

Student's Uncertainty Modeling through a Multimodal Sensor-Based Approach

Imène Jraidi and Claude Frasson

Université de Montréal, Department of Computer Science and Operations Research, 2920 chemin de la tour, H3T-1J8 QC, Canada // jraidiim@iro.umontreal.ca // frasson@iro.umontreal.ca

(Submitted May 1, 2011; Revised September 25, 2011; Accepted December 1, 2011)

ABSTRACT

Detecting the student internal state during learning is a key construct in educational environment and particularly in Intelligent Tutoring Systems (ITS). Students' uncertainty is of primary interest as it is deeply rooted in the process of knowledge construction. In this paper we propose a new sensor-based multimodal approach to model users' uncertainty from their affective reactions and cognitive and personal characteristics. An experimental protocol was conducted to record participants' brain activity and physiological signals while they interacted with a computer-based problem solving system and self-reported their perceived level of uncertainty during the tasks. We study key indicators from affective reactions, trait-questionnaire responses, and individual differences that are related to uncertainty states. Then we develop models to automatically predict levels of uncertainty using machine learning techniques. Evidence indicated that students' uncertainty is associated to their mental and emotional reactions. Personal characteristics such as gender, skill level, and personality traits also showed a priori tendencies to be more or less in particular uncertainty states. The SVM algorithm demonstrated the best accuracy results for classifying students' uncertainty levels. Our findings have implications for ITS seeking to continuously monitor users' internal states so they can ultimately provide efficient interventions to enhance learning.

Keywords

Student uncertainty, EEG, Physiological sensors, Affect, Intelligent tutoring systems

Introduction

Research in distance education, and more precisely in Intelligent Tutoring Systems (ITS), tended to privilege cognitive aspects of teaching in which the learning process was considered as a set of information processing steps devoid of affective aspects, until studies in cognitive science, artificial intelligence, and neuroscience show that the brain mechanisms associated with emotions are not only related to cognitive mechanisms (Cytowic, 1989), but also solicited in perception, problem solving, decision making, and other cognitive processes (Cytowic, 1989; Damasio, 1994; Picard, 1997). Since then, various research areas including education, psychology, computational linguistics, and artificial intelligence devote a growing interest in the close links between affect and learning (Breazeal, 2003; Conati, 2002; Lester, Towns, & FitzGerald, 1999; Picard, 1997) as emotions have an impact on attention, motivation, memorization, and information processing (Goleman, 1996; Pekrun, 1992).

Affective user modeling has become a key construct in human-computer interaction and particularly in ITS (Conati & Maclaren, 2005; Picard, 1997). Even if there is no validated universal theory on emotions neither any consensus or agreement about which emotional states are related or pertinent to learning (Picard et al., 2004; Woolf et al., 2009), some studies rely on theoretical models explicitly linking emotions to learning (D'Mello et al., 2008; Kort, Reilly, & Picard, 2001) while other research focus on particular states such as frustration (Burlison, 2006; McQuiggan, Lee, & Lester, 2007), stress (Prendinger & Ishizuka, 2005) or attention (Rebolledo-Mendez et al., 2009).

Students' uncertainty is of primary interest as it is considered among the most recurrently observed states during computer tutoring and due to its theorized relationship to learning (Craig et al., 2004; Graesser & Olde, 2003; Kort et al., 2001; Pon-Barry et al., 2006; VanLehn et al., 2003). Indeed, the state of uncertainty is deeply rooted in the process of knowledge construction; it is related to a state of confusion or hesitation that one inevitably goes through when a misconception or a lack of knowledge or understanding arises. It can also signal a lack of confidence with regards to one's self-efficacy in performing specific tasks. Depending on the case and the frequency of this state, a tutor can decide either to intervene by providing appropriate aid or hints to help the student to clear up his misconception, encouraging him to be more confident in the learned concepts, or to let him get over this state by himself.

Hence, a tutoring system has to efficiently identify the student's state in order to formulate the appropriate response and adapt his pedagogical strategies accordingly. Nevertheless, actual ITS still cannot compete with human tutors who can readily detect from a glance that a student appears uncertain. In most work so far, uncertainty modeling relied on acoustic-prosodic, lexical, or discourse features extracted from utterance/dialogue-based system interactions (D'Mello et al., 2008; Liscombe, Hirschberg, & Venditti, 2005; Pon-Barry et al., 2006). However, uncertainty has barely been linked to learners' emotional and mental manifestations. We consider that this state inevitably involves these dimensions, and that current discourse/utterance features can be insufficient or even imprecise, as they cannot always reflect users' uncertainty. We also believe that uncertainty encompasses as well, cognitive factors and is specific to each individual and context.

In this paper we propose a new multimodal sensor-based approach to model learner uncertainty by integrating these factors. We use various data sources from learners' electrophysiological activity assessing their mental and emotional reactions as well as cognitive and personal characteristics. The hypotheses we establish is that (1) these features are related to the state of uncertainty and (2) can effectively predict a learner's uncertainty level. An experimental study was conducted to test these hypotheses. In this experiment, an acquisition protocol was established to record student affective reactions, trait-questionnaire responses and individual differences while they interacted with a logical problem solving system and expressed various levels of uncertainty.

Our approach is two-fold: First we analyze key trends that are associated to uncertainty, then we develop predictive models to automatically assess uncertainty states, which involves training machine learning algorithms using reported levels of uncertainty to supervise the classification process. Evidence indicated significant correlations between the electrophysiological sensor data and students' reported uncertainty levels. Personal characteristics such as gender, skill level, and personality traits also showed a priori tendencies to be in particular uncertainty states. The SVM algorithm demonstrated the best accuracy results for classifying students' uncertainty levels.

This paper is structured as follows. In the first section we present related work on automatic detection of students' uncertainty and sensor-based affective modeling in the tutoring community. In the second section we describe our experimental setup and methodology. In the third section we present and discuss the obtained results. We conclude in the fourth section and present future work.

Related work

Promising results have been reported on correlating uncertainty to learning (Craig et al., 2004; Graesser & Olde, 2003; Kort et al., 2001; Pon-Barry et al., 2006; VanLehn et al., 2003). VanLehn et al. (2003) view uncertainty as a "learning impasse" that occurs when students realize that they lack knowledge, getting thus more involved to understand the material they are learning and about which they are uncertain. This creates an opportunity for the student to engage in constructive learning. Graesser and Olde (2003) describe uncertainty as a "cognitive disequilibrium" in which learners confront difficulties that fail to match their expectations, which causes deliberation and inquiry aimed at restoring cognitive equilibrium. Other studies show that adapting and responding to student uncertainty can greatly improve learning (Forbes-Riley & Litman, 2010; Pon-Barry et al., 2006).

Besides significant studies have been conducted on automatically recognizing uncertainty in tutoring systems (Carberry & Schroeder, 2002; D'Mello et al., 2008; Liscombe et al., 2005; Pon-Barry et al., 2006). Pon-Barry et al. (2006) for example, used linguistic cues (such as hedges, response latencies, or filled-pause signals) extracted from human tutoring corpus through a frequency analysis to detect users' uncertainty in a tutoring system. Liscombe et al. (2005) used acoustic-prosodic features to classify student uncertainty in a corpus collected from a speech-enabled intelligent tutoring system. Carberry and Schroeder (2002) proposed an algorithm to recognize doubt by examining linguistic and contextual features of dialogue in conjunction with world knowledge including stereotypical beliefs ascribed to the dialogue. We believe, however that uncertainty is a rather complex state, which inevitably involves an affective dimension manifested by particular mental and emotional activations and is specific to each individual and context and that current discourse/utterance features can be insufficient or even imprecise, as they cannot always reflect users' uncertainty. The main contribution of this study is to propose an alternative approach for uncertainty recognition based on new information sources such as users' affective reactions and personal characteristics.

On the other side, the integration of physiological data combined with artificial intelligence techniques in ITS proved their effectiveness in assessing user state, trying to bridge the gap between actual tutoring systems and face to face education and improve technology's adaptability by accurately detecting student's affect, adapting tutorial interventions, and providing appropriate strategies to assist him to foster optimal conditions for learning (D'Mello et al., 2005; McDaniel, et al., 2007; Picard, 1997; Prendinger & Ishizuka, 2005; Woolf et al., 2009). Most of these studies use non-intrusive sensors to analyze a variety of physical cues including observable changes like face expressions, body postures, vocal tones, and physiological signal changes such as heart rate, skin conductivity, temperature, or respiration.

Moreover, with the advent of consumer oriented electroencephalograms (EEG), it is now possible to measure a learner's mental state with a high time resolution and precision and develop systems that directly modulate their tasks to neural indexes of cognition. The growing progress in developing portable, convenient, and low cost EEG headsets and devices allows using EEG technology within operational educational environments (Chaouachi, Jraidi, & Frasson, 2011; Stevens, Galloway, & Berka, 2007). Neural research established various EEG-based mental gauges of alertness, engagement, or executive load using features extracted from power spectral density (PSD) bands or event related potential (ERP) components (Pope, Bogart, & Bartolome, 1995; Prinzl, Freeman, & Scerbo, 2000; Serman et al., 1993). More precisely, EEG studies on mental concentration and attention defined an EEG indicator of attention to internal processing during performance of mental tasks (Harmony et al., 1996). They have found that an increase of the brain activity within the delta and low theta frequency band is related to an increase in subjects' mental concentration.

In this paper, we propose a new multimodal sensor-based approach to model students' uncertainty by integrating affective indicators using neurological and physiological sensors to track users' emotional activations and mental concentration as well as cognitive and personal criteria within a problem solving context. We seek to identify key trends/indicators that are related to uncertainty states and develop a predictive model to assess students' uncertainty levels.

Methodology and experimental design

The experimental setup consists of a problem solving system, a 6-channel EEG headset, physiological sensors, and two video feeds. Data were synchronized using necessary time markers in order to integrate the recorded signals with the rest of the instrumental setup under specific (un)certainty states. The problem solving system consists of a series of logical tasks that do not require particular prerequisites but involve a high level of attention. These tasks imply inferential skills on information series and are typically used in brain training exercises or tests of reasoning. The system is composed of 3 modules. Each module is concerned with specific forms of data: the first module deals with geometrical shapes, the second module with numbers, and the third module with letters. Each module starts with a tutorial explaining the task and giving examples to get users accustomed with the types of problems. Then, 5 multiple-choice questions related to each tutorial are given. Learners were asked to respond as quickly and efficiently as possible. They were informed that a correct answer is rewarded 4 points, -1 point is given for a bad answer, whereas 0 point is given for a no-answer. A fixed time limit of 80 seconds for each question was imposed. Failing to give an answer within the allowed time was considered as a no-answer. We detail our methodology and protocol in the following subsections.

Considerations for uncertainty elicitation

One of the most important points in this study was to obtain accurate data related to specific uncertainty states. Thus problem tasks were selected in a way that potentially causes uncertainty. To choose the right answer, learners needed to deduce a logical rule. Without this rule, the learner was not able to be sure of his answer. Moreover, problems had different difficulty levels and some of them involved a second rule to decide between two answers that both match the first rule. For instance in the geometrical module, three shapes were successively presented in the interface. The first shape represented a black triangle, the second a white rectangle and the third a black pentagon. The learner was then asked to deduce the fourth element by choosing one answer among five possibilities. In this example, the rule that one should deduce is to add a side in each shape and the correct answer would be a hexagon. Two hexagons (black and white) were included among the propositions and only one matches to the second rule that one should also

deduce (i.e., alternating between the two colors) and the correct answer would be the white hexagon. Other questions were designed to systematically mislead the learners. For instance in the number-based module, two perpendicular data series were presented. In the vertical series all the numbers were multiples of seven and in the horizontal series all the numbers were multiples of five. In this task, one should deduce the missing intersection element, which should be a multiple of both five and seven. But no such element was given among the possible answers.

After each question, the system interacted with the learners and prompted them to report how they answered to the question by choosing between the following: “I was certain about my response” or “I was uncertain about my response.” Furthermore to assess uncertainty granularity levels, learners were prompted to choose between the following: “I was certain at 50% or more” or “I was certain at less than 50%”, whenever an uncertain response was reported. Hence three possibilities can be registered for each question: certain (Cert), uncertain (Uncert) and no-answer (No_Resp) with two possible granularity levels for Uncert, namely certain at 50% or more (Low_Uncert) or certain at less than 50% (High_Uncert).

Electrophysiological recordings

Three types of sensors were used during the experiment namely electroencephalogram (EEG), skin conductance (SC), and blood volume pulse (BVP) sensors. Data were digitized using the ProComp Infinity multi-channel data acquisition system (Thought Technology Ltd., 2007).

EEG is a measurement of the electrical brain activity produced by the synaptic excitations of neurons. During the session, learners wore a stretch electro-cap and EEG was recorded from sites P3, C3, Pz, and Fz as defined by the international 10-20 electrode placement system (Jasper, 1958). Each site was referenced to Cz and grounded at Fpz. Two more active sites were used namely A1 and A2 typically known respectively as the left and right earlobe. This setup is known as the “referential linked ear montage” and is depicted in figure 1. In this montage, roughly speaking, the EEG signal is equally amplified throughout both hemispheres. Moreover, the “linked-ear” setup yields a more precise and cleaner EEG signal by calibrating each scalp signal to the average of the left and right earlobe sites. For example, the calibrated C3 signal is given by $(C3 - (A1 + A2) / 2)$. Each scalp site was filled with a non-sticky proprietary gel from Electro-Cap and impedance was maintained below 5 Kilo Ohms. Any impedance problems were corrected by rotating a blunted needle gently inside the electrode until an adequate signal was obtained. The recorded sampling rate was at 256 Hz. Due to its weakness (on the order of micro volts), the EEG signal needs to be amplified and filtered. Besides, the electrical brain signal is usually contaminated by external noise such as environmental interferences caused by surrounding devices. Such artifacts alter clearly the quality of the signal. Thus a 60-Hz notch filter was applied during data acquisition to remove these artifacts. In addition, the acquired EEG signal easily suffers from noise caused by user body movements or frequent eye blinks. Thus a 48-Hz high pass and 1-Hz low pass de-noising filters were applied.

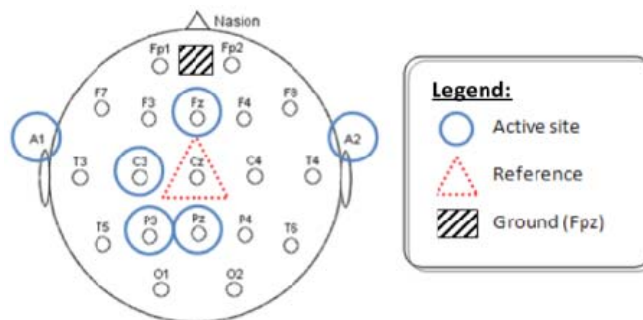


Figure 1. EEG channel electrode placement

BVP and SC sensors were placed in the resting left hand fingers. Data were recorded at a sampling rate of 1024 Hz. SC measures changes in the resistance of the skin produced by the sweat gland activity. A tiny voltage is applied through two electrodes strapped to the first and middle fingers on the palm side. This establishes an electric circuit and allows us to quantify the skin's ability to conduct the electricity. BVP sensor was placed on the tip of the ring

finger. It emits an infrared light and measures the amount of light reflected by the surface of the skin. This amount varies with the amount of blood present in the skin and thus with each heartbeat.

Affective data gathering

From the EEG raw signals, we computed mental concentration. As previously mentioned, this neural index is given by the brain activity within the delta and low theta (delta_low_theta) frequency band (Harmony et al., 1996). An EEG power spectrum was calculated for each electrode site using a Fast Fourier Transformation and the needed frequency band was extracted (1.56 - 5.46 Hz). We then computed a relative power value from the transformed signal by calculating the rate of the delta_low_theta sub-band range over the total EEG frequency band range (1.56 – 48 Hz). EEG relative power values were then summed from the electrode sites P3, C3, Pz, and Fz to compute the global ratio. A mean relative power band rate was measured for each task of the logical test.

SC signals were used to derive the galvanic skin response (GSR) widely known to linearly vary with the arousal ratings (Lang, 1995). It increases as a person becomes more stressed. From the BVP signal, the heart rate (HR) was calculated by measuring the inverse of the inter-beat intervals (distance between successive pulse peaks). The HR is extensively applied to understand the autonomic nervous system function and has shown a close correlation to valence (Lang, 1995). Both mean HR and mean GSR values were recorded for each entry. Normalization was done by subtracting current values from the baseline, and dividing the difference by the standard deviation.

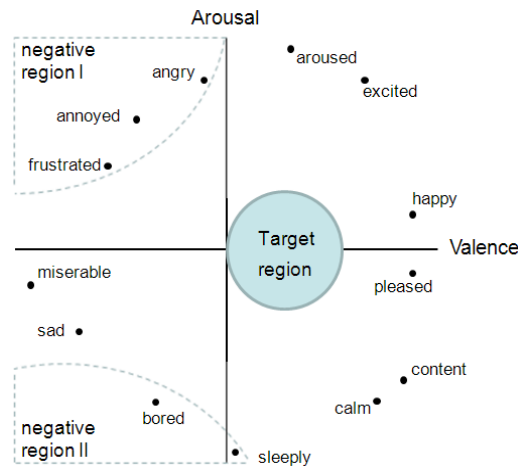


Figure 2. Russell's Circumplex model of emotions with regions

HR and GSR were jointly used to measure specific emotional activations as emotions can be characterized in terms of judged valence (negative to positive) and arousal (low to high) (Lang, 1995). We used Russell's Circumplex model of emotions (Russell, 1980) that classifies emotions within the two-dimensional arousal/valence emotional space. Two strategic emotional regions were defined during learning as depicted in figure 2. The first region involves negative emotions like frustration, boredom, or anger (negative region I and II) and should be avoided. The second region is the target emotional region specified by a slight positive valence and neutral arousal. This region is assumed to provide a maximum of efficiency and productivity in learning (Kaiser, 2006). In our study, we focused on the proportions of positive emotions within the target region for each question. We weighted then the number of HR and GSR recordings corresponding to this region by the total number of recordings.

Participants and protocol

Thirty-eight learners (14 women) with a mean age of 27.31 ± 6.87 years ranging from 19 to 47 years were recruited for the experiment. Participation was compensated with 10 dollars. Upon arrival at the laboratory, participants were briefed about the experimental objectives and procedure and asked to sign a consent form. Learners were then outfitted with the sensors and a 5-minute baseline was recorded to establish a neutral state for the electrophysiological parameters. Problem solving tasks were then completed and response time was recorded for

each question. Learners were then asked to fill in information about their age, gender, skill level in logical based problem solving (low, or medium to high), and scales on a personality test, namely the Big Five Inventory (BFI). This test scales personality traits according to five dimensions, namely Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism (OCEAN) (John, Naumann, & Soto, 2008).

Results and discussion

Our main hypotheses were that users' brain activity, emotional reactions, and cognitive and personal characteristics (1) are related to their uncertainty level, and (2) can be effectively used to predict their actual uncertainty state. Figure 3 shows the general architecture of our approach. After completing the recording process (described in the previous section), we first identify through correlational analyses, key indicators from the recorded data that could be associated to uncertainty (I). Then we develop predictive models to detect levels of uncertainty using machine learning algorithms (II).

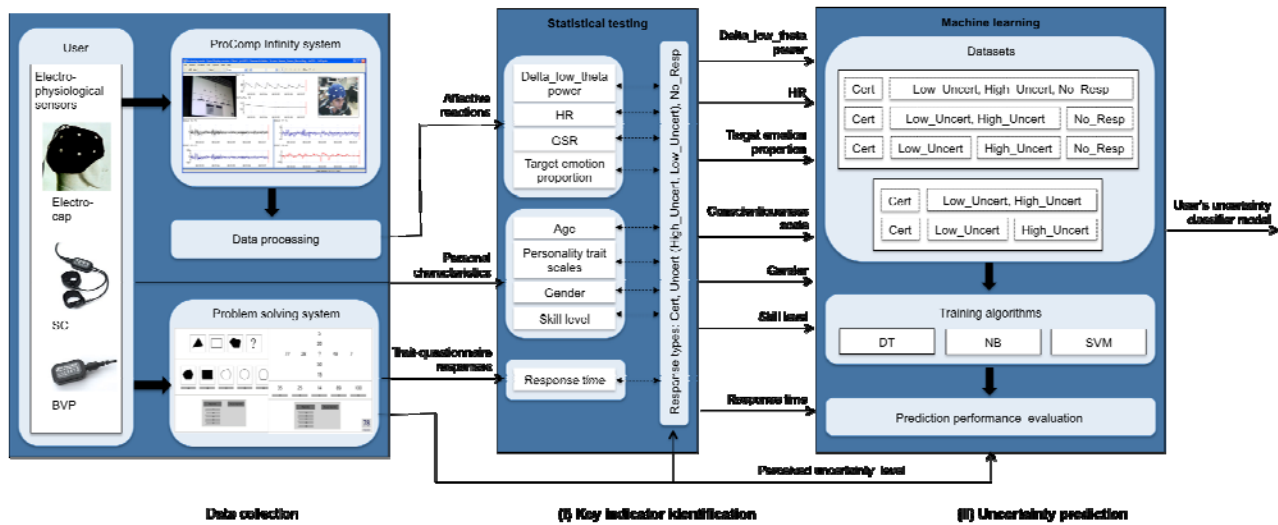


Figure 3. General architecture of the sensor-based approach

A total of 570 entries (15 questions x 38 participants) were gathered from this experiment: 323 for Cert responses (56.67%), 189 for Uncert (33,16%) with 103 entries for High_Uncert (54.50%) and 86 Low_Uncert (45.50%), and 58 No_Resp (10.17%). We detail in the following subsections the results obtained from both analyses.

Key trends in learners' uncertainty

We started by investigating the relationships between learners' uncertainty and their affective reactions according to the electrophysiological signals recorded across the four response types of the problem solving questionnaire. Statistical testing was performed using one-way analyses of variance (ANOVA). Figure 4 shows the results for the means of delta_low_theta relative power rates, HR, and proportions of emotions within the target region of Russel's Circumplex model of emotions.

First, a main effect of response type was found for the delta_low_theta relative power values. An although small but significant difference was observed across the four conditions ($F(3, 566) = 3.559, p < 0.05$). This suggests that a statistically significant difference of mental concentration (signaled by the rates of the delta_low_theta power band) exists between the four reported levels of uncertainty. We observed the highest rates of delta_low_theta for the No_Resp and the High_Uncert groups of answers. This suggests that a state of certainty (i.e., the learner is sure of his reasoning and hence his response) does not necessarily imply a higher level of mental concentration and that being uncertain does not mean a lack of mental concentration but can be instead a sign of that if we also consider that in the No_Resp, learners were indeed very uncertain and did not take the risk to respond so they do not lose one point from the final score of the quiz, or that they did not find the answer within the allowed time. One can explain

this from another perspective, i.e., in case of uncertainty, the learner tends to be more focused and involved, trying harder to reach the solution of the problem and having difficulty in finding the logical rule between the data, which costs him a higher level of mental concentration as opposed to a state of certainty in which he is more at ease with the exercise. This confirms previous studies about the theorized relationship between learning and uncertainty (Craig et al., 2004; Graesser & Olde, 2003; Kort et al., 2001; Pon-Barry et al., 2006; VanLehn et al., 2003) where it is suggested that uncertainty can signal the advent of constructive learning, since that students tend to be more engaged to understand and clarify the fuzzy knowledge and concepts causing their uncertainty.

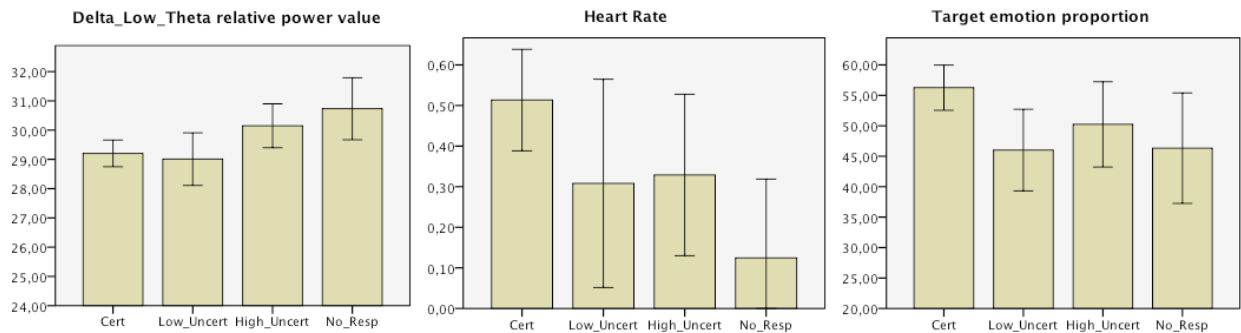


Figure 4. Comparison of affective variables for the four different response types.

The error bars denote the standard error

Second, a statistically significant difference between the types of responses was found for the HR signals ($F(3, 566) = 2.709, p < 0.05$). The group of Cert responses was significantly associated with the highest HR values suggesting that the more certain the students were about their answers, the more likely they tended to have positively valenced emotions. Positive valence for affective modeling -even if there is no wide agreement upon its interpretation- is more associated to positive emotions (Lang, 1995). This interpretation strengthens the intuitive fact that when a student is certain about his response, he tends to manifest a calm attitude and express positive emotions like satisfaction or joy. However, a state of uncertainty is usually related to negative emotions like confusion, dissatisfaction, or frustration.

No significant differences were found between the groups of answers for the GSR data ($F(3, 566) = 1.623, p = n.s.$), suggesting that the types of responses were not related to the intensity of the emotional reactions (arousal) but rather to their valence. In order to go further within this analysis, we considered the proportions of positive emotions within the target region of the Circumplex model, given by a slight positive valence and neutral arousal. We found a statistically significant difference of the mean of target emotional proportions across the four types of responses ($F(3, 566) = 3.361, p < 0.05$). We observed the highest proportions for the Cert responses, which suggests that the learners were more frequently within the target region when they were certain about their answers.

Based upon the obtained results, different trends in terms of concentration, valence, and positive emotional activations can be related to the state of uncertainty. Other interesting trends were observed looking at the response time across the four groups of responses; we found a statistically significant difference between the four conditions ($F(3, 566) = 137.925, p < 0.01$). The shortest response times were observed for the Cert responses ($M = 29.75, SD = 19.00$) compared to Low_Uncert ($M = 54.39, SD = 20.11$), High_Uncert ($M = 66.42, SD = 12.73$), and No_Resp ($M = 59.10, SD = 20.35$). Indeed, one can expect that a learner responds faster when he is certain about his answers than when he is uncertain and can take more time to try to figure out the solution. This natural tendency confirms that learners' response time should be taken into account for an accurate uncertainty assessment.

In our next investigation we analyzed learners' individual a priori tendencies to be in particular (un)certain states. We examined the impact of personal characteristics namely age, personality traits, gender, and skill level in the logical based problem solving tasks, by looking at the eventual individual associations between these criteria and the number of answers for each level of (un)certainly.

First, we ran bivariate correlations to assess the relationships respectively between participants' age and each of the five personality trait scales (OCEAN) and the number of answers for the different considered types of responses. No significant correlation was found with regards to the age variable. However, for the personality traits, statistically

significant Pearson's correlation coefficients (r) were found for the conscientiousness trait scale. Table 1 summarizes the correlational results. Interestingly, we found a positive although low correlation between the conscientiousness trait scale and the number of Cert responses ($r = 0.364$), a low negative correlation with the number of Uncert responses ($r = -0.399$), and a moderate negative correlation with the number of High_Uncert responses ($r = -0.501$). These correlations were statistically significant ($p < 0.05$) suggesting that the more conscientious the participant was, the more he tended to be uncertain in the logical quiz.

Table 1. Bivariate correlational results

<i>Correlation between the conscientiousness personality trait and the uncertainty levels (N = 38)</i>		
Response Type	r	p
Cert	0.364*	0.025
Uncert	-0.399*	0.013
Low_Uncert	0.076	0.652
High_Uncert	-0.501**	0.001
No_Resp	-0.019	0.908

* Correlation is significant at the 0.05 level (2 tailed).

** Correlation is significant at the 0.01 level (2 tailed).

Second, one-way ANOVAs were performed to test the associations between the number of answers of each response type and respectively, participants' perceived skill level in logical problem solving and the gender variables. A statistically significant effect of the skill level was found for the number of Cert responses ($F(1, 36) = 4.346$, $p < 0.05$) and for the number of High_Uncert responses ($F(1, 36) = 4.268$, $p < 0.05$). Results revealed that learners with moderate to high skill level had more certain answers ($M = 9.18$, $SD = 2.23$) than learners with low skill level ($M = 7.56$, $SD = 2.52$) and less highly uncertain answers ($M = 2.18$, $SD = 1.65$ versus $M = 3.44$, $SD = 2.09$). This suggests that participants with moderate to high skills in logical based problem solving were more certain about their answers in the logical quiz than participants with a low skill level. A significant effect of gender was also found for the number of High_Uncert responses ($F(1, 36) = 4.872$, $p < 0.05$). Women reported more highly uncertain responses ($M = 3.57$, $SD = 2.13$) than men ($M = 2.21$, $SD = 1.64$). These results underline the importance of individual characteristics/differences (such as gender, skill level, or personality traits) for a multidimensional modeling of uncertainty.

Uncertainty prediction

In the previous section we were interested in identifying indicators that can distinguish between learner trends and contribute in assessing levels of uncertainty. It was found that several facets from their electrophysiological activity as well as cognitive and personal parameters were significantly related to their state of uncertainty confirming hence our first hypothesis. In this section we are interested in the second hypothesis of this research that is a combination of these factors can reliably assess a user's uncertainty level. We train classifier models by taking as an input, features that revealed statistically reliable associations with the (un)certainty levels namely delta_low_theta rate, HR, target emotions proportions, response time, gender, skill level and the conscientiousness trait scale.

First, we trained a binary classifier to predict the Uncert from the Cert responses. Then, we extended the analysis to predict users' uncertainty in a more detailed level (High_Uncert, Low_Uncert, Cert). Besides, two separate datasets were considered. In the first dataset, No_Resp samples were either included with the Uncert samples or gathered in a separate class. In the second dataset, No_Resp samples were discarded. This separation is motivated by the ambiguous interpretation of the No_Resp samples (10.17% of the data). Does a no-response mean a high level of uncertainty such that the learner was unable to reach the solution of the problem within the allowed time or did not take the risk to respond? Or does it merely indicate that the learner did not have the time to respond even if he knew the correct answer? Table 2 shows the accuracies of classification results from three machine learning algorithms namely Decision Tree (DT), Naïve Bayes (NB) classifier, and Support Vector Machines (SVM) (Witten & Frank, 2005).

Prediction performance was evaluated using a K-fold cross validation technique (Efron & Tibshirani, 1993). The input dataset is divided into K subsets. The classifier is trained on K-1 subsets and evaluated on the remaining subset. This process is repeated K times, the accuracy estimates are averaged to yield the overall classifier accuracy. This

study employed the Weka software (Witten & Frank, 2005), a collection of machine learning algorithms intended for data mining tasks. We used the software's default parameters for the three algorithms with $K = 20$.

Table 2. Classifier accuracy results 1^{st} dataset (No_Resp included)

Classes	DT	NB	SVM
Cert, Uncert	77.37%	78.42%	79.64%
Cert, Uncert, No_Resp	72.46%	71.58%	73.33%
Cert, Low_Uncert, High_Uncert, No_Resp	64.56%	63.68%	65.08%
<i>2nd dataset (No_Resp excluded)</i>			
Classes	DT	NB	SVM
Cert, Uncert	78.90%	80.86%	83.25%
Cert, Low_Uncert, High_Uncert	73.84%	71.67%	74.46%

As presented in table 2, the SVM classifier has shown the highest prediction rates in all cases with accuracies ranging from 65.08% for the 4-class model (Cert, Low_Uncert, High_Uncert, No_Resp) to 83.25% for the binary model (Cert, Uncert) excluding the no-answers from the training set (2^{nd} dataset). Indeed, we noticed that merging the No_Resp examples in the Uncert category slightly decreases the quality of the model to 79.64% (Cert, Uncert binary model in the 1^{st} dataset), which suggests that trained models are clearly sensitive to the introduced inputs and hence that the no-answers can eventually involve both uncertainty and certainty states, which introduces a bias in the model.

Table 3. Classifier accuracy results without the sensor data

<i>1st dataset (No_Resp included)</i>			
Classes	DT	NB	SVM
Cert, Uncert	69.66%	70.64%	72.07%
Cert, Uncert, No_Resp	68.26%	67.22%	69.28%
Cert, Low_Uncert, High_Uncert, No_Resp	57.80%	60.73%	62.08%
<i>2nd dataset (No_Resp excluded)</i>			
Classes	DT	NB	SVM
Cert, Uncert	70.90%	72.07%	74.66%
Cert, Low_Uncert, High_Uncert	68.58%	67.32%	70.69%

These results confirm our second hypothesis. That is a classifier model of a user's uncertainty state can be built on the basis of a multimodal combination of factors from affective variables, namely mental concentration, valence, and positive emotional activations, trait-questionnaire features such as the response time and individual differences such as gender, skill level, and personality trait scales. Results also suggest that this approach can be further extended to handle several levels of uncertainty. We believe that our method could be an appropriate alternative for an ITS to automatically assess users' uncertainty states using machine learning techniques applied to EEG, GSR, and HR measures using non-intrusive sensors, as well as cognitive and personal criteria, so that ultimately, the prediction could be used to guide learning during computer-based education.

In order to highlight the contribution of these additional sensors, we replicated our analysis by excluding all the sensor data from the prediction models' inputs. Table 3 shows the results of the classification results from the three algorithms using the same above setup. Accuracy rates ranged from 57.80% for the 4-class model (Cert, Low_Uncert, High_Uncert, No_Resp) in the 1^{st} dataset to 74.66% for the binary model (Cert, Uncert) excluding the no-answers in the 2^{nd} dataset. Prediction performance decreased for the three algorithms in all the cases as compared to the previous approach including the sensor data. These results confirm that a multimodal sensor-based method yields more accurate predictive models. For instance, for the best-case binary model (Cert, Uncert), the prediction accuracy of the SVM classifier decreased from 83.25% to 74.66% for the same considered settings. This suggests that there is a non-negligible contribution and an obvious advantage of integrating affective data through these electrophysiological sensors to assess learners' uncertainty states and that endowing ITS with capabilities to track learners' mental and emotional reactions could give rise to more accurate user one-line monitoring and thereby eventually providing intelligent assistance and more efficient automated interventions and tutorial adjustments.

Conclusion and future works

In this paper we have proposed a new multimodal sensor-based approach to assess students' uncertainty on the basis of their cerebral and emotional behavior using electrophysiological data with cognitive and personal variables. An experimental protocol was established by recruiting 38 participants to record EEG, BVP, and SC signals as well as trait-questionnaire responses and individual criteria namely age, gender, skill level, and personality trait scales. Participants interacted with a logical problem solving system designed to elicit uncertainty and reported their perceived level of uncertainty regarding each answer. These responses were used to supervise the classification process.

Results confirmed that students' cerebral activity and emotional reactions in terms of mental concentration, HR, and positive target emotions with regards to the Circumplex model of emotions were significantly associated to different uncertainty levels. We also observed that participants' individual differences contributed to some trends to be in particular uncertainty states. Finally, we developed classifiers to automatically predict levels of uncertainty using machine learning techniques, with the SVM algorithm demonstrating the best accuracy results (83.25%), and showed that a sensor-based modeling approach yields more precise predictions as opposed to a conventional modeling. This work should however be extended with a deeper comparative study with regards to current methods of uncertainty assessment.

Our future research trends will be focused on using this sensor-based approach to track the students' states and guide the teaching process in a way that enhances users' cognitive abilities and learning performance. In the short term, we are planning to extrapolate uncertainty models within more complex learning situations, gathering more data, and refining the models by incorporating further parameters from learners' profiles. In the long term, we will be interested in developing a tutor that will integrate real time model predictions and select appropriate pedagogical strategies according to the classifiers' outputs. The tutor will use associations between user's actions and internal states to adjust the tutoring content. Further variables such as the frequency of user's uncertainty, history of the presented concepts and system's interactions, and answer correctness will be considered to track the subjacent potential cause of uncertainty and adapt the problem difficulty levels and the adequate support to the user.

Acknowledgments

We would like to acknowledge the National Science and Engineering Research Council (NSERC) and the Tunisian Ministry of Higher Education and Scientific Research for funding this work. We are also thankful to Maher Chaouachi for his valuable comments.

References

- Breazeal, C. (2003). *Designing sociable robots*. Cambridge, MA: MIT Press.
- Burleson, W. (2006). *Affective learning companions: strategies for empathetic agents with real-time multimodal affective sensing to foster meta-cognitive approaches to learning, motivation, and perseverance* (Unpublished PhD thesis). Massachusetts Institute of Technology, Cambridge, MA.
- Carberry, S., & Schroeder, L. (2002). Toward recognizing and conveying an attitude of doubt via natural language. *Applied Artificial Intelligence*, 16(7-8), 495 - 517.
- Chaouachi, M., Jraidi, I., & Frasson, C. (2011). Modeling mental workload using EEG features for intelligent systems. In J. Konstan, R. Conejo, J. L. Marzo & N. Oliver, *Proceedings of User Modeling, Adaptation and Personalization* (pp. 50-61). Berlin Heidelberg: Springer.
- Conati, C. (2002). Probabilistic assessment of user's Emotions in educational games. *Applied Artificial Intelligence*, 16(7-8), 555-575.
- Conati, C., & Maclaren, H. (2005). Data-driven refinement of a probabilistic model of user affect. In L. Ardissono, P. Brna, & A. Mitrovic (Eds.), *Proceedings of the 10th International Conference on User Modeling* (pp. 40-49). Berlin Heidelberg: Springer.
- Craig, S., Graesser, A., Sullins, J., & Gholson, B. (2004). Affect and learning: An exploratory look into the role of affect in learning with AutoTutor. *Journal of Educational Media*, 29(3), 241-250.

- Cytowic, R. E. (1989). *Synesthesia: A union of the senses*. New York, NY: Springer-Verlag.
- D'Mello, S., Craig, S. D., Gholson, B., Franklin, S., Picard, R., & Graesser, A. (2005, January). *Integrating affect sensors in an intelligent tutoring system*. Paper presented at the Intelligent User Interfaces Workshop, San Diego, CA.
- D'Mello, S., Craig, S., Witherspoon, A., McDaniel, B., & Graesser, A. (2008). Automatic detection of learner's affect from conversational cues. *User Modeling and User-Adapted Interaction*, 18(1), 45-80.
- Damasio, A. (1994). *Descartes' error: Emotion, reason and the human brain*. New York, NY: Grosset and Putnam.
- Efron, B., & Tibshirani, R. (1993). *An introduction to the bootstrap*. New York, NY: Chapman and Hall.
- Forbes-Riley, K., & Litman, D. (2010). Designing and evaluating a wizarded uncertainty-adaptive spoken dialogue tutoring system. *Computer Speech and Language*, 25(1), 105-126.
- Goleman, D. (1996). *Emotional intelligence: Why it can matter more than IQ*. London, UK: Bantam Books.
- Graesser, A., & Olde, B. A. (2003). How does one know whether a person understands a device? The quality of the questions the person asks when the device breaks down. *Journal of Educational Psychology*, 95(3), 524-536
- Harmony, T., Fernández, T., Silva, J., Bernal, J., Díaz-Comas, L., Reyes, A., ... Rodríguez, M. (1996). EEG delta activity: An indicator of attention to internal processing during performance of mental tasks. *International Journal of Psychophysiology*, 24(1-2), 161-171.
- Jasper, H. H. (1958). The ten-twenty electrode system of the International Federation. *Electroencephalography and Clinical Neurophysiology*, (10), 371-375.
- John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative big-five trait taxonomy: History, measurement, and conceptual Issues. In O. P. John, R. W. Robins & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (pp. 114-158). New York, NY: The Guilford Press.
- Kaiser, R. (2006). *Prototypical development of an affective component for an e-learning system* (Unpublished Master Thesis). University of Rostock, Rostock, Germany.
- Kort, B., Reilly, R., & Picard, R. (2001). An affective model of interplay between emotions and learning: Reengineering educational pedagogy-building a learning companion. In T. Okamoto, R. Hartley, Kinshuk, & J. P. Klus (Eds.), *Proceedings of the International Conference on Advanced Learning Technologies* (pp. 43-46). Los Alamitos, CA: IEEE Computer Society.
- Lang, P. J. (1995). The emotion probe: Studies of motivation and attention. *American Psychologist*, 50(5), 372-385.
- Lester, J. C., Towns, S. G., & FitzGerald, P. J. (1999). Achieving affective impact: Visual emotive communication in lifelike pedagogical agents. *IJAIED*, 10(3-4), 278-291.
- Liscombe, J., Hirschberg, J., & Venditti, J. J. (2005, September). *Detecting certainty in spoken tutorial dialogues*. Paper presented at the 9th European Conference on Speech Communication and Technology, Lisbon, Portugal.
- McDaniel, B. T., D'Mello, S., King, B. G., Chipman, P., Tapp, K., & Graesser, A. (2007). Facial features for affective state detection in learning environments. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th Annual Cognitive Science Society* (pp. 467-472). Austin, TX: Cognitive Science Society.
- McQuiggan, S. W., Lee, S., & Lester, J. C. (2007). Early prediction of student frustration. In A. Paiva, R. Prada, & R. W. Picard (Eds.), *Proceedings of the 2nd international conference on Affective Computing and Intelligent Interaction* (pp. 698-709). Berlin, Heidelberg: Springer-Verlag.
- Pekrun, R. (1992). The impact of emotions on learning and achievement: Towards a theory of cognitive/motivational mediators. *Applied Psychology*, 41(4), 359-376.
- Picard, R. (1997). *Affective Computing*. Cambridge, MA: MIT Press.
- Picard, R., Papert, S., Bender, W., Blumberg, B., Breazeal, C., Cavallo, D., ... Strohecker, C. (2004). Affective learning - A manifesto. *BT Technology Journal*, 22(4), 253-269.
- Pon-Barry, H., Schultz, K., Bratt, E. O., Clark, B., & Peters, S. (2006). Responding to student uncertainty in spoken tutorial dialogue systems. *International Journal of Artificial Intelligence in Education*, 16(2), 171-194.
- Pope, A. T., Bogart, E. H., & Bartolome, D. S. (1995). Biocybernetic system evaluates indices of operator engagement in automated task. *Biological Psychology*, 40(1-2), 187-195.
- Prendinger, H., & Ishizuka, M. (2005). The empathic companion: A character-based interface that addresses users' affective states. *Applied Artificial Intelligence*, 19(3-4), 267-285.

- Prinzel, L. J., Freeman, F. G., & Scerbo, M. W. (2000). A closed-loop system for examining psychophysiological measures for adaptive task allocation. *International Journal of Aviation Psychology*, 10(4), 393-410.
- Rebolledo-Mendez, G., Dunwell, I., Martínez-Mirón, E., Vargas-Cerdán, M., de Freitas, S., Liarokapis, F., & García-Gaona, A. (2009). Assessing NeuroSky's Usability to Detect Attention Levels in an Assessment Exercise. In J. A. Jacko (Ed.), *Proceedings of the 13th International Conference on Human-Computer Interaction: New Trends* (pp. 149-158). Berlin, Heidelberg: Springer-Verlag.
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161-1178.
- Sterman, M. B., Kaiser, D. A., Mann, C. A., Suyenobu, B. Y., Beyma, D. C., & Francis, J. R. (1993, October). *Application of Quantitative EEG Analysis to Workload Assessment in an Advanced Aircraft Simulator*. Paper presented at the Human Factors and Ergonomics Society Annual Meeting, Seattle, Washington, USA.
- Stevens, R., Galloway, T., & Berka, C. (2007). EEG-related changes in cognitive workload, engagement and distraction as students acquire problem solving skills. In C. Conati, K. McCoy, & G. Paliouras (Eds.), *Proceedings of the 11th International Conference on User Modeling* (pp. 187-196). Berlin Heidelberg: Springer.
- Thought Technology Ltd. (2007). <http://www.thoughttechnology.com/bioinf.htm>
- VanLehn, K., Siler, S., Murray, C., Yamauchi, T., & Baggett, W. B. (2003). Why do only some events cause learning during human tutoring? *Cognition and Instruction*, 21(3), 209-249.
- Witten, I. H., & Frank, E. D. (2005). *Data mining: Practical machine learning tools and techniques*. San Francisco, CA: Morgan Kaufmann.
- Woolf, B., Burleson, W., Arroyo, I., Dragon, T., Cooper, D., & Picard, R. (2009). Affect aware tutors: Recognising and responding to student affect. *International Journal of Learning Technology*, 4(3/4), 129-164.