A Theoretical and Empirical Approach in Assessing Motivational Factors: From Serious Games To an ITS

Lotfi Derbali, Pierre Chalfoun and Claude Frasson

Département d'informatique et de recherche opérationnelle
Université de Montréal, 2920 Chemin de la Tour, Montréal, CANADA
{derbalil, chalfoup, frasson}@iro.umontreal.ca

Abstract
This study investigates Serious Games (SG) to assess motivational factors appropriate to an Intelligent Tutoring System (ITS). An ITS can benefit from SG’ elements that can highly support learners' motivation. Thus, identifying and assessing the effect that these factors may have on learners is a crucial step before attempting to integrate them into an ITS. We designed an experiment using a Serious Game and combined both the theoretical ARCS model of motivation and empirical physiological sensors (heart rate, skin conductance and EEG) to assess the effects of motivational factors on learners. We then identified physiological patterns correlated with one motivational factor in a Serious Game (Alarm triggers) associated with the Attention category of the ARCS model. The best result of three classifiers run on the physiological data has reached an accuracy of 73.8% in identifying learners’ attention level as being either above or below average. These results open the door to the possibility for an ITS to discriminate between attentive and inattentive learners.

Keywords: Motivational factors, game elements, ITS, physiological sensors, EEG attention ratio.

Introduction
An Intelligent Tutoring System (ITS) focuses on learners’ specific needs by assessing learning difficulties and offering help and support through the use of various components. One important component is known as the Learner Model (LM). It attempts to describe learners’ behaviors and evaluate their knowledge. Another important, and complementary, component is called the Tutor Model (TM). Through the use of the LM, the tutor customizes learning environments by adapting learning strategies in order to respond intelligently to learners’ needs, objectives and interests. It does not come as a surprise then to see the massive body of work aimed at enriching the LM by modeling various learners’ cognitive and emotional states such as goals and moods (for a extensive review, see Conati and Maclaren 2009). What is surprising, however, is the existence of only a handful of papers that have considered modeling motivational states in the LM. Yet, motivation plays an important role in both learners’ performance and persistent use of an ITS over time. One possible explanation is the fact that learners’ interaction with an ITS has always been considered to be intrinsically motivating. However, learners’ negative emotions or motivational states such as boredom or disengagement have been known to appear following a certain period of interaction with an ITS (Arroyo et al. 2007; de Vicente and Pain 2002). These affective states can cause motivational problems, decrease the learning benefits of the ITS or even cause the learners to start “gaming” the system (Baker et al. 2006). In addition, learners have also been known to experience a lower sense of relatedness to the ITS (Rovai and Lucking 2003), thus increasing their feeling of isolation and possibly leading to further motivational issues. In such a context, it seems both relevant and wise to investigate further research avenues in an attempt to reduce, and eventually repair these issues.

In fact, within the researchers who have begun tackling these problems, some have found that non classical environments, such as Serious Games (SG), seemed to show a promising potential from a motivational standpoint. Indeed, several studies reported that key elements can make games motivational, such as fun and fantasy (Garris et al. 2002; Malone and Lepper 1987). The key elements become motivational factors only if they support or enhance motivation in SG. However, in order to identify and “extract” the successful motivational factors from SG, it becomes important to identify which of those factors may be relevant for learning and if so, be easily integrated into an ITS. Drawing on recent research that attempted to theoretically identify motivational factors in SG that may support learners’ motivation (Huang et al. 2010; Hung et al. 2009; McNamara et al. 2009), we aim in this study to
identify motivational factors both in a theoretical and empirical manner by combining psychometric instruments with physiological recordings, namely heat rate (HR), skin conductance (SC) and brainwaves (EEG). We ask in this paper the two following research questions: Can we assess motivational factors in SG by quantifying their impact on learners both theoretically and empirically? If so, can we construct a classifier to identify physiological patterns related to learners’ motivational states?

The organization of this paper is as follows: in the first section, we present previous work related to our research. In the second section, we explain our approach in assessing motivational factors in SG. In the third section, we detail our experimental methodology. In the fourth section, we present the obtained results and discuss them, in the last section, as well as present future work.

Related Work

Serious Games (SG) are computer applications that combine serious intent, learning and training possibly by using video environments or computer simulations. They are considered suitable teaching tools to support learning experiences (Prensky 2001). For instance, (Johnson and Wu 2008) have used a Serious Game called Tactical Language and Culture Training System (TLCTS) to help learners quickly acquire functional skills in foreign languages and cultures. TLCTS includes interactive lessons that focus on specific communicative skills and interactive play to help apply those skills. For over 20 years now, several researches have been using computer games as a platform for studying intrinsic motivation for learning. They have reported important factors responsible for the positive effect created by computer games such as challenge, curiosity, control, sensory stimuli, interaction and fantasy (Garris, et al. 2002; Malone and Lepper 1987; Prensky 2001).

Recently, SG have been used in an attempt to overcome learners’ motivational problems. Ryan and colleagues have stated for example that the motivational pull of computer games is attributed to the combination of optimal challenge, and informational feedback (Ryan et al. 2006).

In addition to studying the key factors in SG that motivate learners, other researchers have assessed learners’ motivation in an attempt to highlight the importance of these factors in overcoming motivational issues in learning. Indeed, the assessment of learners’ motivation, or lack of, has been the subject of several studies (Boyer et al. 2008; de Vicente and Pain 2002). One such study by (Arroyo, et al. 2007) evaluated the impact of a set of non-invasive interventions in an attempt to repair students’ disengagement while solving geometry problems in a tutoring system. They claimed that showing students’ performance after each problem re-engages students, enhances their learning, and improves their attitude towards learning as well as towards the tutoring software.

Nevertheless, the effectiveness of any study regarding the assessment of learners’ motivational state changes depends on two important factors: (1) the choice of proper assessment tools and (2) the accuracy of the selected tools. Assessment of motivation has been classically done through the use of self-reports in the form of a questionnaire (Guay et al. 2000; Keller 1987; Malone and Lepper 1987). Recently, numerous researchers have integrated physiological reactions in assessing learners’ motivational state. As a matter of fact, the physiological effects of motivation can, and have been, measured in learners in terms of peripheral nervous system activity expressed by changes in heart rate (HR), skin conductance (SC) and brainwaves by recording waveform patterns through the use of electro-encephalography, or EEG (Brogni et al. 2006; Derbali and Frasson 2010; Rebollo-Mendez et al. 2010).

We aim in this study to identify Serious Game elements susceptible of providing motivational support to learners and assess their impact by using both theoretical and empirical measuring tools, namely the ARCS questionnaire and physiological sensors (HR, SC and EEG). We will start by detailing our approach in the following section.

Motivational Factors

Theoretical approach. The key issue in this paper is related to the identification and assessment of motivational factors in Serious Games (SG) that support and enhance learners’ motivation. In order to define motivational factors in SG, we first need to present the tools used to measure motivation itself. In the present study, the ARCS model of motivation (Keller 1987) has been chosen to theoretically assess learners’ motivation in SG. Keller’s model has been used in learning, training and games (Dempsey and Johnson 1998; Gunter et al. 2006) and therefore is of particular interest in our study. Keller used existing research on motivational psychology to identify four categories and twelve sub-categories of motivation to constitute the ARCS model of motivation:

Attention: attracts learners’ attention at the beginning and during the process of learning. Diverse activities should be considered to maintain students’ feelings of novelty, thus the attention can be sustained.

Sub-categories: Variability (A1), Perceptual arousal (A2) and Inquiry arousal (A3).

Relevance: informs learners of the importance of learning and explains how to make it meaningful and beneficial.

Sub-categories: Familiarity (R1), Goal orientation (R2) and Motive matching (R3).

Confidence: allows learners to know the goal and to believe that the goal can be achieved, if enough effort (physical and/or intellectual) has been made.

Sub-categories: Success opportunities (C1), Personal control (C2) and Learning requirements (C3).

Satisfaction: provides feedback on performance and allows learners to know how they are able to perform well and apply what is learned in real life situations.

Sub-categories: Positive satisfaction (S1), Natural consequences (S2) and Equity (S3).
The justification for using Keller’s ARCS model in SG is based on the work of Grunter and colleagues (Gunter, et al. 2006). The authors proposed a formal paradigm for SG design where they established a mapping between Keller’s ARCS Model and common game design elements. Thus, we define a motivational factor as being a game element susceptible of providing motivational support for learners.

For example, the sub-category A1 (variability) can be considered a motivational factor related to the attention category of the ARCS model only if it maintains learners’ interest by sufficiently varying the instructional elements in SG. Otherwise, it is simply a Serious Game element and not a motivational factor. After identifying the various motivational factors in SG, we then proceed to evaluating their impact on learners.

**Empirical approach.** Performance, time spent in a game, response time, and physiological reactions are examples of various indicators that can be used to determine the close relationship between motivational factors and learners’ motivational state. We decided to evaluate the impact of motivational factors on learners in this study by using non-invasive physiological sensors (HR and SC). These sensors are typically used to study human affective states (Lin et al. 2007). However, we decided to add another interesting and important sensor: EEG. Indeed, brainwave patterns have long been known to give valuable insight into the human cognitive process and mental state (Wilson and Fisher 1995). This paper explores the intricate relationship between the Attention category in the ARCS model and its corresponding cerebral fingerprint expressed in the form of a ratio known as the “attention ratio” or Theta/Low-Beta (Putman et al. 2010). Also, it is common knowledge throughout the neuro-scientific community that investigations of cerebral activity limited to one area of the brain may offer misleading information regarding complex states such as attention. We have therefore investigated different cerebral areas to study simultaneous brainwave changes. Furthermore, seeing motivation as a state of both cognitive and emotional arousal (Williams and Robert 1997), we have decided to combine the three physiological sensors when empirically evaluating the impact of motivational factors on learners. The idea is to analyze, in a joint venture, both physiological and cerebral signals to determine, or at least estimate, their correlations with motivational factors in SG. To that end, various machine learning algorithms will be constructed using theoretical and empirical results in order to classify learners in two distinct classes based on their self-reported Attention category score of the ARCS model: class “Above” for those with a reported score above group average and class “Below” for the rest. A description of data collection and analysis is given in the experiment and results sections.

**Experiment**

**Methodology.** The study invited participants to play the freely available SG called FoodForce, an initiative of the World Food Program of the United Nations, intended to educate players about the problem of world hunger. FoodForce is comprised of multiple arcade-type missions, each intended at raising players’ awareness towards specific problems regarding world-wide food routing and aid. FoodForce also presents players’ objectives in a short instructional video before the beginning of each mission. A virtual tutor also accompanies the player throughout each mission by offering various tips and lessons relative to the obstacles and goals at hand. Following the signature of a written informed consent form, each participant was placed in front of the computer monitor to play the game. A baseline was also computed before the beginning of the game. A pre-test and post-test were also administered to compare learners’ performance regarding the knowledge presented in the Serious Game. Questions pertained to general knowledge regarding problems of world hunger.

Fig. 1 presents a flow diagram of the five missions.

**Data collection.** The motivational measurement instrument called Instructional Materials Motivation Survey IMMS (Keller 1987) was used following each mission to assess learners’ motivational state. Due to time constraints, we used a short IMMS form which contained 16 out of the 32 items after receiving the advice and approval from Professor Keller. Two cameras were also used to simultaneously record learners’ facial expressions and game progress. Physiological data was also recorded in synchrony to both camera feeds throughout the experiment. The SC and HR sensors were attached to the fingers of participants’ non-dominant hands, leaving the other free for the experimental task. An EEG cap was also conveniently fitted on learners’ heads and each cerebral sensor spot slightly filled with a saline non-sticky solution. EEG recordings were managed by the Thought Technology Pro-Comp Infinity Encoder.

**Data analysis.** EEG was recorded by using a cap with a linked-mastoid reference. The sensors were placed on three selected areas (F3, C3 and Pz) according to the
international 10-20 system (see Fig. 2). The reference sensor was located at Cz and the ground at Fpz. Impedance at each area was maintained below 5 KΩ. The EEG was sampled at a rate of 256 Hz. A Power Spectral Density (PSD) was computed to divide the EEG raw signal into the two following frequencies: Theta (4-8 Hz) and Low-Beta (12-20 Hz) in order to compute the Attention ratio (Theta/Low-Beta) as previously described. To reduce artefacts, participants were asked to minimize eye blinks and muscle movements during recording. A necessary normalization technique (z-score) was applied to all physiological data (HR, GSR and EEG). Indeed, normalizing the data keeps the physiological patterns for individual participants and establishes a common metric for inter-participant comparison. Fig. 2 illustrates the recorded data and the EEG real-time computed ratios.

Figure 2 - Game screen shot and physiological data

Participants. Thirty three volunteers (22 male) took part in the study in return of a fixed compensation. Participant’s mean age was 26.7 ± 4.1 years. Four participants (2 male) were excluded from the EEG analysis due to technical problems at the time of recording. The next section will detail the experimental results and findings.

Experimental Results and Discussion

Before presenting our results, we considered it necessary to quickly explain the statistical approach used in this section. Indeed, we could not rely on the usual parametric statistical tools such as ANOVA and t-test because (1) our sample population is small (29 participants), (2) no justifiable assumptions could be made with regards to the normal distribution of the data, and (3) normality tests run on our data confirmed its non-normal distribution. Hence, non-parametric Friedman’s ANOVA by ranks (counterpart of the parametric one-way ANOVA) and non-parametric Wilcoxon’s signed ranks test (counterpart of paired sample t-test) have been used. However, p-value is interpreted in the same manner in both approaches and to that effect, reported significant p-values were all computed at the 0.05 significance level (95% confidence).

Performance and ARCS. We first report general results regarding learning and motivation for learners. We administered pre-tests and post-tests questionnaires pertaining to the knowledge taught in the Serious Game and compared results. The Wilcoxon signed ranks test showed a significant positive change in learner’s performance in terms of knowledge acquisition (Z = 4.65, p < 0.001). By applying Friedman’s ANOVA by ranks, significant differences for the general motivational score as well as each category of the ARCS model were also observed between missions, except for Relevance (Motivation overall score, F(1,4) = 10.16, p < 0.05; Attention, F(1,4) = 19.51, p < 0.001; Relevance, F(1,4) = 7.38, p = 0.12; Confidence, F(1,4) = 16.8, p < 0.05; Satisfaction, F(1,4) = 10.85, p < 0.05). Non-significance of the Relevance results can be explained by the fact that the instructional videos presented between missions were roughly the same: video segments explaining goals of each mission (R2: Goal orientation) or the real application of each mission in the field (R1: Familiarity). Conversely, the other three categories (Attention, Confidence, and Satisfaction) have been implemented and presented by various game elements throughout the missions. This fact is especially valid for the Attention category which showed the strongest difference and rank. Since Attention showed the most significant results in our study and, more importantly, is one of the most natural and relevant categories to implement in an ITS, we have decided to answer our first research question by investigating in details game elements that may serve as motivational factors related to this category. We have specifically chosen to investigate the fifth mission of the Serious Game that contained various elements regarding all three Attention sub-categories:

<table>
<thead>
<tr>
<th>Attention sub-category</th>
<th>Elements in mission 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1: Variability</td>
<td>Different tasks</td>
</tr>
<tr>
<td>A2: Perceptual arousal</td>
<td>3 alarm triggers</td>
</tr>
<tr>
<td>A3: Inquiry arousal</td>
<td>solving problems</td>
</tr>
</tbody>
</table>

Table 1 – Example of motivational factors in mission 5

Physiological analysis. We have decided to investigate the three alarm triggers as game elements supporting motivation in mission 5. An example of an alarm trigger is shown in Fig. 2. Using the self-reported Attention score following the mission, we have divided participants in two classes: a “Below” class representing participants who reported an Attention score below that of the overall average and an “Above” class presenting the opposite (a score above average). The physiological trends are presented in Fig. 3. Each dot on the graph represents the average difference for a 5 second window computed before and after each alarm (average for 5 seconds after the alarm minus average 5 seconds before the alarm). Fig. 3 shows almost complete opposite trends for all physiological data between the “Below” and “Above” classes, except for SC. A Wilcoxon signed two-sides ranks test was ran on each presented dot and significant results were obtained for all.

The data points towards the fact that the “effect” of an alarm trigger seems to decrease over time. We can see on the HR & SC upper-left sub-figure of Fig. 3 that the effect of those alarms on SC seems to slowly fade after the second alarm, contrary to popular belief. Indeed, one may...
think that intervening with color and sound tends to capture learners' attention, but our findings seem to indicate that this is only partially true. There seems to be a certain "adaptation" on the part of the learner with regards to SC at the very least. Nevertheless, any permanent diagnosis regarding learners' attention level in reaction to an alarm trigger based only on SC at this point may be hasty or even wrong for there are numerous other physiological trends to consider first. Indeed, even if no clear trends were found in HR, the cerebral data provided clarity in distinguishing between the two classes.

In fact, variations in the Attention ratio are clearly evident for both classes. These results seem to show the relevance and importance of adding the EEG in assessing learners' attention change, even more so when this change cannot be clearly established by the use of HR and SC alone. We found numerous occasions when two participants from different classes had the same SC and HR trends but have shown very opposite trends in EEG sites, especially C3 area. An example of this situation is illustrated in Fig. 4: two participants had the same HR and SC trends but only an opposite trend in C3 helped us identify their respective attention classes.

Classification. In order to validate our findings and hopefully answer our second research question, we have chosen to build three classifiers using all the physiological data presented in this section as input and the self-reported attention score after mission 5 as output. The output is a binary value (0 for class "Below" and 1 for class "Above"). The dataset contained 87 instances (29 participants x 3 alarms). The imbalanced dataset problem is a special type of classification problem. We have therefore used an over-sampling method that balances training classes by properly increasing the number of minority class data points. All classifiers were trained on 67% of dataset and validated on 33% of the dataset. Table 2 shows the obtained results of the validation phase. The best overall classification accuracy (73.8%) was achieved using a Multilayer Perceptron with one hidden layer containing 5 neurons. We can see by the results that all three classifiers were able to successfully classify the "Below" class with a high classification accuracy.

The obtained classifier results point towards the fact that it is possible to (1) model learners’ physiological reactions and trends in terms of attention levels towards motivational factors in SG, (2) distinguish between two learner classes relevant to a tutor in an ITS and (3) identify with a relatively high accuracy rate learners susceptible of showing a lack of attention towards a learning process.

Conclusion and Future Work

In this paper, we have assessed the effects of motivational factors relevant to an ITS in SG using the ARCS theoretical model as well as three empirical physiological sensors: HR, SC and EEG. We have successfully answered our first research question by assessing motivational factors in SG related to the Attention category of the ARCS model with the use of physiological sensors. Results have shown that motivational factors seem to elicit specific physiological trends in learners, especially observable in the EEG attention ratios. We then successfully answered our second research question by using these physiological trends to train three different classifiers in order to identify excessive Theta and consequently inattentive state. Conversely, a low Theta/Low-Beta ratio is normally correlated with excessive Low-Beta brainwave activity reflecting normal state in adults. Thus, the EEG Attention ratio generally increases for participants who reported a low Attention category score (class "Below") whereas the same ratio decreases for the learners in the class "Above".

Table 2 – Classifiers

<table>
<thead>
<tr>
<th>Classifier name</th>
<th>Classification accuracy</th>
<th>Above</th>
<th>Below</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multilayer Perceptron</td>
<td>63.6 %</td>
<td>88.9%</td>
<td>73.8%</td>
<td></td>
</tr>
<tr>
<td>Logistic regression</td>
<td>42.9%</td>
<td>93.8%</td>
<td>64.9%</td>
<td></td>
</tr>
<tr>
<td>Naive Bayes</td>
<td>47.1%</td>
<td>87.5%</td>
<td>66.7%</td>
<td></td>
</tr>
</tbody>
</table>

The power of the EEG Attention ratio can be explained by Putman and colleagues (Putman, et al. 2010). According to the authors, a negative correlation exists between the attention ratio and learners’ Attention level. A high Theta/Low-Beta ratio is usually correlated with
learners’ attention level based on self-reported score. The Multilayer Perceptron classifier was able to successfully distinguish attentive from inattentive learners and all three classifiers were able to identify, with a high accuracy, the physiological trends related to inattentive learners. The obtained results are very encouraging to the future integration of such motivational factors in an ITS because (1) it is now possible to assess the impact of the integration of motivational factors related to learners’ attention, (2) we can rely on this assessment as a substitute for self-reports that can disrupt a learning session, and (3) it is possible to enrich the LM with a motivational component based on our results, thus enabling the TM to properly adapt its interventions.

However, one limitation in this work is the assumption that the ARCS categories are independent from each other. Simultaneous factors in SG can be related to different categories of the ARCS. We plan to address these dependencies in a complementary study in order to highlight other distinctive, or even common, physiological patterns related to Relevance, Confidence and Satisfaction.

Acknowledgments

We thank the CRSNG, the FQRSC and the Tunisian Government for their support.

References


