Multimed Tools Appl (2015) 74:3183–3206
DOI 10.1007/s11042-013-1779-z

Affective-aware tutoring platform for interactive digital television

Sandra Baldassarri · Isabelle Hupont · David Abadía · Eva Cerezo

Published online: 6 December 2013

Abstract Interactive Digital TeleVision (IDTV) is emerging as a potentially important medium for learning at home. This paper presents a novel affective-aware tutoring platform for IDTV which makes use of automatic facial emotion recognition to improve the tutor-student relationship. The system goes further than simply broadcasting an interactive educational application by allowing the personalization of the course content. The tutor can easily access academic information relating to the students and also emotional information captured from learners' facial expressions. In this way, depending on their academic and affective progress, the tutor can send personal messages or extra educational contents for improving students' learning. In order to include these features it was necessary to address some important technical challenges derived from IDTV hardware and software restrictions. The system has been successfully tested with real students and tutors in a non-laboratory environment. Our system tries to advance in the challenge of providing to distance learning systems with the perceptual abilities of human teachers with the final aim of improving students learning experience and outcome. Nevertheless, there is still relatively little understanding of the impact of affect on students' behaviour and learning and of the dynamics of affect during learning with software. Systems like ours would make it possible to attack these relevant open questions.

Keywords Affective computing · Learning technologies · Facial recognition · Facial expressions · Emotions · Interactive digital tv

S. Baldassarri (✉) · E. Cerezo
GIGA-AffectiveLab, Computer Science Department, Engineering Research Institute of Aragon (I3A), Universidad of Zaragoza, Zaragoza, Spain
e-mail: sandra@unizar.es

E. Cerezo
e-mail: ecerezo@unizar.es

I. Hupont · D. Abadía
Multimedia Technologies Division, Aragon Institute of Technology, Zaragoza, Spain

I. Hupont
e-mail: ihupont@ita.es

D. Abadía
e-mail: dabadia@ita.es

1 Introduction

Nowadays, digital television is present in the daily lives of most Americans and Europeans. Until a few years ago, digital TV had been viewed as an emulation of analog TV, since the interaction is the same from the users' point of view: the television contents were emitted by broadcast and displayed in the TV. However, more recently the concept of interactivity has arisen: TV sets start to include Internet connection, allowing to create a direct communication with the user through the "return channel". This has opened the door to a new digital TV paradigm: Interactive Digital Television (IDTV).

The IDTV concept brings new challenges as well as new opportunities to the existing world of TV services. Interactive TV applications enable users to abandon their passive habits and to actively interact with the broadcast contents [40]. Within this context, IDTV is emerging as a potentially important medium for creating opportunities for learning at home. The convergence of interactive television and distance learning systems, called "t-learning" [24], gives viewers the opportunity of adapting and personalizing courses and contents according to their preferences [14]. In fact, experience has proved that interactive applications delivered through Digital TV must provide personalized information to the viewers in order to be perceived as a valuable service [43, 51].

Broadcasting interactive learning applications through the digital TV promises to open new pedagogical perspectives given the wide penetration of the medium, since about 98 % of European homes have at least one television set, whereas the penetration of Internet enabled computers is lower than 60 % [1]. This value is even less in countries in process of development [20]. Apart from wide-world usage, TV is considered by the viewer trustworthy in reference to broadcast content and easy to operate [58].

However, since IDTV is still in its beginnings, the t-learning field has to face many technical challenges [23, 47]. T-learning is not just an adaptation for IDTV of e-learning techniques used on the Internet [19]. It has its own distinctive characteristics, mostly related to the usability and technological constraints imposed by the television set and the return channel, operated through the set-top-box (STB). These constraints include having to use a simple remote control, which reduces the possibilities of interaction with the student, and the fact that STBs have lower computer power and connectivity capabilities than a personal computer. Moreover TV screens usually have lower resolution and colour management capabilities than a computer screen and are watched from a higher distance, which greatly conditions the design of user interfaces. For these reasons, most t-learning applications via IDTV have been more about edutainment than formal learning [6, 57]: learning contents are broadcasted and the interaction with the learner is limited to a simple pre-defined navigation within the application. More personalized t-learning interactive applications are needed to achieve more efficient learning, such as tutoring systems where the tutor can track the student's advances and personalize educational contents depending on each learner's progress.

Tutoring is a learning interactive and guided process that is continuously supervised with the aim of giving help in case the student is confused, or giving incentives if the student is frustrated [46]. The tutoring process is based in a double cycle of actions, in two complementary levels [69]. In the external level the global aspects of the process, as selection of the learning items, problems to propose, or learning assessment, are considered [5, 22]. Once the external level is determined, the internal level sets the steps to take inside each action of the external cycle according to the following action pattern: the tutor does a tutorial action, like an explanation or a suggestion; the student does an action as answer; and the tutor supplies feedback [56]. For offering an appropriate feedback, it is essential that the tutor can detect the emotional state of the student. Emotions are central to human experience, influencing cognition, perception and everyday tasks such as learning, communicating or even making rational decisions [53]. One of the big differences between an expert human teacher and a distance learning tutoring tool is that the former has the capability of recognizing and addressing the emotional state of learners and, based upon that observation, is able to take some action that positively impacts learning (e.g. by providing support to a frustrated learner who is likely otherwise to quit or increasing the level of difficulty in the exercises in which the student appears to be bored since the tests are too easy). Providing these kinds of perceptual abilities to distance tutoring systems would considerably benefit the learning process [3, 4, 54].

In this paper, we present an affective-aware tutoring platform for IDTV that goes further than simply broadcasting an interactive educational application by allowing the visualization of the affective state of the student and, therefore, improving the relationship between the distance tutor and the student. The tutor can access the students' academic information and, furthermore, extract emotional information from an analysis of their facial expressions. According to the students' behaviour, the tutor can send personal messages or extra educational contents for improving their learning. To guarantee both global delivery of the learning contents and personalized communications, the proposed platform has been developed under an IP-based open architecture that overcomes the technical limitations imposed by STBs. Finally, our system has been successfully tested with real students and tutors in a non-laboratory environment.

The structure of the paper is the following: Section 2 presents the related work, considering the three main areas involved in our work: recognizing user facial emotions, affective learning and adaptive learning. Section 3 presents the description of our system, giving special attention to the technical challenges that emerged from using an IDTV platform. In Section 4 the personalization capabilities considered in our system are exposed while Section 5 is focused on the system developed for emotion recognition. Section 6 presents an assessment carried out with end users and a discussion about these results. Finally, Section 7 comprises concluding remarks and a description of future work.

2 Related work

The last half-century of technological acceleration has yielded a massive incursion of digital technology into the learning environment, making dramatic differences to the practice of learning. In this overview of related work we will focus on the recognition of student's affective state from his/her facial expressions, on the challenges of considering affect in the learning process, and on the way of personalizing e-learning environments.

2.1 Recognizing user facial emotions

Facial expressions are the most powerful, natural and direct way used by humans to communicate affective states. Thus, the interpretation of facial expressions is the most extensively used method for determining user's emotions. For this reason, making a t-learning tutoring system able to interpret facial expressions would allow it to be affectively 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 [25] which include "happiness", "sadness", "fear", "anger", "disgust" and "surprise" [17, 67]. There have been a few tentative efforts to detect non-basic affective states, such as "fatigue", "interested", "thinking", "confused" or "frustrated" [39, 75]. 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 intrinsically ignores the intensity of emotions. To overcome the problems cited above, some researchers, such as Whissell [72], Plutchik [55] and Russell [60], prefer to view affective states as not independent of one another but rather as systematically related. They consider emotions as a continuous 2D space whose dimensions are evaluation/valence and activation/arousal. The evaluation/valence dimension measures how a human feels, from positive to negative. The activation/arousal dimension measures whether humans are more or less likely to take some action under the emotional state, from active to passive. This broader perspective of emotions has opened the door to the consideration of mixed cognitive/affective states relevant in learning contexts. Baker et al. situate their learning-centred user states in Russell's 2D framework: for example "boredom" has a negative valence and low level of arousal, "confusion" has negative valence and a moderate level of arousal, "frustration" has a high negative valence and a high level of arousal; and "delight" has a positive valence and a high level of arousal [5].

As shown, and compared to the categorical approach, dimensional representations are attractive because they provide a way of describing a wide range of emotional states and intensities. However, in comparison with category-based descriptions of affect, 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) [28] or a four class (2D space quadrants) classification [11], thereby losing the descriptive potential of 2D space.

For many years, substantial effort was dedicated to recognizing facial expressions in still images. Given that humans inherently display facial emotions following a continuous temporal pattern [52], more recently attention has been shifted towards sensing facial affect from video sequences. The study of the dynamics of facial expressions reinforces the limitations of the categorical approach, since it represents a discrete list of emotions with no real link between them and has no algebra: every emotion must be studied and recognized independently. The dimensional approach is much more able to deal with 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.

Besides the problem of how to correctly describe affect, one of the main drawbacks of existing affective analysers is related to the type of emotional information representation they provide as output. When a categorical approach is used, the great majority of studies usually show a histogram or pie chart representing the distribution-percentages or confidence values-of the studied emotional labels at each sampled time [26, 31]. On the other hand, most systems following a dimensional approach use to provide "activation vs. time" and "evaluation vs. time" graphs [49]. In some works, the affective analysis results are even limited to simple text logs, which turn out really difficult to interpret without any kind of graphical visualization [17]. Very few works propose more sophisticated, visual and intuitive emotional reports. For instance, McDuff et al. [45] present the result of a smile analysis to the user by means of emoticons and time graphs, comparing his/her smile track with an aggregate track. The systems developed in [36, 44] provide a continuous visual representation of the user affective evolution inside the evaluation-activation space. Another interesting example is the Affdex© commercial software [2], that allows to watch emotional information time graphs, together with the user facial expressions and the video contents used as stimuli.

2.2 Affective learning

Over the last few years there has been work towards incorporating assessments of the learner's affect into intelligent tutoring systems. Kort et al. proposed a comprehensive four-quadrant model that explicitly links learning and affective states [41]. This model was used in the MIT's Affective Learning Companion [10], able to mirror several student non-verbal behaviours believed to influence persuasion and liking. Sarrafdzeh et al. developed Easy with Eve, an affective tutoring system in the domain of primary school mathematics that detects students' emotions through facial analysis and adapts to students states [62]. De Vicente and Pain developed a system that could track several motivational and emotional states during a learning session with an ITS [21]. Conati developed a probabilistic system [15] that can reliably track multiple affective states (including joy and distress) of the learner during interactions with an educational game, and use these assessments to drive the behaviour of an intelligent pedagogical agent [16]. The Turing Research Group at the University of Memphis is adding an emotional component to their ITS AutoTutor, a natural language-based tutor successfully tested on about 1000 computer literacy and physics students.

Though there have already been multiple attempts to detect affect in learning environments, there is still relatively little understanding of the impact of affect on students' behaviour and learning [30]. The natural dynamics of affect during learning with software is similarly impoverished with few studies going on [5].

There is in fact a complex relationship between affect and cognition and the extension of cognitive theory to explain and exploit the role of affect is still in its infancy [54]. Research has demonstrated, for example, that a slight positive mood does not just make you feel a little better but also induces a different kind of thinking, characterised by a tendency toward greater creativity and flexibility in problem solving, as well as more efficiency and thoroughness in decision making [37]. The influences in cognition are not limited to positive mood, in general, it is not yet clear how harmful or persistent the affective states currently thought to be negative actually are. For example, confusion, a cognitive–affective state that is often perceived as being negative, has been shown to be positively correlated with learning [5] and even though some research has focused on reducing users' frustration [33], it is not clear that frustration is always a negative experience, at least in all types of human-computer interaction [29].

2.3 Adaptive learning

Snow and Farr suggested that sound learning theories are incomplete or unrealistic if they do not include a whole person view [65], integrating both cognitive and affective aspects, implying that no educational program can be successful without due attention to the personal learning needs of individual students. Russell suggested that educators should identify and acknowledge learning differences and make maximum use of the available technology to serve them accordingly [61].

When we talk about providing a personalized service in e-learning systems, there are two main research directions: adaptive educational systems and intelligent tutors (agents). Adaptive educational systems adapt the presentation of content and the navigation through these contents to a student's profile. This profile may comprise the student's learning style, knowledge, background, goals, among other features. Examples of these systems are: MLTutor [64], KBS-Hyperbook [32] and ELM–ART [8]. On the other hand, intelligent tutors talk to the student, recommend educational activities and deliver individual feedback according to the student's profile, which generally includes the student's knowledge or activities within the she is taking. Examples/she is taking. Examples of intelligent tutors in different domains are: Auto Tutor [18]; Easy with Eve [62]; eTeacher [63] or KERMIT [68].

Traditional teaching resources, such as textbooks, typically guide the learners to follow fixed sequences to other subject-related sections related to the current one during learning processes. Web-based instruction researchers have given considerable attention to flexible curriculum sequencing control to provide adaptable, personalized learning programs [50]. Curriculum sequencing aims to provide an optimal learning path to individual learners since every learner has different prior background knowledge, preferences and often various learning goals [9]. In an e-learning adaptive system, an optimal learning path aims to maximize a combination of the learner's understanding of courseware and the efficiency of learning the courseware [59]. Moreover, as numerous e-learning systems have been developed, a great quantity of hypermedia in courseware has created information and cognitive overload and disorientation, such that learners are unable to learn very efficiently. To aid more efficient learning, many powerful personalized/adaptive guidance mechanisms, such as adaptive presentation, adaptive navigation support, curriculum sequencing, and intelligent analysis of student's solutions, have been proposed [12].

In the DTV domain, efforts have also been made to personalize information. Lopez-Nores et al. [43] explore the possibilities of running downsized semantic reasoning processes in the DTV receivers supported by a pre-selection of material driven by audience stereotypes. Rey-Lopez et al. [58] are working in building a model for personalised learning using Adaptive Hypermedia techniques and Semantic Reasoning. Bellotti et al. [7] use a template-based approach to configure flexible courses through a personalization manager that keeps track of the dynamic user profiles and of the persistent profiles.

In next sections we will present a t-learning tutoring system in which the detection of user's affective state opens the door to new adaptive learning strategies to be performed by the human tutor in charge of the course. Even though Soleymani and Pantic [66] have already foreseen the importance of a TV aware of viewers' emotions, to our knowledge, there is no t-learning tutoring system in the literature with the ability of intelligently recognizing affective cues from the student.

3 System description and technical challenges

The general architecture of the platform as well as the communication flows between its different components is shown in Fig. 1. As it can be seen, the system has three main actors: the student, the tutor and the contents creator. All the actors are linked with each other through Moodle (Modular Object Oriented Developmental Learning Environment), a free open source Learning Management System that helps educators to create effective on-line learning communities. Moodle is in charge of storing all the course materials, such as the learning contents, exercises, glossary, tests, bibliography, etc., and also contains a description of the course organization, i.e. number of modules, contents associated to each module and navigation rules between the course materials. It also centralizes, manages and keeps synchronized all the information related to the communication between actors and between the actors and the platform. This section provides an overview of the system, by describing the role of each actor in the platform and the way they access and interact with it, and the technical challenges faced to bring a learning system, an IDTV application and affective computing technology together.

3.1 System overview

The tutoring tool has three main actors: the student and the tutor, who communicate with each other through the application, and the contents creator. Each actor accesses the system through different interfaces and has a different role. There follows a brief description of the role of each actor.

The student is located at home (see Fig. 2a) and accesses a broadcast t-learning interactive MHP (Multimedia Home Platform) application through a set-top box, by logging in with his/

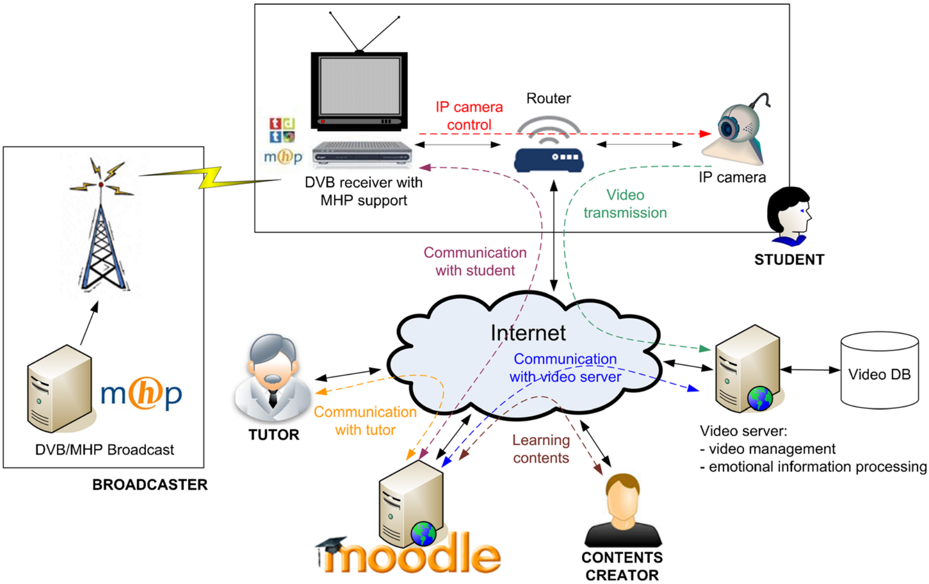


Fig. 1 T-EDUCO general architecture and communication flows

her personal student ID and password. This MHP student-side application connects to Moodle to obtain the course description and materials, and is in charge of displaying them in an appropriate and usable format especially developed for a TV device (e.g. to enable navigation through TV's remote control coloured buttons). It also offers a personalized communication between the student and the tutor through emails and videos. The application is able to control the management of an IP camera that records videos of the student that are then processed to automatically extract facial emotional information about his/her affective state (as it is explained in Section 3.2.1).

The tutor (see Fig. 2b) interacts with T-EDUCO directly via Moodle's web interface. After performing some necessary adaptations in Moodle's source code, T-EDUCO has also been provided with the capability of both sending communications to the students' set-top boxes and receiving information from the interactive t-learning application via IP. In this way, T-EDUCO keeps the t-learning application and Moodle synchronized. This allows the tutor to access the academic and emotional information of each student and to send personal messages and contents, as will be further detailed in this paper.

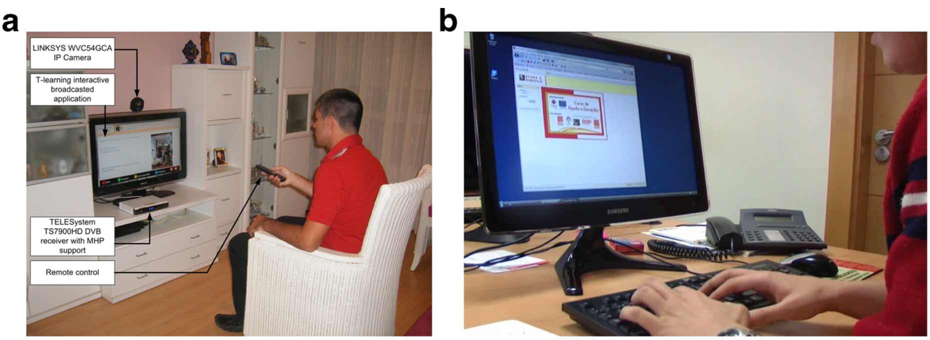


Fig. 2 Typical actors' environment. **a** Student. **b** Tutor

The contents creator, also accessing the system directly via Moodle's web interface, is responsible for uploading all the initial material required for the development of each course: slides, notes, exercises, references and bibliography, self-assessments, tests, glossary, etc. These contents can be later modified or changed by the tutor depending on each student's individual needs and progress, as will be explained in the next section.

3.2 Technical challenges

In other platforms, like Smartphone, tablets, PCs, etc., there are no technical challenges on setting an architecture for our tool, since they use to have embedded or integrated cameras, connectivity facilities, answer bandwidth, etc. However, IDTV, presents till now, specific technical challenges that are tackled in this section. Although nowadays new TVs with embedded cameras are being introduced in the market, they are not easily accessed yet.

Our platform includes novel features, not yet exploited in any existing IDTV application, that provide substantial added value to t-learning services: communication between users through video messages, the distribution of dynamic and personalized contents for each participant, and the detection of emotions. These new features provide t-learning systems, innovative in themselves, with a higher pedagogical value. These features have not previously been considered in IDTV systems because of important technical restrictions, both in hardware and software:

- Hardware restrictions:

– Commercial STBs have very restricted hardware capacities in terms of memory, connectivity and processor level. This implies an inability to run complex data processing algorithms, such as emotion recognition algorithms, or to connect any peripheral appliances such as cameras. It also makes it impossible to store advanced and heavy multimedia material (images, video, etc.) in local memory for reproduction in IDTV applications.
– In IDTV all the interaction is performed through a hardware device: the remote control. Its interactive capacities are very limited, and therefore browsing through learning contents has always been too linear and the activities offered by the t-learning applications very restricted.

- Software restrictions:

– Until now, IDTV standards –such as MHP, OCAP or HbbTV- have lacked several communication capacities (such as Bluetooth) and have not allowed most multimedia formats (audio, video) to be reproduced.

In this section we explain how we faced and overcame the restrictions imposed by STBs and IDTV in order to provide the system with the required performance.

3.2.1 Video capture and display

The capture of students' facial expressions and the recording of video-messages require the control of a webcam from the MHP learning application: the start/end of the recording must be commanded, the video flow must be stored, etc. However, current IDTV standards do not allow direct control of the camera and STB storage capacity does not allow a large number of videos or long length videos to be stored in the memory.

Therefore, in order to solve these technical challenges, we opted for a camera with the following features:

- IP Connection. The camera has an IP connection, so that it can be turned on/off from the t-learning application through HTTP commands.
- Video streaming. The camera allows the video stream that is being recorded to be sent to a static IP address in ASF format. In this way, the video recorded by the user can be directly sent, without going through the STB, to an external video server in which it will be stored and/or processed in order to extract the emotional information of the student.
- Universal Plug and Play (UPnP). The camera implements a UPnP discovering protocol so that it emits signals to the net informing about its presence. The STB, through a *UPnP client* library adapted to be executed over MHP, listens to these UPnP messages receiving information from the camera, such as its IP address.
- Affordable price and commercially available.

The visualization of the video messages received from the tutor by the student in the STB raises another technological challenge: the standard MHP does not support the direct reproduction of videos stored in the server. The standard only allows the visualization of MPEG-2 videodrips [70], a high-compression video format that codifies the frames as predictive (p-frames) and interpolated (i-frames) and without audio support.

This problem has been resolved by converting the video stream that comes from the IP camera in the external server, recoding separately the image component in videodrips and the audio component in MP2, which is a format compatible with MHP. Once the video is converted, the t-learning application is able to download the video and audio streams from the external server, synchronize and reproduce them.

3.2.2 Platform adaptability

As pointed out previously, the developed broadcast t-learning interactive application is based on the MHP (Multimedia Home Platform, version 1.1.3) digital TV interactive standard. The choice of the MHP standard is justified by two principal factors. First, for the development of our applications, we have access to an equipped online laboratory of interactive digital television with a MHP playout and an STB farm provided with different features, both hardware and software (different MHP versions). This STB farm brings together most of the MHP decoders that can be found in the European market. Broadcast emission of MHP applications can be emulated and tested online in any available STB in our laboratory environment, accessing the laboratory via a web site. More information about the laboratory environment can be found in [1]. The second reason for choosing the MHP standard is that most Spanish broadcasters only allow MHP compliant interactive applications to be coded and emitted. This is the case of the Spanish regional television service used for a pilot broadcast emission of the t-learning application that made it possible to assess our application on a large scale with real users.

However, MHP is not the only standard that exists nowadays for IDTV. There are other standards such as OCAP or, the most recent, HbbTV that are well-established depending on the geographic location. In any case, all the IDTV solutions have something in common, that is, their capacity for making broadcast and broadband technologies converge. This is exactly where the main strength of our platform lies: the system is open and scalable enough to be adapted to any other standards or future new trends in interactive TV, since both the management of the camera and the communications with the Moodle platform are IP-based. This is especially important taking into account that nowadays the digital interactive television sector is growing, interactivity trends are constantly changing and the future of the IDTV is still not clear: will it converge to a single solution or, on the contrary, will more solutions emerge, either standard or proprietary.

4 Personalization capabilities

This section presents the personalization capabilities of the tool that allow the tutor to adapt the learning contents depending on each student's personal evolution (both academic and emotional) throughout the course.

Distance learning tutors, like traditional classroom teachers, should not limit their role to passing or failing students solely on the basis of marks obtained in final assessments. Tutors should be pedagogues, following the progress of students throughout the course and facilitating their learning. They should be aware of difficulties encountered by students (noticing if they appear to be frustrated, confused, etc.) or, on the other hand, of any signs of boredom resulting from the course contents being below their level and the need to make faster progress. Tutors need to develop strategies to benefit and foster progress. In classroom teaching, these issues are addressed intuitively. However, it is not so easy to follow this type of continuous monitoring in a distance learning application. Most such applications tend to ignore these human factors. In our system, the learning contents can be adapted taking into account the academic and affective tracking of the student, personalizing the contents and messages.

4.1 Academic follow-up

A T-EDUCO t-learning interactive course consists of a set of modules. Each module is composed of several lessons and a final evaluation test. Since the t-learning application and Moodle are synchronized, the distance tutor can keep detailed logs of all the activities the students perform: what materials have been accessed, which lessons accomplished, which marks have been obtained in the final tests, which questions have been incorrectly answered… All these options can be filtered by course, pupil, date or module (see Fig. 3).

4.2 Affective follow-up

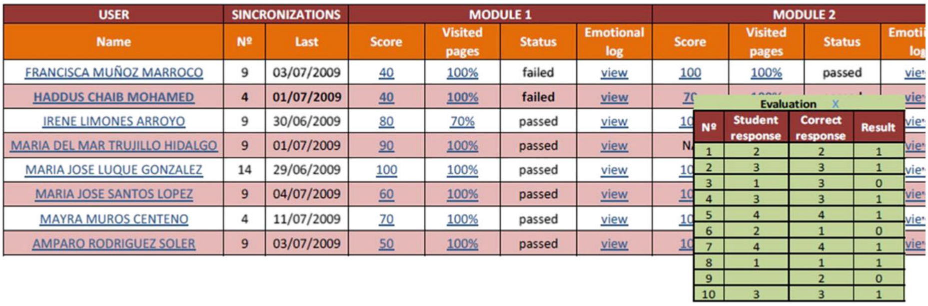
Tutors are often in charge of a large number of students across different courses. Therefore, keeping close contact with each learner is difficult, especially because t-leaning applications usually do not allow personal contact. For this reason, it would be interesting to be able to automatically extract emotional information from the student and to present it to the tutor in a simple and efficient way, so that problems during the learning process may be easily detected.

T-EDUCO has the capability of automatically extracting affective information by analyzing the learner's video sequences captured by an IP camera. Before starting the final test of each module, the application suggests that the student be recorded by the camera while answering the evaluation questions. This recorded video carries useful affective information that is extracted and presented in the form of emotional logs (detailed in Section 5). These logs are made accessible to the tutor through Moodle (7th column in Fig. 3).

4.3 Other personalization capabilities

T-EDUCO allows the student and the tutor to interchange personal communications in the form of e-mail messages and video messages.

The learner can write/access e-mail messages through the t-learning application. For writing the mails, students are provided with a virtual keyboard that can be handled by means of the TV remote control. They can also record a video message and attach it to the mail by simply clicking a "record video" button. If desired, the recorded video can be watched, re-recorded or



USER	SINCRONIZATIONS		MODULE 1				MODULE 2			
Name	Nº	Last	Score	Visited pages	Status	Emotional log	Score	Visited pages	Status	Emoti log
FRANCISCA MUÑOZ MARROCO	9	03/07/2009	40	100%	failed	view	100	100%	passed	vie
HADDUS CHAIB MOHAMED	4	01/07/2009	40	100%	failed	view	7(			vie
IRENE LIMONES ARROYO	9	30/06/2009	80	70%	passed	view	10			vie
MARIA DEL MAR TRUJILLO HIDALGO	9	01/07/2009	90	100%	passed	view	N			vie
MARIA JOSE LUQUE GONZALEZ	14	29/06/2009	100	100%	passed	view	10			vie
MARIA JOSE SANTOS LOPEZ	9	04/07/2009	60	100%	passed	view	10			vie
MAYRA MUROS CENTENO	4	11/07/2009	70	100%	passed	view	10			vie
AMPARO RODRIGUEZ SOLER	9	03/07/2009	50	100%	passed	view	10			vie

Evaluation X

Nº	Student response	Correct response	Result
1	2	2	1
2	3	3	1
3	1	3	0
4	3	3	1
5	4	4	1
6	2	1	0
7	4	4	1
8	1	1	1
9		2	0
10	3	3	1

Fig. 3 Example of student information that the tutor can access through the Moodle platform

removed by the student before sending the message. At the other end, the tutor can read and write e-mails via the Moodle platform, in the same way as in most webmail applications.

The interactive t-learning course is broadcast to every user with the same initial contents. Throughout the course, the academic and the affective follow-up offered by T-Educo allow the tutor the possibility of sending additional personalized learning contents, such as extra modules and/or evaluation tests, according to the student's academic progress and emotional requirements. These are made available to the student in the t-learning application. Moodle's modular design makes it easy for the tutor to create new modules and tests that will engage learners. The extra contents are sent to the student's set-top-box in the SCORM (Sharable Content Object Reference Model) standard format, a standard specification that allows the development of structured pedagogical contents that can be imported within different learning applications.

5 Recognizing learners' emotions

A distinguishing factor of T-EDUCO is its capability of extracting emotional information from the learners' recorded video sequences. This makes it possible for the tutor to better analyze student behavior and to detect boredom, frustration or dejection at an early stage. Moreover, it gives tutors the possibility of establishing a more humanized relationship with their students.

5.1 Facial expression recognizer

The video sequences recorded from the learners are analyzed, frame by frame, by a 2D continuous emotional images recognition system [34]. The initial inputs of our system are a set of distances and angles obtained from 20 characteristic facial points (eyebrows, mouth and eyes). 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 [31, 67]. The points are obtained thanks to faceAPI [27], a commercial real-time facial feature tracking program that provides Cartesian facial 3D coordinates.

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. This output is subsequently expanded in order to be able to work in a continuous emotional space and thus to consider intermediate emotional states.

In order to select the best classifiers to achieve discrete emotional classification, the Weka tool was used [74]. 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. 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.

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 the psychologist Cynthia Whissell. The “Dictionary of Affect in Language” (DAL) [72] developed by Whissell is a tool to quantify the emotional meanings of words. Words included in the DAL were selected in an ecologically valid manner from an initial corpus of more than one million words [73]. Whissell assigns a pair of values<evaluation; activation>to each word of the DAL. The evaluation dimension goes from positive to negative whereas the activation dimension ranges from active to passive. The emotion-related words corresponding to each of Ekman's emotions have a specific location (x_i; y_i) in the Whissell space. Figure 4 shows the position of some of these words in the evaluation-activation space.

Based on this, the output of the classifiers (confidence value of the facial expression to each emotional category) can be mapped onto the dimensional space. This emotional mapping is carried out considering each of Ekman’s six basic emotions plus “neutral” as weighted points

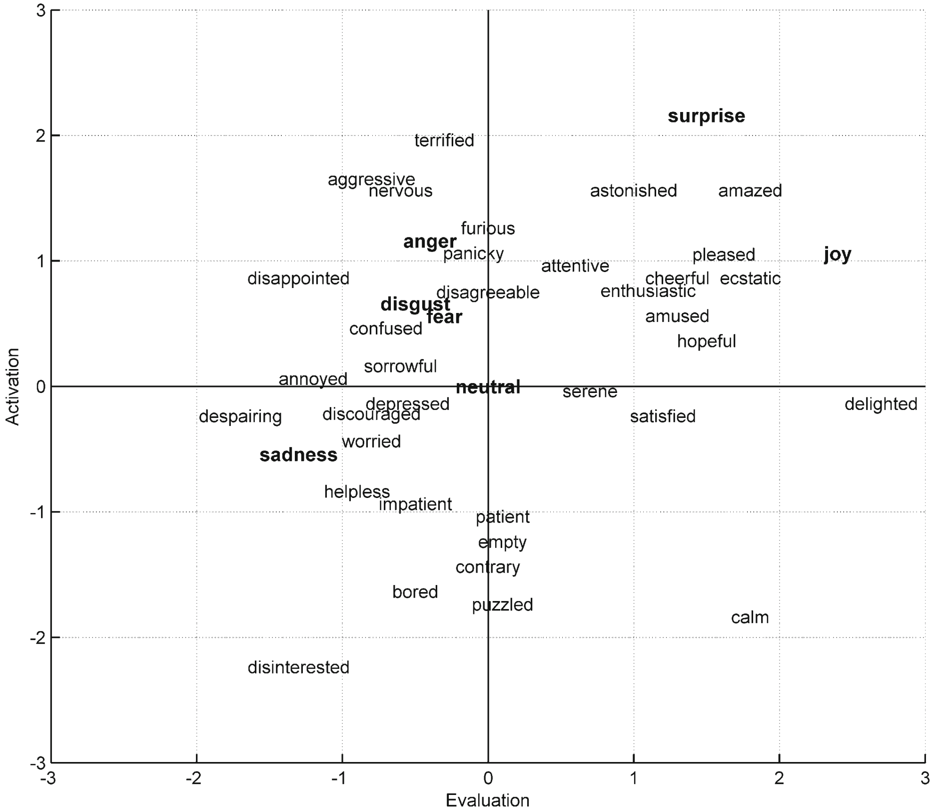


Fig. 4 Simplified Whissell’s evaluation-activation space

in the evaluation-activation space. The weights are assigned depending on the confidence value $CV(E_i)$ obtained for each emotion. The final evaluation and activation (x; y) coordinates of a given image are calculated as the centre of mass of the seven weighted points in the Whissell space, as follows:

$$x = \frac{\sum_{i=1}^{7} x_i CV(E_i)}{\sum_{i=1}^{7} CV(E_i)} \quad and \quad y = \frac{\sum_{i=1}^{7} y_i CV(E_i)}{\sum_{i=1}^{7} CV(E_i)}$$

Figure 5 explains graphically the emotional mapping process of a facial image in the Whissell space. In this way, the output of the system is enriched with a larger number of intermediate emotional states.

The database used to train the system is an extensive universal (with males and females of different ages and ethnicities) one of 1500 images [71]. Figure 6 shows some of the images of the database located in the Whissell space thanks to the coordinates calculated by our emotional recognizer.

The database provides the images labelled with one of Ekman's universal emotions plus "neutral", but there is no a-priori known information about their location in the Whissell 2D space. In fact, the main difficulty when working with a dimensional emotional approach comes

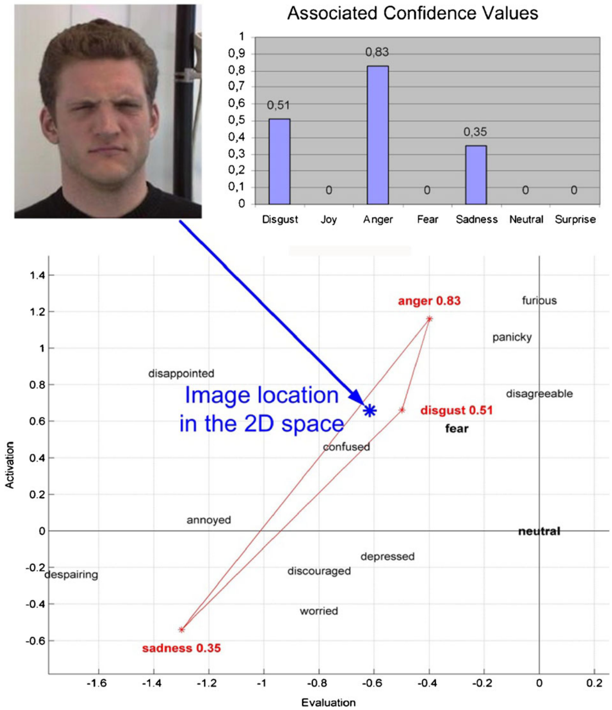


Fig. 5 Example of emotional mapping. The facial images discrete classification method assigns to the facial expression shown a confidence value of 0.83 to "anger", 0.51 to "disgust" and 0.35 to "sadness" (the rest of Ekman's emotions are assigned zero weight). Its location in the 2D space is calculated as the center of mass of those weighted points (blue asterisk)

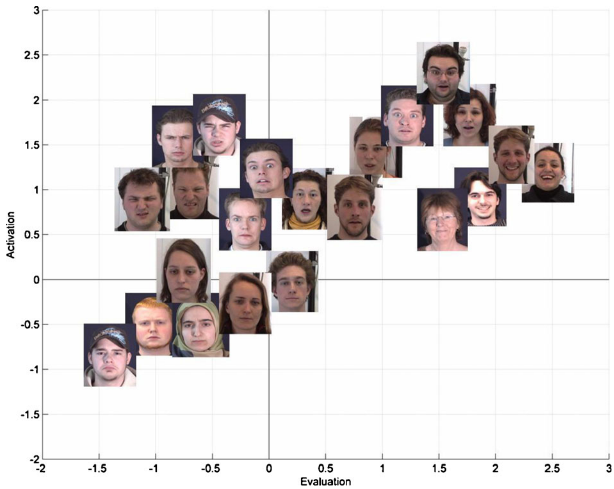


Fig. 6 Some images located in different points of the 2D continuous space after applying our emotional recognition system

from the labelling of ground-truth data. For this reason, human assessment has used to validate the system. The proposed model has been tested with a set of complex video sequences recorded in an unsupervised setting (VGA webcam quality, different emotions displayed contiguously, facial occlusions, etc.). A total of 15 videos from 3 different users were tested, ranging from 20 to 70 s, from which a total of 127 key-frames were extracted to evaluate different key-points. These key-points were annotated in the Whissell space thanks to 18 volunteers. The collected evaluation data have been used to define a 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) and follows the algorithm described by Kumar and Yildrim [42]. MVE looks for the ellipsoid with the smallest volume that covers a set of data points.

The obtained MVEs are used for evaluating results at four different levels, as shown in Table 1. As can be seen, the success rate is 61.90 % in the most restrictive case, i.e. with ellipse criteria and rises to 84.92 % when considering the activation axis criteria.

Once affect is detected and intuitive and meaningful visualization is essential, as pointed out in the introduction. Our facial emotion recognizer outputs simple but effective emotional logs as it will be shown in next section.

5.2 Building emotional logs

Humans inherently display facial emotions following a continuous temporal pattern. With this starting postulate and thanks to the use of the 2-dimensional affect sensing method, the

Table 1 Results obtained in an uncontrolled environment

	Ellipse criteria (success if inside the ellipse)	Quadrant criteria (success if in the same quadrant as the ellipse centre)	Evaluation axis criteria (success if in the same evaluation semi-axis as the ellipse centre)	Activation axis criteria (success if in the same activation semi-axis as the ellipse centre)
Success%	61.90 %	74.60 %	79.37 %	84.92 %

recorded student's affective facial video can be viewed as a point (corresponding to the location of a particular affective state in time t_k) moving through the evaluation-activation space over time. In this way, the different positions taken by the point (one per frame of the video sequence) and its velocity over time can be related mathematically and modelled, finally obtaining an "emotional path" in the 2D space that reflects intuitively the emotional progress of the student during the interaction.

A Kalman filtering technique is proposed to model the "emotional kinematics" of that point when moving through the 2D evaluation-activation space and thus enabling its trajectory to be smoothed and kept under control [35]. Kalman filters are widely used in the literature for estimation problems ranging from target tracking to a function's approximation [48]. Their purpose is to estimate a system's state by combining an inexact (noisy) forecast with an inexact measurement of that state, so that the most weight is given to the value with the least uncertainty at each time t_k.

Analogously to classical mechanics, the "emotional kinematics" of the point in the 2D continuous space (x-position, y-position, x-velocity and y-velocity) are modelled as the system's state in the Kalman framework at time t_k. The output of the 2D emotional classification system is modelled as the measurement of the system's state. In this way, the Kalman iterative estimation process-that follows the well-known recursive equations detailed in Kalman's work [38]-can be applied to the recorded student's emotional video sequence, so that each iteration corresponds to a new video frame (i.e. to a new sample of the computed emotional path). For the algorithm initialization at t_0, the predicted initial condition is set equal to the measured initial state and the 2D point is assumed to have null velocity.

One of the main advantages of using the Kalman filter is that it can tolerate small occlusions or inaccurate tracking. In general, existing facial trackers do not provide highly accurate levels of detection: most are limited in terms of occlusions, fast movements, large head rotations, lighting, beards, glasses, etc. Although the tracker used, faceAPI, deals with these problems quite robustly in laboratory environments, tracking videos taken in "real" settings, such as students' homes, have lower performance. For this reason, tracker measurements include a confidence weighting from 0 to 1; in our case, when a low level of confidence is detected (lower than 0.5), the measurement is not used and only the filter prediction will be taken as the next 2D point position.

Figure 7 summarizes the complete processing of the emotional video sequences. The final emotional log presented to the tutor is a continuous "emotional path" that also includes timestamps and information about the emotional state of the student while he/she is doing the exercises of each module. For facilitating the understanding, the emotional path of each exercise is drawn with a different colour. The tutor can easily access every emotional log through the Moodle platform, clicking in the "view" label of the emotional log column of each module (see Fig. 3) and consequently take pedagogical decisions.

However, an automatic analysis of this information had to take into account, for each module, all the aspects involved in Fig. 3, such as: emotional log, status (failed or passed), score obtained, time spent doing each exercise, amount or percentage of visited pages, etc.

6 Evaluation: pilot study

After extensive technical testing of our platform in a laboratory environment, a final assessment of the t-learning application was performed with end users and with real broadcast emissions.

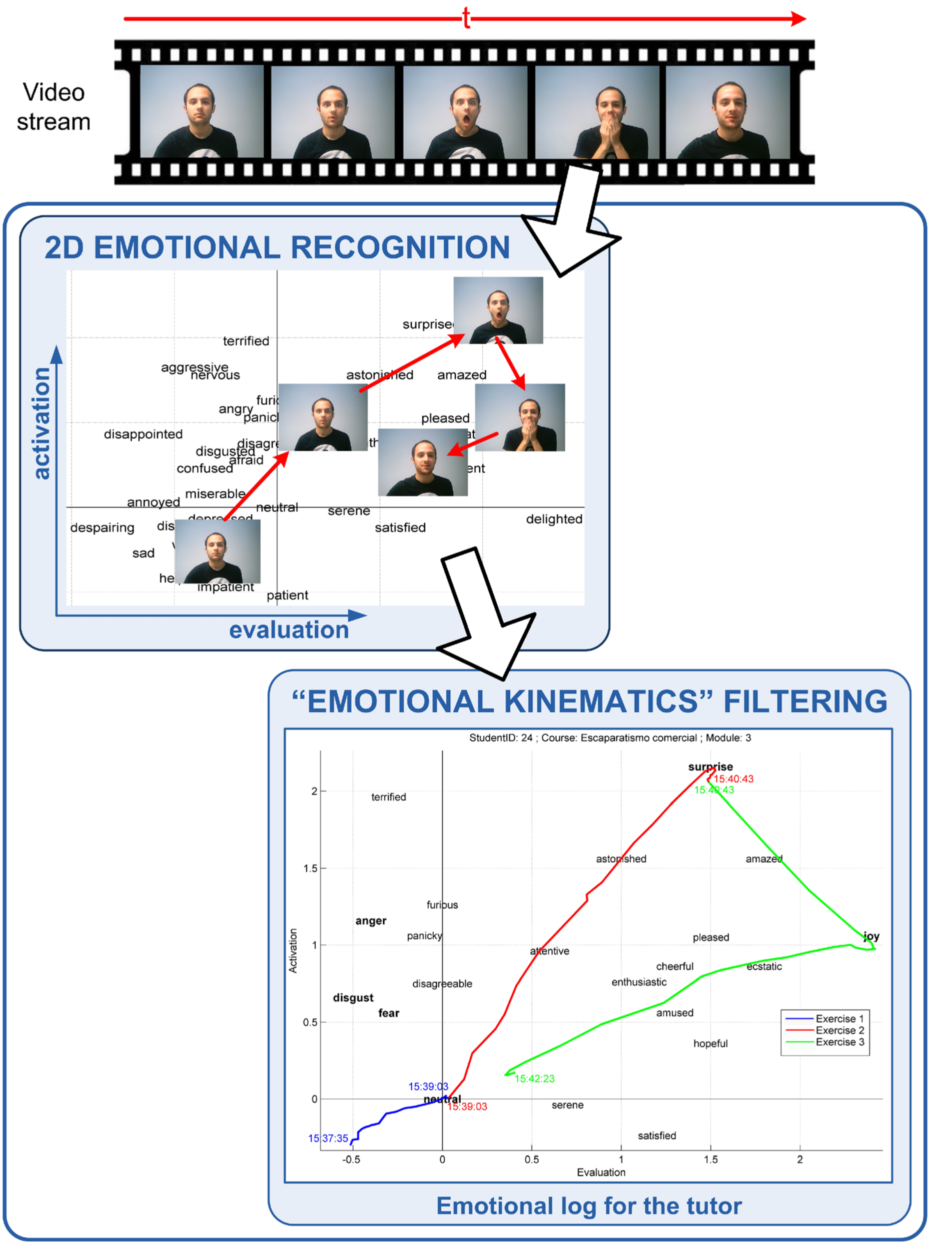


Fig. 7 Emotional video sequences processing

6.1 Hypothesis and set up

A pilot study was carried out with the aim of studying the impact of IDTV in the student-learning content interaction and of the affective information in in the tutor-students interaction. Our hypothesis was that the addition of affective information to the academic one and the personalization of contents would allow a more a humanized academic follow-up, leading to an improvement in student learning.

The pilot study was emitted in Seville (Spain) by RTVA (*Radio Televisión de Andalucía*), through the TV channel "*Canal Sur*" where the t-learning course tested could be accessed 24 h a day and 7 days per week, from the 5th of June to the 30th of October 2012. The course contents had been given to the TV operator 10 days before, for doing the corresponding tests in order to solve possible disruptions without altering the development of the experiment and the TV emission.

The case study was carried out with the contents of a "Home Assistance" course and with 30 students, 8 males and 22 females, with ages ranging from 32 to 47, all of them active workers in the health sector. The students accessed the t-learning application from their homes, and were provided with MHP-compliant STBs. Before beginning the experience, there was a first face-to-face session with the students for training them in the STB use, the contents and other technical aspects. In that session the STBs were given and instructions (user manuals) and technical support resources were facilitated in order to solve possible doubts when the equipments were set up in the students' homes. This first session included a demo of installation, connection and configuration of the equipment, necessary for the use and synchronization between the students and the tutor application. It was mandatory for the students to attend all the different modules and practical exercises that comprised the 28 h course. They also had to use the different utilities of the course, such as the video message communication and the access to the glossary.

The pilot test also included the participation of 5 home assistance service professionals, 3 females and 2 males, with ages ranging from 43 to 51, that had the role of tutors. They were in charge of monitoring the students through Moodle, using all the facilities offered by the platform (generation of personalized contents for each student, video messages, statistical queries and emotional logs, etc.).

In the first session, as much tutors as students received some questionnaires that had to be given later, in a final course session in which a group meeting was carried out to monitor and assess the course. During the experience, three students (women) gave up because they didn't have enough time to follow the course. Therefore, 30 students began the course, but only 27 finished it. The assessment of the experience presented in the following section responds to the data provided by 27 students.

6.2 Results and discussion

Table 2 gathers the final results of the enquiries put to students and tutors after finishing the t-learning course (answers had to be marked between 0 and 10, in a Likert scale). As interaction is one of the great challenges of IDTV, special attention was paid to analyzing interaction factors, between: Student and Tutor, Student and Learning Contents, and Tutor and Student.

The mean mark of the evaluation of our platform slightly exceeds 7 and there are no indicators under 5. Especially good scores were obtained from students for those factors that distinguish our application from other t-learning applications: personalized contents and video messages. Moreover, IDTV is perceived as a good tool for the autonomous work

Table 2 Results of the enquiries put to students and tutors after finishing the t-learning course

Factor	Min	Max	Mean
Student→tutor interaction			
The student can ask for help through the application's features.	5	10	8.40
The interaction between student and tutor with the application is simple and intuitive.	2	10	6.00
The video messages facilitate the interaction between student and tutor.	6	10	8.70
The personalized contents sent by the tutor improve the student's learning.	5	10	7.67
Student-learning contents interaction			
The IDTV allows interaction with the learning contents.	5	10	7.00
The IDTV allows the students to study following their own scheduling requirements.	5	10	8.00
The IDTV makes it easier to carry out the interactive activities.	5	10	7.00
Tutor→student interaction			
The interaction between student and tutor with the application is simple and intuitive.	5	10	8.70
The information provided about the students' progress is enough for a good and humanized academic follow up.	5	10	7.80
The video messages facilitate the interaction between student and tutor.	7	10	8.80
The personalized contents sent by the tutor improve the student learning.	6	10	7.80

of the student, and good scores were obtained for those issues related with the interaction with the learning contents. Regarding the tutors, all the scores are relatively high; above all, the possibility offered by the T-EDUCO platform of humanizing distance education is especially well-considered. These results confirm our initial hypothesis about the improvement in the interaction of the students with learning contents and tutors and in students' learning.

Apart from the numerical results exposed before, it is interesting to analyze the comments that emerged from the participants during the final group meeting. Regarding the general evaluation of the course, the group agreed that the experience was very interesting and enriching, that they liked the course, and that they felt comfortable with this t-learning system, accessing from their homes. In this sense, 81.5 % of the students confirmed that they would do a new course with this system, and this information coincides with 88.9 % that thought that the course met their expectations.

About the advantages and disadvantages that this learning modality presents, the students highlighted the following aspects:

Advantages:

- The course contents are interesting and complete and the course is useful and enjoyable ("*I learnt the same than a face to face course*")
- The opportunity to do it at home, without moving, and with a flexible schedule ("*I like to do the course at home and with my learning speed*", "*I make something to drink and I begin to do the course comfortably*")
- The tutor is close, to motivate and to correct the mistakes ("*When I read the assessment, after doing each exercise, I felt very happy*", "*I like to receive messages, help and a personal assessment from the tutor*", "*I feel close to the tutor*")
- In relation with the user interface, 60.6 % of the students said that the use of the remote control for interacting with the contents had been easy. In this sense, there is divergence of opinions: some users admit they didn't send to much messages because of the problems for writing through the remote control ("*to write messages with the remote control was not*

easy for me"). However, there is agreement in the opinion that the video messages considerably reduced this problem and became very useful.
- A computer is not required.

Disadvantages:

- The main disadvantages came from the experimental technologies: isolated connection fails and STB blocks.
- Problems related with usability and access to the contents ("*I don't like to go through all the pages of a lesson instead of directly accessing to one specific page*", "*It is annoying to register in the application each time it connects*", "*I don't like that you must do the assessment of a lesson for being allowed to see the following lesson*")
- A TV is required and the telephone connection must be close to the TV. If not, you need a large telephone cable or you must stay in rooms shared with the family that complicate the concentration.
- It interferes with the usual entertainment use of the television in the family environment.

Regarding the tutors, all them familiar with e-learning applications, there exists an absolute consensus in positively stressing the personalization possibilities and the affective follow-up offered by our platform ("*It is the first time that I feel close to the student in a Moodle-based platform*", "*It is great to have emotional information of the student, I've never thought that this could be possible in an e-learning application*", "*The impact for learning is completely positive*").

7 Conclusions

In this paper we describe T-EDUCO, the first interactive t-learning affective-aware tutoring platform. The main novelty of the work lies in humanizing t-learning tutoring systems by detecting the learner's affective state and enabling the tutor and the student to communicate through video messages.

The architecture of our system takes advantage of the broadband capabilities of set-top boxes both to synchronize the interactive application with the Moodle platform and, by means of an IP camera, to record and store the facial expressions of the student in a video server. The recorded videos obtained are processed by applying a facial affect recognition method in order to extract affective information from the learner from which to create a simple and effective emotional log. The tutor can easily access both academic and emotional information about the students and consequently personalize the t-learning application contents. The system has been tested in a real broadcast setting with real users. Although the system seemed to have a very good acceptance, a rigorous assessment, comparing the difference between the uses or not of affective information, is still pending. Moreover, it will be very interesting to evaluate the impact of using the emotional recognition system for improving (or not) learning. The next steps in our work will be done in these directions.

The tool can be adapted to future interactive Digital TV standards since both the camera management and the communications with the Moodle platform are based on the Internet Protocol (IP). Other future steps in our research will consist of providing the tutoring tool with the ability to automatically analyze 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. In this way, virtual tutors may be added to the platform to offer help or automatically change the level of difficulty of tests. Moreover, the tool could be very helpful in Massive Open Online Courses

(MOOC) in which the number of students makes almost impossible to follow the students in a personalized way [13, 76].

Acknowledgments This work has been partly financed by the Spanish "Dirección General de Investigación", contract number TIN2011-24660, by the CYTED, contract number 512RT0461, by the Mechatronics and Systems Group (SISTRONIC) of the Aragon Institute of Technology and by the Spanish "Ministerio de Ciencia e Innovación" in the context of the QuEEN project, contract number IPT-2011-1235-430000.

References

1. Abadía D, Navamuel JJ, Vea-Murguia J, Alvarez F, Menendez JM, Sanchez FA, Fernandez G, Rovira M, Domech A (2009) i-LAB: Distributed laboratory network for interactive TV set-top box and services testing, *IEEE 13th International Symposium on Consumer Electronics* 773–776
2. Affdex (2013) Available: http://www.affectiva.com/affdex/
3. Alexander S, Sarrafzadeh A (2004) Interfaces that adapt like humans, computer human interaction 641–645
4. An KH, Chung MJ (2009) Cognitive face analysis system for future interactive TV. IEEE Transactions on Consumer Eectronics 55(4):2271–2279
5. Baker RR, D'Mello SK, Rodrigo MMT, Graesser AC (2010) Better to be frustrated than bored: The incidence, persistence,a nd impact of learners'cognitive–affective states during interactions with three different computer-based learning environments. Int J Human-Computer Studies 68:223–241
6. Bates PJ (2005) Learning through iDTV–results of t-learning study. Europ Confer Interact Telev 137–138
7. Bellotti F, Pellegrino M, Tsampoulatidis I, Vrochidis S, Lhoas P, Bo G, De Gloria A, Kompatsiaris I (2010) An integrated framework for personalized T-Learning in *Cases on transnational learning and technologically enabled environments* 118
8. Brusilovsky P, Schwarz E, Weber G (1996) ELM–ART: An intelligent tutoring system on World Wide Web. In Proceedings 3rd international conference on intelligent tutoring systems, ITS 96, pp. 261–269.
9. Brusilovsky P, Vassileva J (2003) Course sequencing techniques for large-scale web-based education. International Journal of Continuing Engineering Education and Lifelong Learning 13(1–2):75–94
10. Burleson W (2006) Affective learning companions: Strategies for empathetic agents with real-time multimodal affective sensing to foster meta-cognitive and meta-affective approaches to learning, motivation, and perseverance. Doctoral Thesis, Massachusetts In-stitute of Technology.
11. Caridakis G, Malatesta L, Kessous L, Amir N, Paouzaiou A, Karpouzis K (2006) Modeling naturalistic affective states via facial and vocal expression recognition. Intern Confer Multim Interf 146–154
12. Chen C (2008) Intelligent web-based learning system with personalized learning path guidance. Computers & Education vol 51:787–814
13. Clark D (2013) MOOCs: Taxonomy of 8 types of MOOC. Donald Clark Paln B. http://donaldclarkplanb.blogspot.com.es/2013/04/moocs-taxonomy-of-8-types-of-mooc.html
14. Cmolik L, Mikovec Z, Slavik P, Mannova B (2007) Personalized e-learning in interactive digital television environment, in *IADIS International Conference WWW/Internet*, 35–39
15. Conati C (2002) Probabilistic assessment of user's emotions in educational games. Journal of Applied Artificial Intelligence 16:555–575
16. Conati C, Zhao X (2004) Building and evaluating an intelligent pedagogical agent to improve the effectiveness of an educational game, Proceedings of the Ninth International Conference on Intelligent User Interface, pp. 6–13
17. D'Mello S, Jackson T, Craig S, Morgan B, Chipman P, White H, El Kaliouby R, Picard RW, Graesser A (2008) Auto tutor detects and responds to learners affective and cognitive states. In: Proceedings of the Workshop on Emotional and Cognitive Issues at the International Conference of Intelligent Tutoring Systems, pp. 23–27
18. D'Mello SK, Lehman B, Graesser A (2011) A motivationally supportive affect-sensitive autotutor. In *New Perspectives on Affect and Learning Technologies*, 113–126. New York: Springer
19. Damásio M, Quico C (2004) T-learning and interactive television edutainment: The Portuguese case study. *ED-Media* 4511–4518
20. IPEA: Instituto de Pesquisa Economica Aplicada (2013) Panorama da Comunicacao e das telecomunicacoes no Brasil, (in portuguesse): http://www4.planalto.gov.br/brasilconectado/noticias/panorama-da-comunicacoes-e-telecomunicacoes-no-brasil/

21. de Vicente A, Pain H (2002) Informing the detection of the students' motivational state: An empirical study, Proceedings of the Sixth International Conference on Intelligent Tutoring Systems, pp. 933–943
22. Desmarais MC, d Baker RS (2012) A review of recent advances in learner and skill modeling in intelligent learning environments. User Model User-Adap Inter 22(1–2):9–38
23. dos Santos DT, do Vale DT, Meloni LG (2007) Digital TV and distance learning: Potentials and limitations. In: *Proceedings of the 36th Annual Conference on Frontiers in Education* pp. 1–6
24. Dosi A, Prario B (2004) New frontiers of T-learning: The emergence of interactive digital broadcasting learning services in Europe. *ED-Media*, pp. 4831–4836
25. Ekman P, Friesen WV, Hager JC (2002) Facial action coding system, Research Nexus eBook
26. eMotion©: emotion recognition software, 2013. Available: http://www.visual-recognition.nl/Demo.html
27. Face API technical specifications brochure. Available:http://www.seeingmachines.com/pdfs/brochures/faceAPI-Brochure.pdf
28. Fragopanagos N, Taylor JG (2005) Emotion recognition in human–computer interaction. Neural Netw 18: 389–405
29. Gee JP (2004) Situated language and learning: A critique of traditional schooling. Routledge, Taylor & Francis, London
30. Graesser AC, D'Mello S, Person NK (2009) Metaknowledge in tutoring. In: Hacker D, Donlosky J, Graesser AC (eds) Handbook of metacognition in education. Taylor & Francis, Mahwah
31. Hammal Z, Couvreur L, Caplier A, Rombaut M (2007) Facial expression classification: An approach based on the fusion of facial deformations using the transferable belief model. Int J Approx Reason 46:542–567
32. Henze N, Nejdl W (2001) Adaptation in open corpus hypermedia. Int J Artif Intell Educ 12(4):325–350
33. Hone K (2006) Empathic agents to reduce user frustration: the effects of varying agent characteristics. Interact Comput 18:227–245
34. Hupont I, Baldassarri S, Cerezo E (2013) Facial emotional classification: From a discrete perspective to a continuous emotional space. Pattern Anal Applic 16(1):41–54
35. Hupont I, Ballano S, Baldassarri S, Cerezo E (2011) Scalable multimodal fusion for continuous affect sensing, Proc. IEEE Workshop on Affective Computational Intelligence (WACI 2011), Symposium Series on Computational Intelligence, ISBN 978-1-61284-082-6, pp. 68-75
36. Hupont I, Ballano S, Cerezo E, Baldassarri S (2013) From a discrete perspective of emotions to continuous, dynamic and multimodal affect sensing, Advances in Emotion Recognition, ISBN 978-1118130667, Wiley-Blackwell
37. Isen AM (2000) Positive affect and decision making. In: Lewis M, Haviland J (eds) Handbook of emotions. Guilford, New York
38. Kalman RE (1960) A new approach to linear filtering and prediction problems. Transactions of the ASME - Journal of Basic Engineering, Series D 82:34–45
39. Kapoor A, Burleson W, Picard RW (2007) Automatic prediction of frustration. International Journal of Human-Computer Studies 65:724–736
40. Kim J, Kang S (2011) An ontology-based personalized target advertisement system on interactive TV. *IEEE International Conference on Consumer Electronics*, pp. 895–896
41. Kort B, Reilly R, Picard R (2001) An affective model of interplay between emotions and learning: reengineering educational pedagogy—building a learning companion, Proceedings IEEE International Conference on Advanced Learning Technology: Issues, Achievements and Challenges. IEEE Computer Society, pp.43–48.
42. Kumar P, Yildirim EA (2005) Minimum-volume enclosing ellipsoids and core sets. J Optim Theory Appl 126:1–21
43. López-Nores M, Blanco-Fernández Y, Pazos-Arias JJ, García-Duque J, Ramos-Cabrer M, Gil-Solla A, Díaz-Redondo RP, Fernández-Vilas A (2009) Receiver-side semantic reasoning for digital TV personalization in the absence of return channels. Multimed Tools Appl 41:407–436
44. Mandryk RL, Atkins MS (2007) A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. International Journal of Human-Computer Studies 65(4):329–347
45. McDuff D, Kaliouby RE, Picard RW (2012) Crowdsourcing facial responses to online videos. IEEE Trans on Affective Computing 3(4):456–468
46. Merrill DC, Reiser BJ, Ranney M, Trafton JG (1992) Effective tutoring techniques: A comparison of human tutors and intelligent tutoring systems. J Learn Sci 2(3):277–305
47. Montpetit M, Klym N, Mirlacher T (2011) The future of IPTV. Multimedia Tools and Applications 53:519–532
48. Morrell DR, Stirling WC (2003) An extended set-valued Kalman filter, In *Proceedings of ISIPTA*, 396–407
49. Nicolaou MA, Gunes H, Pantic M (2011) Continuous prediction of spontaneous affect from multiple cues and modalities in valence-arousal space. IEEE Trans Affect Comput 2(2):92–105
50. Papanikolaou KA, Grigoriadou M (2002) Towards new forms of knowledge communication: The adaptive dimension of a webbased learning environment. Computers and Education 39:333–360

51. Pazos-Arias JJ, López-Nores M, García-Duque J, Díaz-Redondo RP, Blanco-Fernández Y, Ramos-Cabrer M, Gil-Solla A, Fernández-Vilas A (2008) Provision of distance learning services over Interactive Digital TV with MHP. Comput Educ 50:927–949
52. Petridis S, Gunes H, Kaltwang S, Pantic M, 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
53. Picard RW (2000) Affective computing, *The MIT Press*
54. Picard RW, Papert S, Bender W, Blumberg B, Breazeal C, Cavallo D, Machover T, Resnick M, Roy D, Strohecker C (2004) Affective learning-a manifesto. BT Technology Journal 22(4)
55. Plutchik R (1980) Emotion: A psychoevolutionary synthesis. Harper & Row, New York
56. Porayska-Pomsta K, Mavrikis M, Pain H (2008) Diagnosing and acting on student affect: The tutor's perspective. User Model User-Adap Inter 18(1–2):125–173
57. Rey-López M, Díaz-Redondo RP, Fernández-Vilas A, Pazos-Arias JJ, López-Nores M, García-Duque J, Gil-Solla A, Ramos-Cabrer M (2008) T-MAESTRO and its authoring tool: Using adaptation to integrate entertainment into personalized t-learning. Multimedia Tools and Applications 40:409–451
58. Rey-López M, Fernández-Vilas A, Díaz-Redondo RP (2006) A model for personalized learning through IDTV. In: Wade V, Ashman H, Smyth B (eds.) *AH 2006*, LNCS 4018, pp. 457–461
59. Roland H (2000) Logically optimal curriculum sequences for adaptive hypermedia systems. In International conference on adaptive hypermedia and adaptive web-based system, pp. 121–132.
60. Russell JA (1980) A circumplex model of affect. Journal of personality and social psychology 39(6):1161–1178
61. Russell T (1997) Technology wars: Winners and losers. Educom Review 32(2):44–46
62. Sarrafzadeh A, Alexander S, Dadgostar F, Fan C, Bigdeli A (2008) How do you know that I don't understand? A look at the future of intelligent tutoring systems, Computers in Human Behavior vol 24: 1342–1363
63. Schiaffino S, Garcia P, Amandi A (2008) eTeacher: Providing personalized assistance to e-learning students. Computers & Education vol 51:1744–1754
64. Smith ASG, Blandford A (2003) MLTutor: An application of machine learning algorithms for an adaptive web-based information system. Int J Artif Intell Educ 13(2–4):233–260
65. Snow R, Farr M (1987) Cognitive-conative-affective processes in aptitude, learning, and instruction: An introduction. Conative and affective process analysis 3:1–10
66. Soleymani M, Pantic M (2013) Emotional Aware TV, *Proceedings of TVUX-2013: Workshop on Exploring and Enhancing the User Experience for TV at ACM CHI 2013*
67. Soyel H, Demirel H (2007) Facial expression recognition using 3D facial feature distances. Lect Notes Comput Sci 4633:831–883
68. Suraweera P, Mitrovic A (2002) KERMIT: A constraint-based tutor for database modeling. In *Intelligent Tutoring Systems*, 377–387. Springer Berlin Heidelberg
69. VanLehn K (2006) The behavior of tutoring systems. International journal of artificial intelligence in education 16(3):227–265
70. Videodrips. Available: http://www.interactivetvweb.org/tutorials/mhp/content_referencing
71. Wallhoff F (2006) Facial expressions and emotion database, *Technische Universität München* Available: http://www.mmk.ei.tum.de/~waf/fgnet/feedtum.html
72. Whissell CM (1989) The dictionary of affect in language, Emotion: Theory, research and experience, vol 4. The Measurement of Emotions, New York
73. Whissell C, Whissell's dictionary of affect in language technical manual and user's guide. Available: http://www.hdcus.com/manuals/wdalman.
74. Witten I, Frank E (2005) Data mining: Practical machine learning tools and techniques, 2nd edn. Morgan Kaufmann, San Francisco
75. Yeasin M, Bullot B, Sharma R (2006) Recognition of facial expressions and measurement of levels of interest from video. IEEE Transactions on Multimedia 8:500–508
76. Zapata-Ros M, *El diseño instruccional de los MOOCs y el de los nuevos cursos online abiertos personalizados (POOCs)*, 2013 (In Spanish) [Preprint] http://hdl.handle.net/10760/19744

Sandra Baldassarri received a B.Sc. in Computer Science from University of Buenos Aires, Argentina, in 1992 and a Ph.D. in Computer Science Engineering from the University of Zaragoza, Spain, in 2004. She is Assistant Professor at the University of Zaragoza (Spain) and founder member of the AffectiveLab of the Advanced Computer Graphics Group (GIGA) at the University of Zaragoza. Her research interests include virtual humans, affective computing, multimodal interfaces and tangible and natural interaction.

Isabelle Hupont obtained a M.Sc. in Telecommunications Engineering from the University of Zaragoza (Spain) in 2006 and a Ph.D. in Computer Science Engineering in 2010. Since 2007 she has been working as a researcher at the Instituto Tecnológico de Aragón. She is also an Associate Professor at the Computer Science Department of the University San Jorge of Zaragoza. Her research interests include Human Computer Interaction, Computer Vision, emotional interfaces, multimodal interaction, Artificial Intelligence and computer graphics.

David Abadia received a B.Sc. degree in Industrial Engineering in 2001 and an Ms.C. degree in Computer Science in 2007 from the University of Zaragoza, Spain. Since 2002, he works as a senior engineer and a scientific researcher at the Aragon Institute of Technology, where he leads both private and public funding projects. Since 2004, he is also a Computer Science lecturer at the University of Zaragoza. Previously, he worked at SIEMENS AG in R&D for Software Architectures for Distributed Systems and at the German Aerospace Center in research with last generation robots for aerospace environments. His areas of interest are Video Scene Understaning (PhD focus), Computer Vision, Artificial Intelligence and Networked Interactive Multimedia Systems.

Eva Cerezo received a Ph.D. degree in Computer Science in 2002 from the University of Zaragoza. She is currently an Associate Professor at the Computer Sciences and Systems Engineering Department at the University of Zaragoza and a founding member of GIGA AffectiveLab. Her research areas are affective multimodal human computer interaction, tangible interaction and virtual humans. She is author of more than 70 international publications that take in conjunction more than 300 citations. She is editor of the Advances in Human-Computer Interaction Journal and regular reviewer of several national and international conferences like the “ACM CHI Conference On Human Factors in Computing Systems” and the “International Conference on Intelligent User Interfaces”.