# Managing Student Emotions in Intelligent Tutoring Systems

**Roger Nkambou**

GDAC Laboratory – University of Québec at Montréal
nkambou.roger@uqam.ca

**Abstract**

In the classic educational context, observing and identifying learner's emotional response allow the teacher to adapt the lesson, with the aim of improving the quality of the learning process. In computer environment, Intelligent Tutoring Systems (ITS) should consider relationships between emotion, cognition and action in order to improve the interaction with the learner. Hence, inspired by cognitive and behavioral models, ITS equipped with emotional recognition capability will be able to provide relevant tools for efficient and pleasant learning. In order to provide ITS the capacity of considering and managing the emotions of the learner, we propose an architecture which includes several components for emotion recognition.

## Introduction

In the social context, emotions play a prominent role in verbal end non-verbal communication. With their multi-modal perception, the aptitude of an interlocutor in identifying them, through a diversity of behaviors, like face movements, gestures, speech, is a fundamental aspect as judgment and decision making are influenced by mood, feelings, which facilitates human adaptation and social integration. In a cognitive activity such as learning, communication (between learners and tutors) is one of the fundamental issues, and its quality may influence the learning. Therefore, learning process implies cognitive aspects as well as socio-emotional aspects: in real world, teaching also implies to observe the student's affective behavior in order to detect affective responses which can express interest, excitation, confusion, etc. and suggest a review of the actual interaction flow.

In the same perspectives, Intelligent Tutoring Systems (ITS) must consider and manage affective factors (Picard, 1997). Considering the affective model of the student, an Affective ITS (AITS) – an ITS equipped with emotional management capabilities - could efficiently adapt content planning, learning/tutoring strategies and even tutoring dialogues. Nevertheless, new technical and methodological problems appear with new research issues; in this case the problem of detecting and evaluating learner's emotional state. The learner's emotional state is not directly observable by a machine, and, until now, the most objectives approaches in this field use external captors in order to examine the body position, the look and facial expressions of the learner (Burleson et al. 2004).

In order to promote a more dynamic and flexible communication between the learner and the system, we integrate two adaptive emotional agents in a multi-agent AITS. The first one allows the tutor to express emotions in response to the student's actions. An emotional tutor, called Emilie-1 (Nkambou et al. 2002) has been successfully integrated in a learning environment for on-line teaching of science. The second one aims at capturing and managing the emotions expressed by the learner during a learning session. This agent (Emilie-2) works according to a pedagogical loop which includes the following actions: 1) capture, extraction and recognition of emotions through a given emotional channel; 2) analysis, diagnosis and interpretation of the recognized emotions; 3) remediation through relevant pedagogical actions. In this paper, we focus on emotions' capture, extraction and classification functions of Emilie-2 from facial expression channel. Firstly, we present our architecture of AITS with a stress on emotions management components. Secondly, we present some details on the implementation of Emilie-2, mainly its emotions recognition component (perception layer). We then present some experiments in order to validate the quality of this component. The results of these experiments are presented and discussed. Finally, we present the current focus of our work.

## An Affective Tutoring System Architecture

This research focuses on emotional management in learning context. We believe that taking the learner emotional state into account can enhance the learning's quality. Also, tutor feedbacks could be improved if the tutor can express emotions (Kapoor and Picard, 2005). Hence, our research aims at extending both the learner and the tutor models which emotional aspects. This extension is done by integrating two emotion management agents, one for the tutor's emotions management and the other for the student's emotions management. Figure 1 presents the multi-agent architecture of an AITS which includes those two emotional agents. In this figure, the learner's model represents both his cognitive state (knowledge, skills and

performances' history) and his affective state (mood, emotions and psychological profile). We could also see that a specific agent manages the cognitive state: the profiler. This agent updates the acquired knowledge, skills and performances of the learner, and maintains the cognitive model integrity. It also helps the diagnosis of knowledge or skills incorrectly learned, or missing, and permits remediation with the help of tutoring agents.

The affective state contains short-term information (resp. medium and long-term), which corresponds to emotions (resp. mood and psychological profile) of the learner. The Emilie-2 agent manages this part of the model. Tutoring agents are the agents that contribute to the training: a planner for the selection of relevant learning activities, a coach which helps students during problem solving activities, etc.

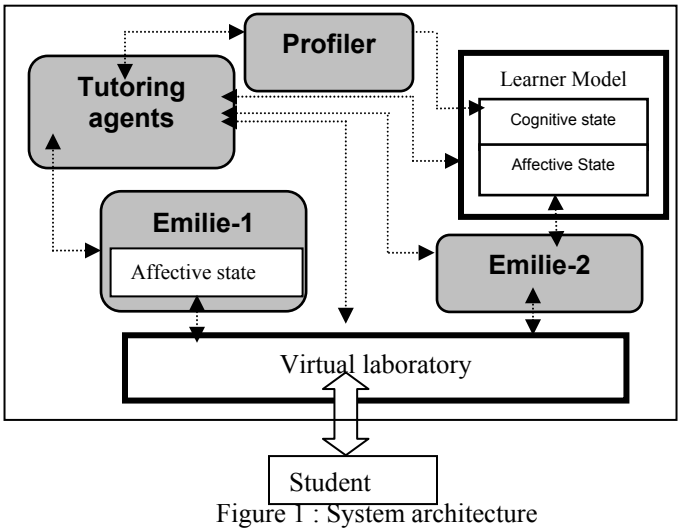


Figure 1 : System architecture

During interaction, some tutorials agents may want to include emotional dimension. Thus, they use Emilie-1 which is able, with a specific activity, to translate through a character (2D or 3D), the emotions of the tutoring agent. It has to be aware of the concerned task and of the desired emotional reaction (by the designer or the concerned tutoring agent). The emotional state of Emilie-1 is a short-term memory which represents the current emotional reaction. The action is carried out through the virtual laboratory which includes interactive tools (objects and resource) related to the content.

## Emilie-2: an agent for emotion recognition

### Architecture of Emilie-2

Emilie-2 aims, via a digital camera placed over the screen, at carrying out the acquisition of face image and analyze the facial expressions, in order to identify the emotions. This agent is made of three layers (modules) (figure 2): the first one (perception layer) captures and extracts the facial expressions (image acquisition and face tracking) and proceeds to its categorization (classification); the second one (cognition layer) analyses and diagnoses the perceived learner's emotional state and the third one (action layer) takes decision on remedy pedagogical actions to carry out in response to the actual emotional state. Tutoring agents are then informed and may access information in the new affective state (updated by Emilie-2) to adapt the current tutoring flow accordingly.

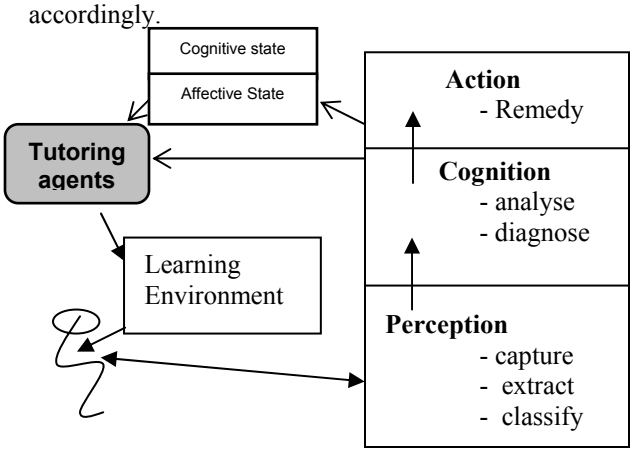


Figure 2 : Emilie-2 pedagogical cycle

The cognitive layer includes two main processes named analysis and diagnosis. The analysis of an emotional state recognized by the perception layer makes it possible to translate the meaning of this emotion in the learning context. It is carried out by taking into account several elements including the recognized emotion, the current affective profile, the historic of the actions realized before the emotion expression, the cognitive state of the student, the emotion evolution and the social context (if it corresponds to social or a collaborative learning). The emotion analysis may reveal if the student feels "satisfaction", "confidence", "surprise", "confusion", or "frustration". These states are more precise in educational context and appropriated pedagogical actions can be taken in order to influence those emotions. Another important process in the cognition layer is the diagnosis of the analyzed emotional state. This process determines the possible causes which has led to this situation (success/failure in an exercise, difficulty of the tasks, lack of knowledge, incorrect command of the knowledge, etc.). This is done using the learner's cognitive state and the historic of his actions. The action layer of the architecture includes processes that take decision about relevant pedagogical actions which can control the observed emotional state, in order to place the student in a better state (Chaffar and Frasson, 2005). Diagnosis and remediation processes in Emilie-2 are similar to those already implemented in the profiler agent, because we suppose that the emotional reactions of the learner exclusively come from learning factors. In these conditions, our hypothesis is that, the expression of negative emotions can 'only' be justified in terms of cognitive problems (lack of knowledge, misconception, misunderstanding, mal-rules… Details on diagnosis and remediation processes in connection with cognitive

problems can be found in Tchetagni et al. 2004. The next sections of this paper present some detail related to the perception layer.

## The perception layer of Emilie-2

This layer aims at perceive and recognize the emotion expressed by the student, via the analysis of its facial expressions during learning session. The perception contains three steps: facial expressions extraction, reduction and classification. The next sections present our approach for the implementation of these functions in Emilie-2.

**Facial expression extraction.** In order to promote the agent's autonomy, the facial expressions' extraction consists in two successive tasks, the learner's image acquisition, via a digital camera, and a face tracking, which eliminate a lot of useless information, like the background.

**Facial expression recognition.** In order to realize the recognition, we use the connexionist approach. Indeed, we believe that the facial expression of an emotion cannot systematically be reduced to a set of geometrical characteristics, and that a more global approach, based on the whole face (or subparts of the face), should produce better results. Furthermore, a connexionnist system, following a learning process on pre-categorized patterns (Pantic & Rothkrantz, 2000), should be able to proceed to a better generalization, allowing successful treatment on unknown pattern. Thereby, this module performs a recognition of facial expression, by classifying each image in one of the six basis facial expression categories defined by Paul Ekman (1978) (we also use the neutral expression, as the 7$^{th}$ category). The recognition module consists of two sub-modules. The main sub-module is an artificial neural network which classifies each image in one of the seven facial expression categories. The second module which works upstream of the other, reduces the dimension of the inputs. Indeed, in order to generalize the classification, an artificial neural net might use dimensional reduced inputs. A face image, even without background, still contains too much information (256 per 256 pixels represents 65536 inputs), which make it difficult to generalize.

**Data reduction module.** Among the different existing methods (Donato et al. 1999; Turk & Pentland, 1991; Penev & Atick, 1996), we have decided to use the decomposition by Eigenfaces (Turk & Pentland, 1991; Gross et al. 2004; Cader et al, 2001), a method which is well documented, simple to implement, and which may lead to good recognition results (79.3% according to Donato et al. (1999)). This method is based on the fact that there is a correlation between different pixels in an image, which means that, it is possible to come out with reduced information that can characterize the whole image. By extracting the relevant information, this method constructs a sub-space which clearly shows the significant variations of the picture, and we reduce the initial images by projecting them into this sub-space. Thus, this method constructs a new base on the eigenvectors of the images covariance matrix, and expresses each image as a linear combination of these vectors (called "eigenfaces") (figure 3). We obtain a reduction of the size of the images, because the linear coefficients are sufficient to characterize the image. Once the calculus of the eigenfaces has been made, this method selects the most relevant, those who are associated to the highest eigenvalues.

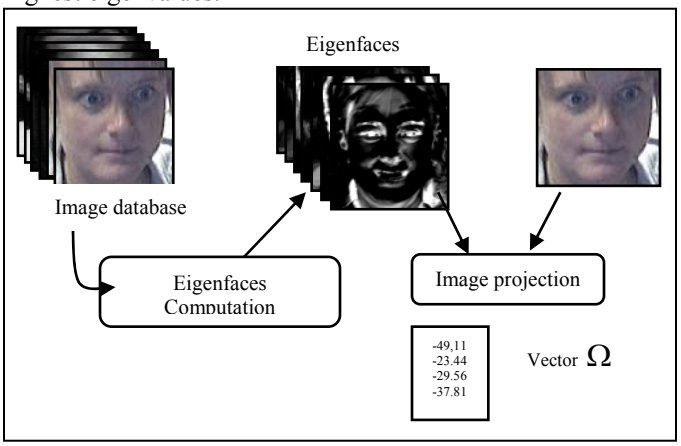


Figure 3. Description of the eigenfaces reduction method

**Facial expression classification module**. Once the extraction of the relevant information has been made, the facial expression recognition is processed using an artificial neural network.

Its architecture is made of three layers: an input layer, which provides the images, and composed by a number of neurons corresponding to the result of the data reduction process, one hidden layer, for which we will search the adequate number of neurons, and an output layer composed by 7 neurons, one for each basis facial expression plus one for the neutral expression.

In order to design the network, we have proceeded to several experimentations presented in the next section.

# Experiments and results

## Objective

The objective of the experimentations is to determine the optimal architecture of the neural network which classify facials expressions. Mainly, we want to find the adequate number of hidden neurons. For the moment, experimentations are carried out on a static images database, out of any real-time constraint. Thereby, we use the Cohn-Kanade face database (Kanade et al. 2000), which contains up to ten thousand images, accompanied by a unit action analysis (FACS). We thus are allowed to use supervised learning.

## Methodology

For the moment, we work only on three different expressions, "neutral", "happiness" and "surprise". Thereby, we have constructed our eigenfaces on sets composed by 97 images (each images represents one of the three studied expressions) of 32 subjects. To be sure of the quality of decomposition, we have set the number of eigenfaces to 75, which leads to a good quality decomposition (evaluated on the reconstruction of the image).

Concerning the network, we construct with Matlab multi-layer perceptron, and we test different architectures, by varying the number of hidden neurons from 16 to 30. This contains the empirical values advised in literature (Lepage & Solaiman, 2003). We use the resilient propagation algorithm.

## Quality testing of the learning process

In order to improve the quality of the learning process and the generalization capacity of the network, we use a testing set. This means that we use, for each learning process, 2 sets of images : the learning set, and the testing set which permits to stop the learning process if the network begins to memorize the learning set (that can be seen when the error rate of the testing set begins to increase). The global error of this learning is chosen as the testing set error rate.

Furthermore, to improve this operation on several testing set, we use a cross-validation method. Thereby, we use a training set to construct the learning set (2/3 of the training set) and the testing set (1/3 of the training set). At the end of each learning process, we change the disposition of different thirds used as testing set. The global error rate of this architecture will thus be calculated as the mean of the different learning errors (figure 4).

Currently, we use a training set of 296 images (105 of happiness, 103 of neutral and 88 of surprise), randomly sorted.

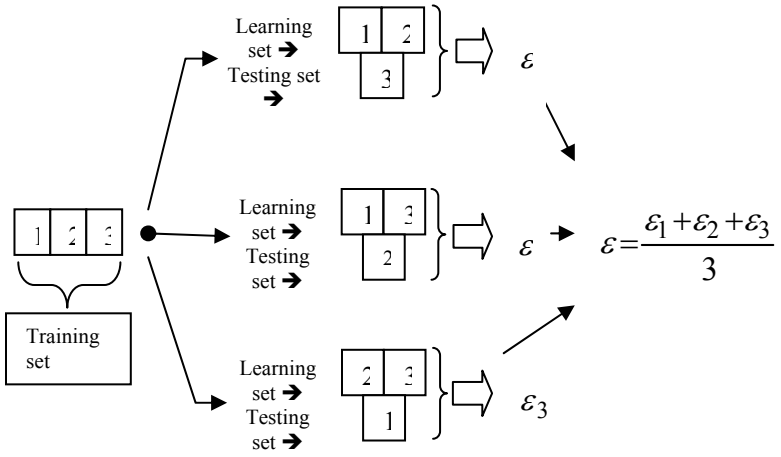


Figure 4 : Description of the learning approach

**Results:** This section presents the different results obtained during the experimentations. It shows the global error rate, and the test results for each architecture, resulting from the resilient propagation algorithm.

In order to simplify the comprehension of the obtained results, we detail the calculus operated on the first architecture, a neural network with 16 hidden neurons (figure 5).

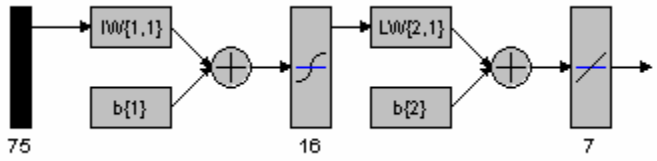


Figure 5 : The perceptron architecture 75 – 16 - 7

Doing the cross validation will lead to three different learning. The first one leads to an error rate $\varepsilon_1 = 0.0401$, which corresponds to the validation error.

The second and the third learning lead to an error rate equal to $\varepsilon_2 = 0.0158$ and $\varepsilon_3 = 0.0135$. Thereby, we obtain a global error rate, with 16 hidden neurons, of $\varepsilon = 0.0231$. In order to obtain several results, we have carried out several simulations, varying the number of hidden neurons. The results are shown in table 1.

This table shows the error rate resulting from a cross-validation learning on each architecture, repeated 10 times. We can see that the bold value, 0.0231 corresponds to the first simulation result, for a network with 16 hidden neurons. This value is the one seen in the example above.

These results show that our networks are able to learn correctly the presented images, and the low error rate for the validation curve is the sign of a good recognition.

| Number of hidden neurons | Simulation number 1 | 4 | 5 | 6 | 10 | **Mean** |
|---|---|---|---|---|---|---|
| 16 | **0.0231** | 0.0269 | 0.0252 | 0.0227 | 0.0202 | 0.0227 |
| 17 | 0.0261 | 0.0267 | 0.0265 | 0.0289 | 0.0242 | 0.0259 |
| 18 | 0.0205 | 0.0242 | 0.0301 | 0.0364 | 0.0236 | 0.0271 |
| 19 | 0.0247 | 0.0281 | 0.0194 | 0.0228 | 0.0182 | 0.0235 |
| 20 | 0.0178 | 0.0294 | 0.0358 | 0.0310 | 0.0396 | 0.0303 |
| 21 | 0.0324 | 0.0208 | 0.0239 | 0.0220 | 0.0355 | 0.0262 |
| 22 | 0.0358 | 0.0217 | 0.0313 | 0.0261 | 0.0239 | 0.0319 |
| 23 | 0.0234 | 0.0417 | 0.0285 | 0.0212 | 0.0394 | 0.0308 |
| 24 | 0.0201 | 0.0197 | 0.0232 | 0.0250 | 0.0517 | 0.0287 |
| 25 | 0.0290 | 0.0227 | 0.0350 | 0.0550 | 0.0273 | 0.0361 |
| 26 | 0.0329 | 0.0325 | 0.0317 | 0.0404 | 0.0418 | 0.0342 |
| 27 | 0.0321 | 0.0340 | 0.0277 | 0.0247 | 0.0341 | 0.0311 |
| 28 | 0.0347 | 0.0243 | 0.0479 | 0.0248 | 0.0514 | 0.0347 |
| 29 | 0.0290 | 0.0381 | 0.0382 | 0.0283 | 0.0393 | 0.0354 |
| 30 | 0.0373 | 0.0656 | 0.0500 | 0.0641 | 0.0478 | 0.0498 |

Table 1 : Error rate of the network

### Quality testing of the classification process

Further to the learning capacity of the network, we would like to test the knowledge of the network in a context of utilization,, with two different test. The first one aims at discover the classification skill of the network with already known images and the second one aims at testing the generalization capability of the network.

**The first test:** We use a set of 48 images (14 of happiness, 17 neutral and 17 surprise) beforehand learned. The test consists simply to simulate the recognition of the images at the end of learning process, and to compare the output with the given output (patterns which has been use for the learning). The results are shown in table2.

These results show that the networks are effectively able to categorize some images which has been learned, which is normal. Nevertheless, we see that these results are less good than what we could expect. We expected up to 100%, but we don't have it. This difference comes from images which are on the boundary between two categories. Several images are similar, and also are classified in different categories. This implies some mistakes.

| Number of Hidden Neurons | Happiness | Neutral | Surpise |
|---|---|---|---|
| 16 | 86% | 94% | 81% |
| 17 | 81% | 100% | 85% |
| 18 | 86% | 92% | 88% |
| 19 | 88% | 92% | 83% |
| 20 | 88% | 100% | 71% |
| 21 | 86% | 92% | 90% |
| 22 | 83% | 98% | 85% |
| 23 | 88% | 100% | 85% |
| 24 | 86% | 94% | 83% |
| 25 | 83% | 96% | 69% |
| 26 | 95% | 96% | 90% |
| 27 | 86% | 92% | 83% |
| 28 | 88% | 98% | 81% |
| 29 | 88% | 96% | 88% |
| 30 | 83% | 96% | 90% |

Table 2 : Known images recognition

| Number of Hidden Neurons | Happiness | Neutral | Surprise |
|---|---|---|---|
| 16 | 34% | 26% | 22% |
| 17 | 39% | 46% | 15% |
| 18 | 30% | 44% | 41% |
| 19 | 33% | 31% | 26% |
| 20 | 45% | 32% | 7% |
| 21 | 17% | 42% | 30% |
| 22 | 39% | 23% | 30% |
| 23 | 45% | 23% | 41% |
| 24 | 29% | 47% | 19% |
| 25 | 10% | 38% | 19% |
| 26 | 22% | 14% | 37% |
| 27 | 18% | 29% | 44% |
| 28 | 13% | 19% | 41% |
| 29 | 22% | 32% | 22% |
| 30 | 25% | 31% | 44% |

Table 3 : Un-known image recognition

**The second test**: The second test aims at establishing the generalization capacity of the network. Therefore, we present to it some new images, un-learned, representing new subjects, with a known expression (happiness, surprise, neutral). We use 70 images (35 *happiness*, 26 *neutral*, 9 *surprise*). The results are shown in table 3.

As we can see, these results are lower than the previous ones, with a recognition rate globally between 10% and 45%, depending on the expression and the architecture. Several causes could explain these results: Firstly, the weak number of images used during the learning step (296) may not lead to a high quality generalization. Secondly, we face the same problems as in the first tests. Finally, a detailed study of the results shows that several images, with a different framing, are systematically classified incorrectly. This is not surprising, artificial neural networks are very sensible to these variations.

## Lessons learned

There are too much problems with the proposed perception component: 1) Face detection using Eigenface method is too much limited and less accurate. The approach is very sensitive to some variations (head position, color...). As these variations are inevitable in learning context, this method becomes very difficult to work in this context. Although we proposed some solutions to overcome variations, it appears that, pre-processing tasks hide some errors that are propagated to the overall perception processes. 2) The facial expression extraction mainly relied on variation of the face. Hence, two identical faces with a position variation of the order of 5 pixels will lead to two different eigenvectors and thus, may be interpreted differently during the classification. 3) The perceptron neural net used is very sensitive to the entry variations and doesn't take into account the variation problem states previously for the 500 eigenvectors that are used in the ideal configuration. Also, the network size is very high given that the number of weights is around 35000 and the value of the acceptable error is 0.01. Thus, up to 3500000 images (what is too much!) are needed for a good training of that network. 4) The current version of the perception component doesn't takes into account the emotion valence which is very important in our application.

## A new approach to emotion perception

Given the problems stated previously regarding the current implementation of Emile-2's perception layer, we developed a new version using feature-based approach for face detection. The approach allows to search for different facial features (eyes, nose, mouth…) and to extract their spatial relationships which are used to compute distance variations (given a reference neutral expression). Thus a vector of those differences is given as input to the neural net. We extend the initial architecture by introducing a new component called 'personal agent' (figure 6) which

makes it possible the adaptation of the neural net training to a specific user. The neural net is a perceptron with 8 entries corresponding to the 8 most important distance variations that have been retained. The network also consider emotional valence with 3 possible values (small, medium, high) for each of the 6 emotions. That means 18 outputs + 1 (for the neutral expression) and 14 hidden layers. If we limit ourselves to 3 relevant emotions for learning context, we will have a network of 8 entries, 10 outputs and 10 hidden layers. The *Emotional_core* is actually implemented using MPT (The Machine Perception Toolbox) developed by the university of California, San Diego. The neural network is configured but should be trained with relevant databases containing normalized faces.

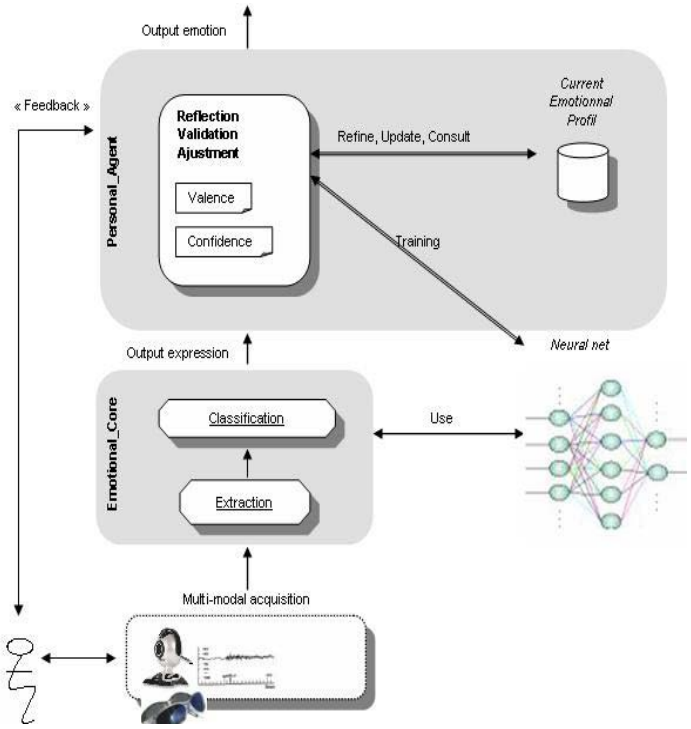


Figure 6: The new perception layer of Emilie-2

## Conclusion

Emotion recognition through facial expression analysis using neural networks has been study by several authors including (Lisetti & Schiano, 2000). However, those works face a huge number of inputs that make it difficult to reach an acceptable performance, especially when this task has to be done in real time. Emilie-2 perception module focuses on a connexionist approach with reduced inputs, in order to improve generalization and performance. We encapsulate the emotional recognition function in a pedagogical loop that allows the AITS to capture student's emotion during learning adapt student-tutor interaction accordingly. As we can see, Emilie-2 currently focuses only on facial expression. However, to accurately capture student affective behavior, information from other channels such as pressure, posture etc. should be considered. We are extending the perception layer to a multimodal emotion recognition capability.

## References


Burleson, W., Picard, R.W., Perlin, K. and Lippincott, J. , 2004. A Platform for Affective Agent Research. *Workshop on Empathetic Agents, International Conference on Autonomous Agents and Multiagent Systems*, Columbia University, NY.

Calder, A.,J., et al. 2001. A principal component analysis of facial expression. *Vision Research*, 41(9), pp. 1179-1208.

Chaffar, S. and Frasson, C. 2005. The Emotional Conditions of Learning. *Proceedings of the FLAIRS Conference 2005*, .pp. 201-206

Donato, G. et al. 1999. Classifying Facial Actions. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 21(10), pp. 974-989.

Ekman, P. and Friesen, W. V. 1978. *The Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Palo Alto: Consulting Psychologists Press.

Gross, R., Matthews, I. and Baker, S. 2004. Appearance-based face recognition and light-fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 26(4), pp. 449-465.

Kanade, T. Cohn, J. F., & Tian, Y. 2000. Comprehensive Database for Facial expression Analysis. In: *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG'00)*. Grenoble, France.

Kapoor, A. and Picard, R.W. 2005. Multimodal affect recognition in learning environments. *ACM Multimedia 2005*: 677-682

Lepage, R., Solaiman, B. 2003. *Les réseaux de neurones artificiels et leurs applications en imagerie et en vision par ordinateur*. Montréal : Presses de l'École de technologie supérieure. ISBN 2-921145-40-5.

Lisetti, C. L., Schiano, D. J. 2000. Automatic Facial Expression Interpretation : Where Human-Computer Interaction, Artificial Intelligence and Cognitive Science Intersect. *Pragmatics and Cognition* (8), pp. 185-235

Nkambou, R., Laporte, Y, Yatchou, R. et Gouradères, G. 2002. Embodied Emotional Agent and Intelligent Training System. In: *Recent Advances in Intelligent Paradigms and Applications*, pp. 233-253. Springer-Verlag

Pantic, M. and Rothkrantz, L. M J. 2000. Automatic analysis of facial expressions : the state of the art. *IEEE Transactions on Pattern analysis and Machine Intelligence*, 22(12), pp. 1424-1445.

Penev, P.S., Atick, J.J. 1996. Local feature analysis : a general statistical theory for object representation. *Network : Computation in Neural Systems*, 7(3), pp. 477-500.

Picard, R.W., *Affective Computing*. 1997, Cambridge, MA: MIT Press.

Tchetagni, J., Nkambou, R. and Kabanza, F. 2004. Epistemological Remediation in Intelligent Tutoring Systems. In : *Proceedings of the 17th IEA-AIE,* pp. 955-966. LNAI 3029, Springer-Verlag, Berlin.

Turk, M., Pentland, A. 1991. Eigenfaces for recognition. *Journal of Cognitive NeuroScience*. 3(1), pp. 71-86.