

Affective computing in a T-Learning application

SANDRA¹BALDASSARRI, ISABELLE HUPONT², EVA CEREZO¹, DAVID ABADÍA²

¹ GIGA AffectiveLab, Computer Science Department, Engineering Research Institute of Aragon (I3A), Universidad de Zaragoza, Zaragoza, Spain
{sandra, ecerezo}@unizar.es

²Interactive Audiovisual Technologies Center, Aragon Institute of Technology, Huesca, Spain
{ihupont, dabadia}@ita.es

Abstract. This paper presents T-EDUCO, the first t-learning affective aware tutoring tool. T-EDUCO goes further than simply broadcasting an interactive educational application by allowing the figure of a tutor to be present and to govern the students' learning process. The tutor can access academic and emotional information about the students through a continuous "emotional path" that includes timestamps and information about the progress made in each exercise. In this way, personal messages or extra educational contents for improving learning can be sent to the students. All this is made possible by a combination of broadcast and broadband technologies

164

Keywords: affective computing, interactive television, t-learning

1 Introduction

IDTV (IDTV) is emerging as a potentially important medium to offer, at home, the access to different kind of applications such as tv-guides, commerce or games, among others. In these cases, the television is used as an interface with the user. To date, the offer has been mostly based on contents available through broadcasting, but the increasing availability of broadband communications through STBs, together with the fact that IDTV users are abandoning their passive habits, implies a new range of highly interactive services that allow the user to access individual contents such as pay-per-view channels, bank services, e-mails, etc. Moreover, being an enhancement added to traditional TV sets,

IDTV is easy to use and familiar to the general public, meeting the socially important need of offering online services to people who cannot afford to buy a computer or lack the knowledge to use one. In this way, IDTV also creates new opportunities for learning at home. Although World Wide Web (WWW) based distance learning methods still seem to be the current dominating trend, the utility of television itself as a learning tool is also well recognized [1, 2, 3]. This is especially true for socially disadvantaged communities where television has far more penetration than WWW interaction. As people are used to a TV environment, and taking into account its high impact on society due to its mass media nature, interactive TV represents a good opportunity to reduce the digital gap and to provide new ways of learning. New learning opportunities can be provided for those social groups that barely have access to traditional forms of education [4]. T-learning (TV-based interactive learning) is not just an adaptation for IDTV of e-learning techniques used on the Internet. It has its own distinctive characteristics, mostly related to the usability and technological constraints imposed by the television set and the STB, 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 STBs have lower computer power than a personal computer. For these reasons, in most t-learning applications learning via IDTV has been more about edutainment (educational entertainment) than formal learning [5, 6]: learning contents are broadcast, and the interaction with the user is limited to a simple sequential and pre-defined navigation inside the application.

More engaged and intelligent t-learning interactive applications are needed to achieve more complete and 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. The main difference 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. However, to our knowledge, there is no t-learning tutoring system in the literature with the ability of intelligently recognizing affective cues from the student. And, as facial expressions are the most extensively used method for determining user's emotions [7], making a t-learning tutoring system able to interpret facial expressions would allow it to be affectively aware and therefore more pedagogical.

This paper presents T-EDUCO, a t-learning tutoring system that goes further than simply broadcasting an interactive educational application by allowing the figure of a tutor to be present and to pedagogically govern the students' learning process. The tutor can access not only the academic information relating to the students but also emotional information and, depending on what is revealed, the tutor can send personal messages or extra educational contents for improving students' learning. The emotional information is extracted from an analysis of the facial expressions of the student taken through an IP camera while performing the evaluations (tests) provided by the broadcast interactive t-learning application at home. This is possible thanks to an open architecture that overcomes the technical limitations imposed by the STBs, combining combine broadcast and broadband technologies to guarantee both global delivery of the learning contents and personalized communications with every pupil.

2 T-EDUCO overview

T-EDUCO is an affective aware t-learning tutoring tool based on DVB-MHP (Multimedia Home Platform, version 1.1.3) [8] digital TV interactive standard. However, as the management of the camera and the communications with the educational platform are IP-based system, it is sufficiently open and scalable to be adapted to future new trends in interactive TV since both. Figure 1 presents a general overview of T-EDUCO.

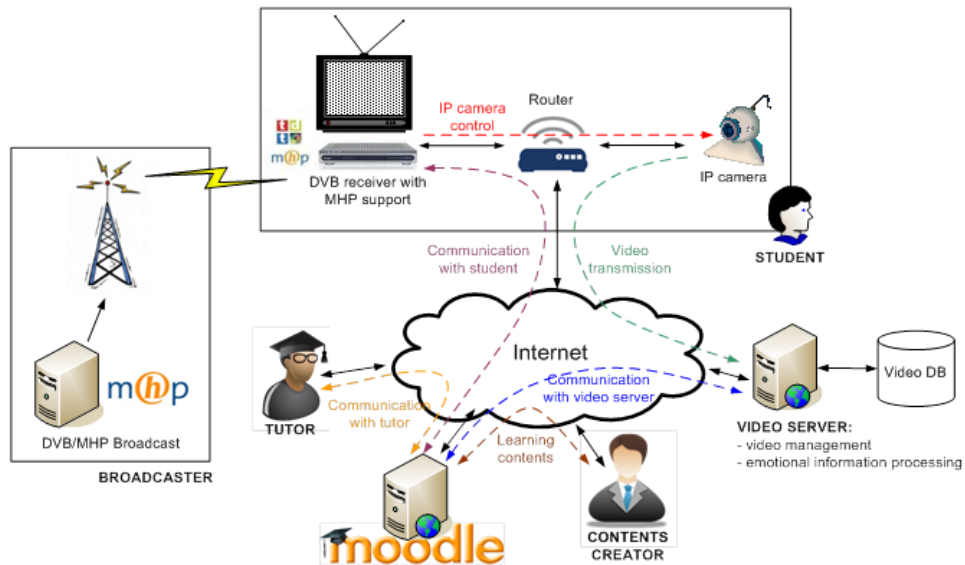


Figure 1. T-EDUCO general architecture and communication flows

The tutoring system has three main actors: the contents creator, the tutor and the student.

The contents creator accesses the t-learning application via Moodle (Modular Object Oriented Developmental Learning Environment) [9]. The contents creators are 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.

The tutor (Figure 2a), also accessing via Moodle, can create, store, organize, and modify the educational contents. 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 and the external video server via IP. In this way, T-EDUCO keeps the t-learning application and Moodle synchronized. This allows the tutor to access to the academic and emotional information of each student and to send personal messages and contents.

The student (Figure 2b) is located at home and accesses a broadcast t-learning interactive application through a set-top box, which also has IP communication capabilities. Besides showing the learning contents, the t-learning application is also able to control the management of an IP camera 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 to automatically extract emotional information.

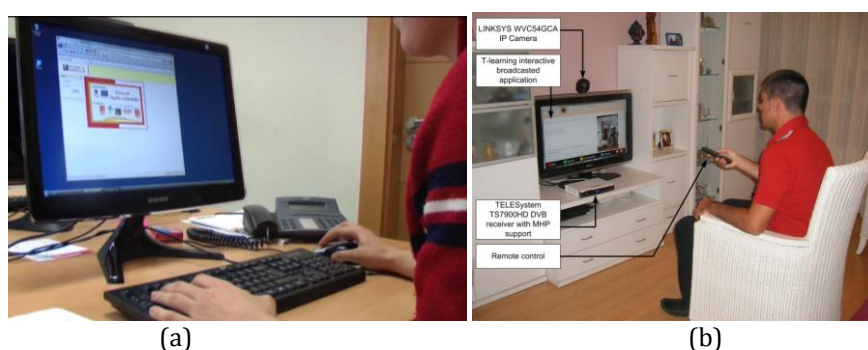


Figure 2. Typical actors' environments. (a) Tutor. (b) Student.

3 T-EDUCO personalization capabilities

The T-EDUCO system includes personalization capabilities in order to allow the tutor to adapt the learning contents depending on each student's personal academic and emotional development throughout the course. Distance learning tutors, like traditional classroom teachers, should be aware of difficulties encountered by students (noticing if they appear to be frustrated, confused, etc.) or, on the other hand, detecting 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, and personalizing the contents and messages.

Academic follow-up: The tutor can keep detailed logs of all the activities the students perform in order to determine who has been active in the course, what students have done and when.

Affective follow-up: The tutor can access the emotional information automatically extracted by the system by analyzing the learner's video sequences captured by the IP camera while interacting with the t-learning broadcast application.

Personalized contents: The interactive t-learning course is broadcast to every user with the same initial contents. Depending on the student's academic and emotional evolution throughout the course, the system offers the tutor the possibility of sending additional personalized learning contents.

Personalized messages: T-EDUCO allows the student and the tutor to interchange personal communications in the form of e-mail messages. The learner can write/access those messages through the t-learning application (see Figure 3), while the tutor can write/access the messages via Moodle.

Figure 4 shows a snapshot of the web-based interface that allows the tutor to build personalized contents for an specific student.

Note that the personalization capabilities of T-EDUCO are possible thanks to the convergence between broadcast and broadband technologies. The key element for achieving this convergence is the set-top box IP connection capabilities, i.e. the return channel that links the set-top box with the Internet and that provides a backchannel for applications that wish to communicate with the outside world. In this way, both a global delivery of the learning contents and personalized communications to every user are ensured.



Figure 3. Snapshots of the broadcasted t-learning application. (a) The student writes an e-mail to the tutor. (b) The students access the personal e-mails received from the tutor

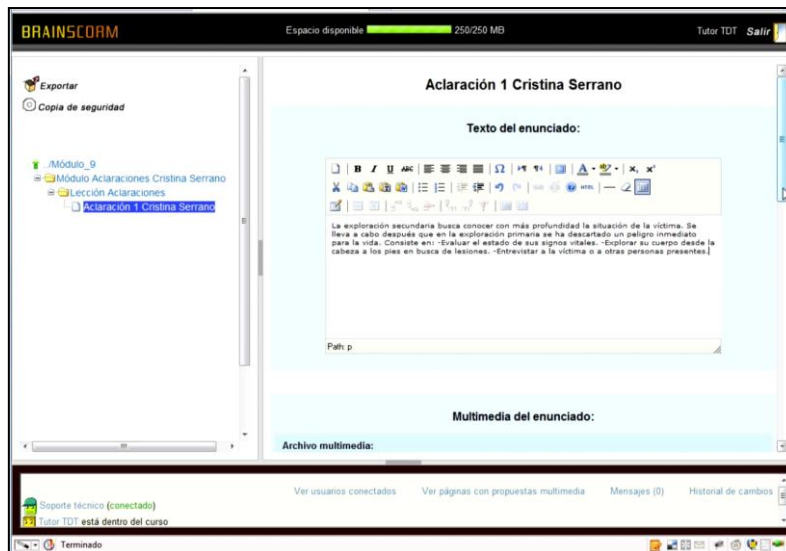


Figure 4. Snapshot of the tool that allows to build personalized contents for the student (in Spanish)

4 Considering student's emotions

Tutors are often in charge of a large number of students across different courses. Therefore, keeping close contact with each learner is difficult, especially when taking into account that t-learning applications don't allow personal contact. For this reason, it

is of interest to be able automatically to 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. This section describes how T-EDUCO uses a continuous facial image affect recognition method to automatically extract emotional information from the learners' recorded video sequences that is later shown to the distance tutor. Section 4.1 explains the context in which the student's facial expressions that serve as input to the method are captured. Section 4.2 details how T-EDUCO makes use of the method to automatically process and extract an "emotional path" from the videos, and how this extracted information is presented to the tutor.

4.1 Capturing facial expressions

Learners have a personal student ID and password to log in the broadcast t-learning interactive application. The course consists of a set of modules and each module has a final evaluation test to be performed within a limited time period.

Before starting the final test of each module, the application suggests that the student can be recorded by the IP camera while answering the evaluation questions. If the student accepts the suggestion, the set-top box middleware sends to the camera the command to start recording and store the video in the Video Server. Each video is identified by <student ID, course ID, module ID> so that it can be unequivocally accessed later. This recorded video carries useful affective information since it captures the student's facial expressions and, if efficiently processed, could be of considerable help to the tutor in following the student's affective progress.

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 clearly showing them [10]. Users' feelings are usually assessed when a task is finished. The users are asked to provide a self-report explaining how they felt during the interaction. However, it has been demonstrated that self-reported feelings at the end of a task are notoriously unreliable. Learners

often try to give the impression that they are fine when they are not [11]. For example, students may feel uncomfortable confessing that they felt frustrated during the assessment. Moreover, humans can verbalize what they perceive while in the process of perceiving. If there is a time span between perceiving and the request to recall, distorted descriptions and explanations may be produced rather than a report of our immediate thoughts [12].

4.2 Processing emotional information

The video sequences recorded from the learners are analyzed, frame by frame, by a 2D continuous emotional images recognition system [13]. 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 points are obtained thanks to faceAPI [14], a commercial real-time facial feature tracking program that provides Cartesian facial coordinates.

The output of the system represents emotions as a specific location (x; y) in a 2D space (Whissell’s space) characterized by evaluation and activation factors [15]. 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. Unlike categorical approaches, where the output is a collection of emotional labels such as “joy”, “sadness” or “fear” [16], dimensional representations are much more able to deal with a continuum of emotions and variations in emotional states over time [15, 17]. In the continuous emotional input enabled by the system, the 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 modeled, 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 [18] is proposed to model the “emotional kinematics” of that point when moving through the Whissell space and thus enabling its trajectory to be smoothed and kept under control. Kalman filters are widely used in the literature for estimation problems ranging from target tracking to a function’s approximation [19]. 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 Whissell space (x-position, y-position, x-velocity and y-velocity) are modeled as the system’s state in the Kalman framework at time t_k . The output of the 2D classification system is modeled as the measurement of the system’s state. In this way, the Kalman iterative estimation process 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 for the 2D point emotional trajectory modeling is that it can tolerate small occlusions or inaccurate tracking. The input facial feature points of the classification method are obtained using the commercial facial tracker faceAPI [15]. Unfortunately, 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 faceAPI deals with these problems quite robustly, on some occasions its performance is poor, especially when working in real-time. For this reason, its measurements include a confidence weighting, from 0 to 1, allowing the acceptability of the tracking quality to be determined. When a low level of confidence is detected (lower than 0.5), the measurement will not be used and only the filter prediction will be taken as the 2D point position. In order to demonstrate the potential of the “emotional kinematics” model, Figure 5 shows an example of its application to a complex captured facial video sequence.

The sequence is complex owing to two main factors:

- Different emotions are displayed contiguously following a neutral→surprise→joy→neutral affective pattern.
- A short facial occlusion occurs during the emotional display, causing the tracking program to temporarily lose the facial features and therefore fail to classify the frame. In these cases, the filter prediction is taken as the 2D point position for those frames.

Figure 5 shows the system's response during the different affective phases of the video sequence studied (including the occlusion period), both with and without applying the “emotional kinematics” filtering technique.

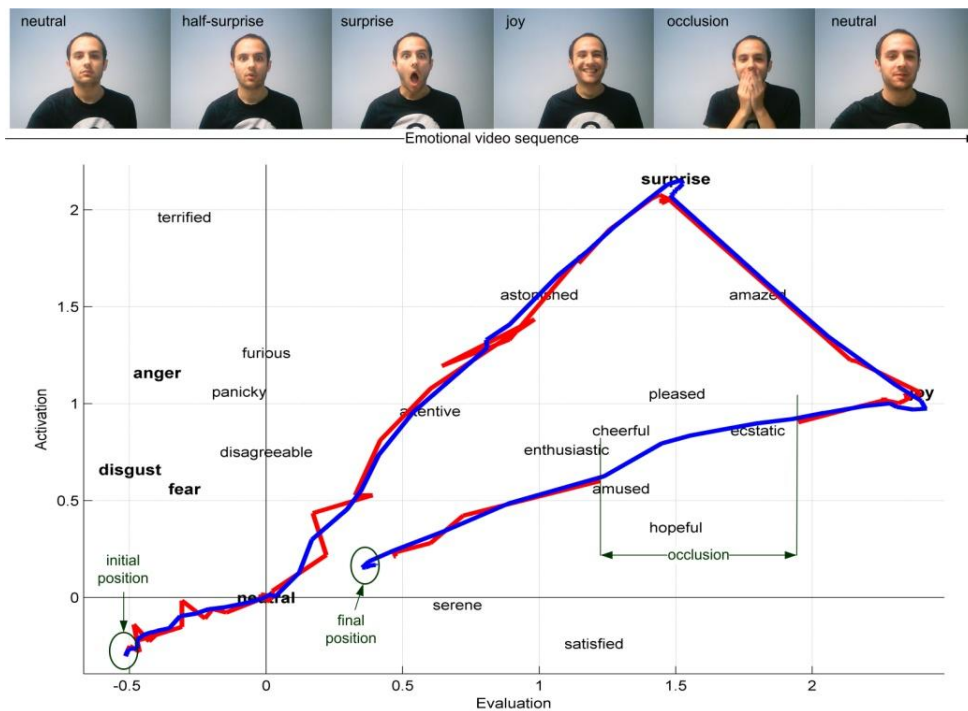


Figure 5. System response during the different affective phases of the video sequence and the occlusion period. In red, emotional trajectory without “emotional kinematics” filtering. In blue, reconstructed trajectory using “emotional kinematics” filtering

As can be observed, the system is able to recognize accurately all the affective phases of the video sequence and ensures emotional continuity even when the 2D classification method does not output

any affective information due to the occlusion. Moreover, the general emotional trajectory is smoothed when “emotional jumps” or discontinuities are detected by the 2D classification method.

The final emotional log presented to the tutor can be seen in Figure 6. It is a continuous “emotional path” that also includes timestamps and information about the beginning and the end of each exercise of the test. Finally, the tutor can easily access every emotional log through the Moodle platform and consequently take pedagogical decisions.

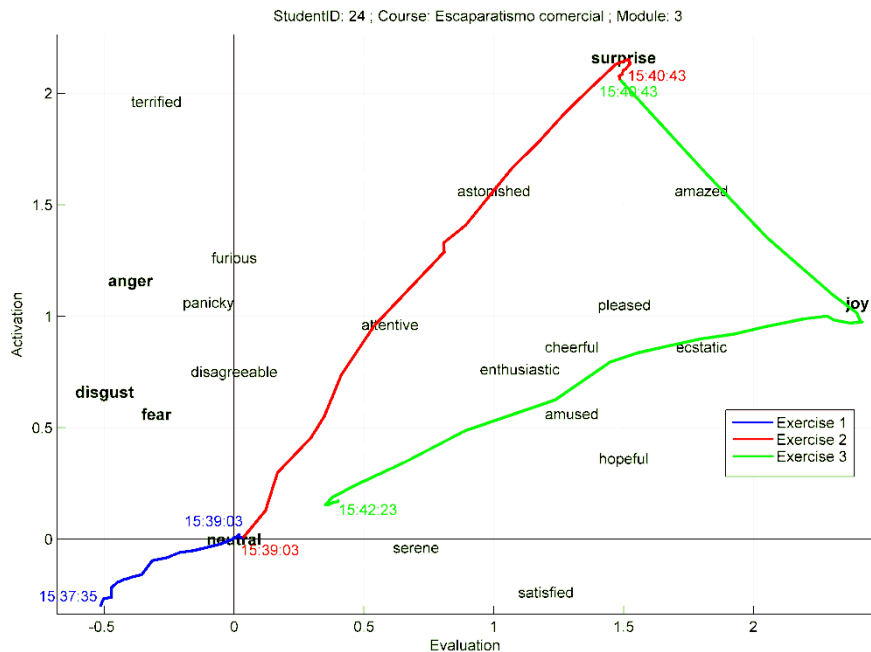


Figure 6. Example of emotional log that can be consulted by the tutor from the Moodle platform. Timestamps showing the beginning and the end of each exercise are included

5 Conclusions

In this paper a t-learning tutoring system that offers tutors emotional information about students while they interact with a distance learning application is presented. As in other t-learning tools, the students work at home through a t-learning broadcast interactive application while the distance tutor accesses T-EDUCO

through the Moodle Learning Management System. But, 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 in a video server the facial expressions of the student while carrying out the different tests of a course. The recorded videos obtained are processed in order to extract affective information from the learner from which to create a simple and effective emotional log. In this way, the tutor can easily access both academic and emotional information about the students and consequently personalize the t-learning application contents. The tool can be easily adapted to future interactive Digital TV standards since both the camera management and the communications with the Moodle platform are based in the Internet Protocol (IP).

The future steps in our research will consist in providing the tutoring tool with the ability of automatically analyzing the student's emotional logs in order to personalize course contents and self-adapt the learning process to the affective state of each student by offering help or automatically changing the level of difficulty of tests. Also, other potential sources as affective information such as statistics about time between answers in the tests or number of successive hits, will be taken into account in order to obtain more accurate information about the student's emotional state.

Acknowledgments. This work has been partly financed by the Spanish "Dirección General de Investigación" contract number N^o TIN2011-24660 and by the project FEDER ATIC.

References

1. 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.
2. P. Aarreniemi-Jokipelto, "T-learning Model for Learning via Digital TV", in *Proceedings of the 16th EAEEIE Annual Conference on Innovation in Education for Electrical and Information Engineering*, 2005.

3. D. T. dos Santos, D. T. do Vale, and L. G. Meloni, "Digital TV and Distance Learning: Potentials and Limitations", in *Proceedings of the 36th Annual Conference on Frontiers in Education*, 2007, pp. 1–6.
4. P. J. Bates, "Learning through iDTV – results of t-learning study", *European Conference on Interactive Television*, Apr. 2005, pp 137-138.
5. M. Damásio and C. Quico, "T-learning and interactive television edutainment: the Portuguese case study", *ED-Media*, 2004, pp. 4511-4518.
6. M. Rey-López, R. P. Díaz-Redondo, A. Fernández-Vilas, J. J. Pazos-Arias, M. López-Nores, J. García-Duque, A. Gil-Solla, and M. Ramos-Cabrer, "T-MAESTRO and its authoring tool: using adaptation to integrate entertainment into personalized t-learning", *Multimedia Tools and Applications*, Vol. 40, 2008, pp. 409–451.
7. R. W. Picard, *Affective computing*, The MIT Press, 2000.
8. MHP Standard Draft TS 102 812 V1.3.1 - MHP 1.1.3, 2007.
9. Moodle - A Free, Open Source Course Management System for Online Learning. Available: <http://www.moodle.org>.
10. A. Kapoor, W. Burleson, and R. W. Picard, "Automatic prediction of frustration", *International Journal of Human-Computer Studies*, vol. 65, 2007, pp. 724–736.
11. K. Schneider, and I. Josephs, "The expressive and communicative functions of preschool children's smiles in an achievement situation". *Nonverbal Behavior* 15. 1991.
12. J. Nielsen, T. Clemmensen, and C. Yssing, C. "Getting access to what goes on in people's heads?: reflections on the think-aloud technique". In *Proceedings of the Second Nordic Conference on Human-Computer interaction, NordiCHI '02*, vol. 31. ACM, New York, NY, 101-110. 2002.
13. I. Hupont, S. Baldassarri, E. Cerezo, "Facial emotional classification: from a discrete perspective to a continuous emotional space", *Pattern Analysis and Applications*, ISSN: 1433-7541, 2012 (in press, DOI 10.1007/s10044-012-0286-6).
14. Face API technical specifications brochure. Available: <http://www.seeingmachines.com/pdfs/brochures/faceAPI-Brochure.pdf>
15. C. M. Whissell, "The dictionary of affect in language", *Emotion: Theory, Research and Experience*, vol. 4, The Measurement of Emotions, New York: Academic, 1989.
16. P. Ekman, W. V. Friesen, and J. C. Hager, "Facial action coding system", *Research Nexus eBook*, 2002.
17. R. Plutchik, *Emotion: A psychoevolutionary synthesis*, New York: Harper & Row, 1980.
18. 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.
19. D. R. Morrell and W. C. Stirling, "An extended set-valued Kalman filter", In *Proceedings of ISIPTA*, 396–407. 2003.