1) Classifiers selection and combination for discrete emotional classification

In order to select the best classifiers to achieve discrete emotional classification, the Weka tool was used [18]. This provides a collection of machine learning algorithms for data mining. From this collection, five classifiers were selected after benchmarking: RIPPER, Multilayer Perceptron, SVM, Naive Bayes and C4.5. The selection was based on their widespread use as well as on the individual performance of their Weka implementation.

A 10-fold cross-validation test over the 1500 training images has been performed for each selected classifier. The success rates obtained for each classifier and each emotion are shown in the first five rows of Table I. As can be observed, each classifier is very reliable for detecting certain specific emotions but not so much for others. For example, the C4.5 is excellent at identifying "joy" (92.90% correct) but is only able to correctly detect "fear" on 59.30% of occasions, whereas Naive Bayes is way above the other classifiers for "fear" (85.20%), but is below the others in detecting "joy" (85.70%) or "surprise" (71.10%). Therefore, an intelligent combination of the five classifiers in such a way that the strong and weak points of each are taken into account appears as a good solution for developing a method with a high success rate.

TABLE I. SUCCESS RATES OBTAINED WITH A 10-FOLD CROSS-VALIDATION TEST OVER THE 1500 TRAINING IMAGES FOR EACH INDIVIDUAL CLASSIFIER AND EMOTION (FIRST FIVE ROWS) AND WHEN COMBINING THE FIVE CLASSIFIERS (SIXTH ROW).

	Disgust	**Joy**	**Anger**	**Fear**	**Sadness**	**Neutral**	**Surprise**
RIPPER	50.00%	85.70%	66.70%	48.10%	26.70%	80.00%	80.00%
SVM	76.50%	92.90%	55.60%	59.30%	40.00%	84.00%	82.20%
C4.5	58.80%	92.90%	66.70%	59.30%	30.00%	70.00%	73.30%
Naive Bayes	76.50%	85.70%	63.00%	85.20%	33.00%	86.00%	71.10%
Multilayer Perceptron	64.70%	92.90%	70.40%	63.00%	43.30%	86.00%	77.80%
Combination of classifiers	**94.12%**	**97.62%**	**81.48%**	**85.19%**	**66.67%**	**94.00%**	**95.56%**

The classifier combination chosen follows a weighted majority voting strategy. The voted weights are assigned depending on the performance of each classifier for each emotion. From each classifier, a confusion matrix formed by elements $P_{jk}(E_i)$, corresponding to the probability of having emotion i knowing that classifier j has detected emotion k, is obtained. The probability assigned to each emotion $P(E_i)$ is calculated as:

$$P(E_i) = \frac{P_{1k'}(E_i) + P_{2k''}(E_i) + \ldots + P_{5k^v}(E_i)}{5} \quad (1)$$

where: k', k'' ... k^v are the emotions detected by classifiers 1, 2 ... 5, respectively.

The assignment of the final output confidence value corresponding to each basic emotion is done following two steps:

1) Firstly, the confidence value $CV(E_i)$ is obtained by normalizing each $P(E_i)$ to a 0 through 1 scale:

$$CV(E_i) = \frac{P(E_i) - \min\{P(E_i)\}}{\max\{P(E_i)\} - \min\{P(E_i)\}} \quad (2)$$

where:

- $\min\{P(E_i)\}$ is the greatest $P(E_i)$ that can be obtained by combining the different $P_{jk}(E_i)$ verifying that $k \neq i$ for every classifier j. In other words, it is the highest probability that a given emotion can reach without ever being selected by any classifier.
- $\max\{P(E_i)\}$ is that obtained when combining the $P_{jk}(E_i)$ verifying that $k=i$ for every classifier j. In other words, it is the probability that obtains a given emotion when selected by all the classifiers unanimously.

2) Secondly, a rule is established over the obtained confidence values in order to eliminate emotional incompatibilities. The rule is based on the work of Plutchik [11], who assigned "emotional orientation" values to a series of affect words. For example, two similar terms (like "joyful" and "cheerful") have very close emotional orientation values while two antonymous words (like "joyful" and "sad") have very distant values, in which case Plutchik speaks of "emotional incompatibility". The rule to apply is the following: if emotional incompatibility is detected, i.e. two non-null incompatible emotions exist simultaneously, that chosen will be the one with the closer emotional orientation to the rest of the non-null detected emotions. For example, if "joy", "sadness" and "disgust" coexist, "joy" is assigned zero since "disgust" and "sadness" are emotionally closer according to Plutchik.

The results obtained when combining the scores of the five classifiers with a 10-fold cross-validation test are shown in Table II. As can be observed, the success rates for the "neutral", "joy", "disgust", "surprise", "disgust" and "fear" are very high (81.48%-97.62%). The lowest result is for "sadness", which is confused with the "neutral" emotion on 20% of occasions, due to the similarity of their facial expressions. Nevertheless, the results can be considered positive as emotions with distant "emotional orientation" values (such as "disgust" and "joy" or "neutral" and "surprise") are confused on less than 2.5% of occasions and incompatible emotions

(such as “sadness” and “joy” or “fear” and “anger”) are never confused.

TABLE II. CONFUSION MATRIX OBTAINED COMBINING THE FIVE CLASSIFIERS.

Emotion --> is classified as	Disgust	Joy	Anger	Fear	Sadness	Neutral	Surprise
Disgust	**94,12%**	0,00%	2,94%	2,94%	0,00%	0,00%	0,00%
Joy	2,38%	**97,62%**	0,00%	0,00%	0,00%	0,00%	0,00%
Anger	7,41%	0,00%	**81,48%**	0,00%	7,41%	3,70%	0,00%
Fear	3,70%	0,00%	0,00%	**85,19%**	3,70%	0,00%	7,41%
Sadness	6,67%	0,00%	6,67%	0,00%	**66,67%**	20,00%	0,00%
Neutral	0,00%	0,00%	2,00%	2,00%	2,00%	**94,00%**	0,00%
Surprise	0,00%	0,00%	0,00%	2,22%	0,00%	2,22%	**95,56%**

2) Emotional mapping to a 2D affective space

To enrich the emotional output information from the system in terms of intermediate emotions, one of the most influential evaluation-activation 2D models has been used: that proposed by Whissell. In her study, Whissell assigns a pair of values <evaluation, activation> to each of the approximately 9000 carefully selected affective words that make up her “Dictionary of Affect in Language” [10]. Fig. 2 shows the position of some of these words in the evaluation-activation space. The next step is to build an emotional mapping so that an expressional face image can be represented as a point on this plane whose coordinates characterize the emotion property of that face.

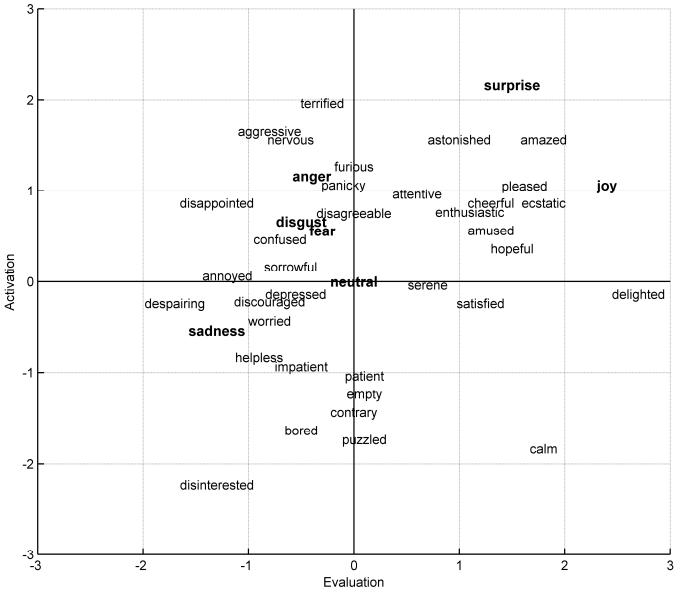

Figure 2. Simplified Whissell’s evaluation-activation space.

It can be seen that the words corresponding to each of Ekman’s six emotions have a specific location in the Whissell space (in bold in Fig. 2). Thanks to this, the output of the classifiers (confidence value of the facial expression to each emotional category) can be mapped in the space. This emotional mapping is carried out considering each of Ekman’s six basic emotions plus “neutral” as weighted points in the evaluation-activation space. The weights are assigned depending on the confidence value $CV(E_i)$ obtained for each emotion. The final coordinates of a given image are calculated as the centre of mass of the seven weighted points in the Whissell space. In this way, the output of the system is enriched with a larger number of intermediate emotional states. Fig. 3 shows several images of the database with their nearest label in the Whissell space after applying the emotional mapping.

Figure 3. Examples of images from the database with their nearest label in the Whissell space.

C. Evaluation with Human Assessment

The database used in this work provides images labeled with one Ekman universal emotions plus “neutral”, but there is no a-priori known information about their location in the Whissell 2D space. The main difficulty when working with a dimensional emotional approach comes from the labeling of ground-truth data since there is any available public facial expression database that provides emotional annotations in terms of evaluation and activation dimensions.

In order to evaluate the system results, it is necessary to establish the region where each image can be considered to be correctly located. For this purpose, a total of 43 persons participated in one or more evaluation sessions (50 images per session). In the sessions they were told to locate a set of images of the database in the Whissell space. As result, each one of the frames was located in terms of evaluation-activation by 16 different persons.

The collected evaluation data have been used to define an ellipsoidal region where each image is considered to be correctly located. The algorithm used to compute the shape of the region is based on Minimum Volume Ellipsoids (MVE). MVE looks for the ellipsoid with the smallest volume that covers a set of data points. Although there are several ways to compute the shape of a set of data points (e.g. using a convex hull, rectangle, etc.), we chose the MVE because of the fact that real-world data often exhibits a mixture of Gaussian distributions, which have equi-density contours in the shape of ellipsoids. The MVE is calculated following the algorithm described by Kumar and Yildrim [19]. The MVEs obtained are used for evaluating results at four different levels:

1) **Ellipse criteria.** If the point detected by the system (2D coordinates in the Whissell space) is inside the defined ellipse, it is considered a success; otherwise it is a failure.

2) **Quadrant criteria.** The output is considered to be correctly located if it is in the same quadrant of the Whissell space as the ellipse centre.

3) **Evaluation axis criteria.** The system output is a success if situated in the same semi-axis (positive or negative) of the evaluation axis as the ellipse centre. This information is especially useful for extracting the positive or negative polarity of the shown facial expression.

4) **Activation axis criteria.** The same criteria projected to the activation axis. This information is relevant for measuring whether the user is more or less likely to take an action under the emotional state.

The results obtained following the different evaluation strategies are presented in Table III.

TABLE III. RESULTS OBTAINED ACCORDING TO DIFFERENT EVALUATION CRITERIA.

	Ellipse criteria	Quadrant criteria	Evaluation axis criteria	Activation axis criteria
Success rate	73.73%	87.45%	94.12%	92.94%

As can be seen, the success rate is 73.73% in the most restrictive case, i.e. with ellipse criteria. It rises to 94.12% when considering the evaluation axis criteria. These results are promising especially when considering that, according to Bassili [20], a trained observer can correctly classify facial emotions with an average of 87%.

III. APPLICATION TO T-LEARNING: A TUTORING TOOL THAT CARES

This section details how the continuous facial affect sensing method presented in section II is applied to a t-learning tutoring tool in order to improve and humanize the tutor's feedback from the students. Section A describes the general architecture of the tutoring tool and section B explains how the emotional information extracted from each student is captured, processed and presented to the tutor.

A. T-learning Tutoring Tool's General Architecture

The tutoring tool has two main actors: the student and the tutor. The system's architecture can be explained from both actors' environments points of view (Fig. 4):

- **Student environment.** The student is located at home. He/she accesses a broadcasted t-learning interactive application through a set-top box (TELESystem TS7900HD), which has also IP communication capabilities. Besides showing the learning contents, the t-learning application is also able to control the management of an IP camera (LINKSYS WVC54GCA) via HTTP commands by accessing the set-top box middleware. This capability is exploited to record videos of the student which are stored in an external video server and further processed in the way described in section III.B to automatically extract emotional information about his affective state. The IP camera is detected by the set-top-box based on UPnP (Universal Plug and Play) service discovery [21]. The set-top box middleware manages the whole recording process as well as recorded video upload to the external video server.

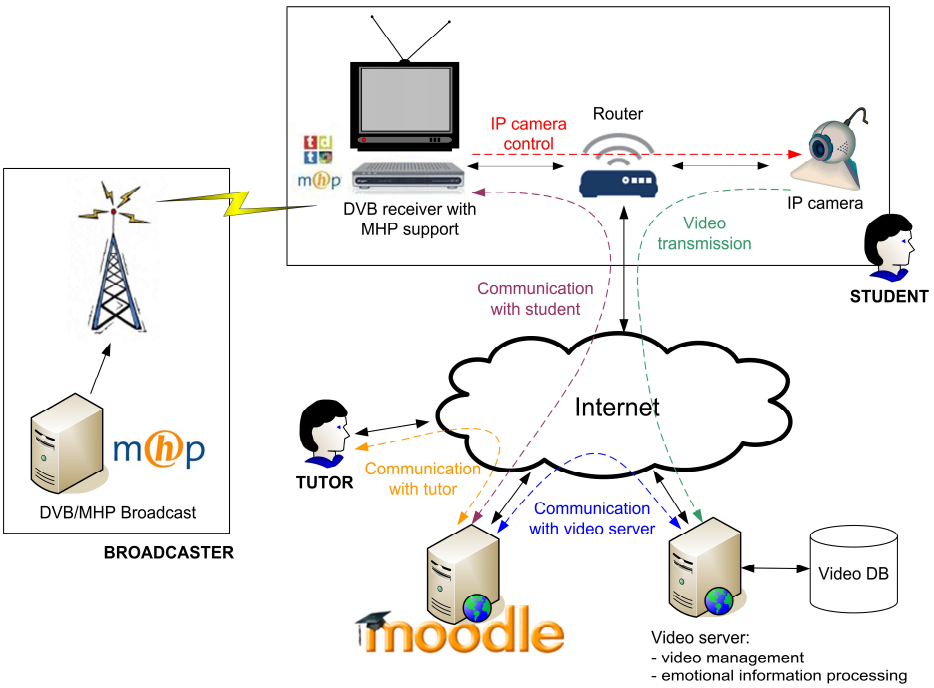

Figure 4. T-learning tutoring tool's general architecture

- **Tutor environment.** The Moodle Learning Management System [22] is used to keep the tutor in communication with the pupils. Moodle allows to create, store, organize, integrate and present educational contents, as well as sending communications to the students set-top boxes. Through Moodle, the tutor access student's information, in particular the extracted emotional information which is available in the video server.

This architecture provides the convergence between broadcast and broadband technologies. The interactive t-learning course is broadcasted to every user with the same contents. Depending on the student's evolution throughout the course (both academic and emotional), the tutor can send additional personalized contents and exercises by means of broadband communications. In that way, both a global delivery of the learning contents and personalized communications to every user are assured.

The broadcasted t-learning interactive application is based on DVB-MHP (Multimedia Home Platform, version 1.1.3) [23] digital TV interactive standard, as it is current digital TV interactive standard approved in the European Union. However, the system is enough open and scalable to be adapted to future new trends in interactive TV, since both the management of the camera and the communications with the tutor through the Moodle platform are IP-based. This is specially important taken into account that nowadays digital interactive television sector is growing and new interactivity technologies trends are arising.

B. Emotional Information Processing and Presentation to the Tutor

Tutors are often in charge of a large number of students across different courses. Therefore, keeping a close contact with each learner is difficult, especially when taking into account that t-leaning applications don't allow personal (i.e. human) contact. For that reason, it turns out interesting to automatically extract emotional information from the student and present it to the tutor in a simple and efficient way, so that problems during the learning process can be easily detected.

The t-learning interactive course consists of a set of modules, and each module is composed of several lessons and a final evaluation test to be performed within a limited time period. Fig. 5 shows a snapshot of a student interacting at home with the t-learning course during the final test. Before starting the final test of each module, the application proposes the student to be recorded while answering the evaluation questions. This recorded video, which is stored in the video server, carries useful affective information since it captures the student's facial expressions and can considerably help the tutor to adopt an appropriate pedagogical strategy (e.g. by offering help or extra contents to a student that has shown frustration during the test). It is important to emphasize the need of getting affective information from the students in the "heat of the moment" while they are feeling emotions and they clearly look so [8].

Figure 5. A student interacts at home with the t-learning course during the module assessment.

Humans inherently display facial emotions following a continuous temporal pattern [24]. With this starting postulate and thanks to the use of the 2-dimensional affect sensing method presented in Section II, which supports continuous emotional input, an affective facial video sequence can be viewed as a point (corresponding to the location of a particular affective state in time t) moving through the evaluation-activation space over time. In that way, each student's emotional record is automatically processed in the video server by applying, frame per frame, the facial affect recognition method and thus forming a continuous "emotional path" log at the output (Fig. 6). In order to provide temporal consistency to the extracted information, a Kalman filtering technique [25] is applied to both smooth the "emotional path" and improve the robustness of the method by predicting future locations (e.g. in cases of temporal facial occlusions or inaccurate tracking). Timestamps and information about the beginning and the end of each exercise of the test are also included in the log. Finally, the tutor can easily access every emotional log (sorted by student, course and module) through the Moodle platform.

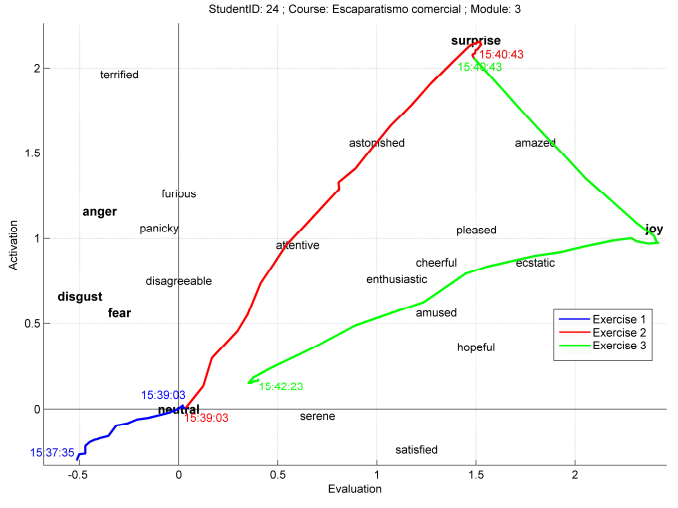

Figure 6. Example of emotional log that can be consulted by the tutor from the Moodle platform. Timestamps showing the beginning and the end of each exercise are included.

IV. Conclusions and Future Work

This paper describes the first t-learning tutoring system with Emotional Intelligence. The student interacts with the tutoring tool at home through a t-learning broadcasted interactive application. The system takes advantage of set-top boxes' broadband capabilities to both communicate the tutor and the student through the Moodle platform and record by means of an IP camera the facial expressions of the student while carrying out evaluation tasks. The obtained recorded videos are processed by applying an intelligent facial affect recognition technique, in order to extract emotional information about the learner that can easily be accessed by the tutor in a usable and simple way. The tool's architecture is enough open to be adapted to future IDTV standards to appear, since both the camera management and the communications with the Moodle platform are IP-based. The main novelty of the work comes from the fact of bringing Artificial Intelligence to t-learning in the form of Emotional Intelligence.

Regarding the facial affect sensing method itself, it combines in a novel manner five most commonly used Artificial Intelligence algorithms in the literature using a weighted majority voting strategy, obtaining at the output the location of the input facial expression in the evaluation-activation space. The final output of the system does not, therefore, simply provide a classification in terms of a set of emotionally discrete labels, but goes further by extending the emotional information over an infinite range of intermediate emotions. The main distinguishing feature of the method compared to others that use the evaluation-activation space for emotional classification is that the system output provides the exact location (coordinates) of facial expression in 2D space. Another noteworthy feature is that it has been tested with an extensive database of 1500 images showing individuals of different races and gender, giving universal results with very promising levels of correctness.

The recent focus on research area of Intelligent Tutoring Systems lies on "adaptive learning", i.e. building self-adaptive learning strategies depending on the learner's progress during the interaction with the tutoring tool. Adaptive learning provides contents and services to meet individual learning needs and thus improve learning achievement and efficiency. The key to successful adaptive learning is finding the multiple sources of personalization (e.g., learning orientation, academic background, etc.) that tells the system how to adapt appropriately. As pointed out in the introduction, the interpretation of student's affective state forms an indispensable part of traditional learning used by human tutors to establish pedagogical strategies with their pupils. For those reasons, we are currently considering to provide the tutoring tool with the ability of automatically analyzing the student's emotional logs in order to personalize the course contents and self-adapt the learning process to the affective state of each student (e.g. by offering help or decreasing the test's difficulty level when the learner is detected to get frustrated). It is also hoped to go beyond facial expression analysis by taken into account more potential sources of affective information (e.g. statistics such as time between responses in the test, number of successive hits, etc).

Acknowledgment

This work has been partly financed by the CETVI project (PAV-100000-2007-307) funded by the Spanish Ministry of Industry, the Grupo de Ingeniería Avanzada (GIA) of the Aragon Institute of Technology, the Spanish DGICYT contract number TIN2007-63025, the Government of Aragon IAF Nº2008/0574 and CyT Nº2008/0486 agreements.

References

[1] A. Dosi and B. Prario, "New frontiers of T-learning: the emergence of interactive digital broadcasting learning services in Europe," ED-Media, 2004, pp. 4831-4836

[2] M. Lytras, C. Lougos, P. Chozos and A. Pouloudi, "Interactive Television and e-learning convergence: examining the potential of t-learning," John Wiley & Sons, 2002.

[3] M. Damásio and C. Quico, "T-learning and interactive television edutainment: the Portuguese case study", ED-Media, 2004, pp. 4511-4518.

[4] R.W. Picard, "Affective computing," The MIT Press, 2000.

[5] P. Ekman, W.V. Friesen, and J.C. Hager, "Facial action coding system," Research Nexus eBook, 2002.

[6] H. Soyel and H. Demirel, "Facial expression recognition using 3D facial feature distances," Lecture Notes in Computer Science, vol. 4633, 2007, pp. 831–383.

[7] Z. Hammal, L. Couvreur, A., Caplier, and M. Rombaut, "Facial expression classification: an approach based on the fusion of facial deformations using the transferable belief model," International Journal of Approximate Reasoning, vol. 46, 2007, pp. 542–567.

[8] A. Kapoor, W. Burleson, and R.W. Picard, "Automatic prediction of frustration," International Journal of Human-Computer Studies, vol. 65, 2007, pp. 724–736.

[9] M. Yeasin, B. Bullot, and R. Sharma, "Recognition of facial expressions and measurement of levels of interest from video," IEEE Transactions on Multimedia, vol. 8, 2006, pp. 500–508.

[10] C.M. Whissell, "The dictionary of affect in language," Emotion: Theory, Research and Experience, vol. 4, The Measurement of Emotions, New York: Academic, 1989.

[11] R. Plutchik, "Emotion: a psychoevolutionary synthesis," New York: Harper & Row, 1980.

[12] N. Fragopanagos and J.G. Taylor, "Emotion recognition in human–computer interaction," Neural Networks, vol. 18, 2005, pp. 389–405.

[13] G. Garidakis, L. Malatesta, L. Kessous, N. Amir, A. Paouzaiou, and K. Karpouzis, "Modeling naturalistic affective states via facial and vocal expression recognition," Int. Conf. on Multimodal Interfaces, 2006, pp. 146–154.

[14] Face API technical specifications brochure. Available: http://www.seeingmachines.com/pdfs/brochures/faceAPI-Brochure.pdf

[15] F. Wallhoff, "Facial expressions and emotion database," Technische Universität München, 2006. Available: http://www.mmk.ei.tum.de/~waf/fgnet/feedtum.html

[16] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, "Web-based database for facial expression analysis," IEEE International Conference on Multimedia and Expo, 2005, pp. 317-321.

[17] M.A. Hall, "Correlation-based feature selection for machine learning," Hamilton, New Zealand, 1998.

[18] I. Witten and E. Frank, "Data Mining: practical machine learning tools and techniques," 2nd Edition, Morgan Kaufmann, San Francisco, 2005.

[19] P. Kumar and E.A. Yildirim, "Minimum-volume enclosing ellipsoids and core sets," Journal of Optimization Theory and Applications, vol. 126, 2005, pp. 1-21.

[20] J.N. Bassili, "Emotion recognition: the role of facial movement and the relative importance of upper and lower areas of the face," Journal of personality and social psychology, vol. 37, 1979, pp. 2049–2058.

[21] UPnP Forum. Available: www.upnp.org.

[22] Moodle - A Free, Open Source Course Management System for Online Learning. Available: http://www.moodle.org.

[23] MHP Standard Draft TS 102 812 V1.3.1 - MHP 1.1.3, 2007.

[24] S. Petridis, H. Gunes, S. Kaltwang, and M. Pantic, "Static vs. dynamic modeling of human nonverbal behavior from multiple cues and modalities," Proceedings of the International Conference on Multimodal Interfaces, 2009, pp. 23-30.

[25] R.E. Kalman, "A new approach to linear filtering and prediction problems," Transactions of the ASME - Journal of Basic Engineering, Series D, vol. 82, 1960, pp. 34-45.