# Assessing Motivational Strategies
# in Serious Games Using Hidden Markov Models

**Lotfi Derbali, Ramla Ghali and Claude Frasson**

Département d'informatique et de recherche opérationnelle

Université de Montréal, 2920 Chemin de la Tour, Montréal, CANADA
{derbalil, ghaliram, frasson}@iro.umontreal.ca

## Abstract

Recent research has extended tutor strategies to model not just interventions to offer information and activities, but also interventions to support learners' wills and motivation. It is important to investigate new ways, intertwined with learners' performance (successful completion of tasks) and judgement (self-report questionnaires), for evaluating tutor intervention strategies. One promising way is the use of physiological sensors. Within this paper, we study some motivational strategies that were implemented in a serious game called HeapMotiv to support learners' performance and motivation. We build several hidden Markov models which use Keller's ARCS model of motivation and electrophysiological data (heart rate HR, skin conductance SC and EEG) and are able to identify physiological patterns correlated with different motivational strategies.

## Introduction

It is widely acknowledged that learners' psychological and cognitive states have an important role in learning. For instance, engagement and motivation or disaffection and boredom obviously affect learners' wills and skills in acquiring new knowledge (Bandura 1986). So, in addition to, but intertwined with, its educational system goal (offer information and activities and support learners' skills in acquiring knowledge), Intelligent Tutoring Systems (ITS) have deployed several actions (or tactics) to assist in maintaining (or even increasing) the learners' will and motivation to learn. In this context, the study of learner's motivation should go through understanding what motivation is, what disrupts it, and how it is supported during the learning process. According to (Keller 2010), "Motivation is generally defined as that which explains the direction and magnitude of behavior, or in other words, it explains what goals people choose to pursue and how they pursue them" (p. 4).

Recently, the benefits of applying digital games as potential learning tools have caught the attention of ITS community. A great interest has been shown to design and experiment Serious Games (SG), for example, computer applications are attempting to combine serious intend with the motivational and goal-based features of games. However, a handful of papers have studied the importance and the design of motivational strategies in SG. This research is concerned with this issue. Indeed, it studies different motivational strategies implemented in a SG called HeapMotiv.

Furthermore, the significant results of recent studies involving physiological sensors to assess motivational learners' states as well as emotional and cognitive strategies (Conati 2002; D'Mello, et al. 2007) make the use of some of these sensors a promising way to evaluate motivational strategies. The aims of this paper are two-fold. First, we assess the effects of adding different motivational strategies in HeapMotiv on learners' performance and motivation. Second, we evaluate the modeling of these strategies by Hidden Markov Models (HMMs). HMMs are fed by psychometric (ARCS model of motivation) and electrophysiological data (heart rate HR, skin conductance SC and electroencephalogram EEG).
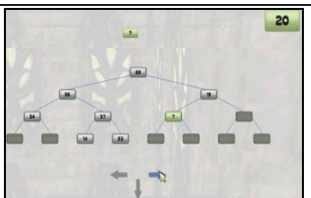
## Related Research

ITS researchers and their counterparts in related fields have argued that learners' negative emotions or non-motivational states such as boredom or disengagement could appear during interaction with computer systems. Several measures (SC, HR, electromyogram EMG, and respiration) have been proposed to deal with those issues. For instance, (Conati 2002; Prendinger et Ishizuka 2005) have developed techniques to detect emotional problems and animated agents stimulating the student to learn better

Other researchers have dealt with the motivation in ITS and other interactive learning environments. They have provided tools to assess motivation and incorporate motivational strategies in their systems, but expended less effort to determine which motivational strategies should be used, how assess their impacts on learners and to what extent they are employed (Boyer, et al. 2008; Rodrigo, et al. 2008). Within the researchers who have tackled this issue, some have found that Serious Games (SG) seemed to show a promising potential from a motivational standpoint. It has been consistently shown that SG include inherent motivational properties and different strategies, allowing them to be used for improving educational applications (Garris, et al. 2002; McNamara, et al. 2009). Furthermore, recent efforts to study the support strategies given to learners have been focused on the question of picking up and modelling these strategies. For instance, (Boyer, et al. 2008) have used HMMs for identifying correlations between local tutoring strategies and student outcomes.

## HeapMotiv and Motivational Strategies

**HeapMotiv.** We developed a SG, called HeapMotiv, to teach binary heap data structure. This SG is a 3D-labyrinth that has many routes with only one path that leads to the final destination. Before obtaining the information signs on the path leading to the final destination, the learner has to play three missions (See Table1) aiming to entertain and educate players about some basic concepts of binary heap.

Table 1. *Different missions of HeapMotiv*

| Mission | Screenshot |
|---|---|
| **Tetris.** It is based on traditional Tetris game. A learner has to move nodes during their falling using the arrows to fill a binary tree without violating the heap property. |  |
| **Shoot.** It is based on shooter games. A learner has to spot violations of shape and heap properties and has then to fix these violations by shooting misplaced nodes. |  |
| **Sort.** It is based on a comparison-sort algorithm. It begins by building a binary heap out of the data set and then removing the biggest item to obtain a sorted array. |  |

**Motivational strategies:** We conduct a survey to determine what motivational strategies to implement in HeapMotiv and to achieve our goal in this research. We use then, the ARCS model of motivation (Keller 1987) to design different motivational strategies. Indeed, John Keller used existing research on psychological motivation to identify four categories of motivation: Attention, Relevance, Confidence, and Satisfaction. He also defines four different motivational strategies associated to each category (Keller 2010): Attention getting strategies $Stg_A$, Relevance producing strategies $Stg_R$, Confidence building strategies $Stg_C$, and Satisfaction generating strategies $Stg_S$.

In the version of HeapMotiv with motivational strategies, a $Stg_A$ is based on submitting challenges as time and errors constraints: (1) a time constraint for each level of difficulty: unlimited, 90 seconds, and 45 seconds for easy, normal, and hard level respectively, and (2) a wild card representing the number of accepted errors committed by the player: unlimited, 3 wild cards, and 1 wild card for easy, normal, and hard level respectively. A $Stg_R$ has been designed before the beginning of each mission by presenting an instructional video to explain and aware learners of the main goal of the mission and its relation to the binary heap data structure. Then, this version of HeapMotiv integrates a $Stg_C$ which allows learners to control the level of each mission (easy, normal, and hard) and to possibly repeat the mission with the same or a different level (at most six trials). Finally, a virtual companion "Sinbad" applies a $Stg_S$ by providing feedback on learners' performance when they find a way out of the labyrinth and meet "Sinbad".

## Hidden Markov Models

**HMM.** The four motivational strategies described above are investigated to feed four Hidden Markov Models (HMMs). HMM is a probabilistic model defined by a tuple $\lambda=(n, m, A, \pi, B)$, where $n$ is the number of hidden states, $m$ is the number of observable states, $A$ is the state transition probability, $\pi$ is the initial state probability and $B$ is the emission probability density function of each state. It is a model representing probability distributions over sequences of observations. (Rabiner 1990) defined three main problems for HMM: evaluation problem, decoding problem and training problem.

**Classification process design:** In our case, we investigate algorithms of the training problem (Baum-Welch B&W) and the evaluation problem (Forward-Backward F&B) to model the four motivational strategies presented above. Indeed, we build 4 HMMs (one for each strategy) and train them using B&W algorithm. For a question of simplicity, binary data encoding is used to feed each HMM. So, each HMM $\lambda$ has 2 hidden states (defined by ARCS model: low and high states of motivation) and 8 observed states (defined by low and high levels of SC, HR, and EEG respectively). Transition and emission probabilities of the model $\lambda$ for a strategy $Stg$ are estimated by applying the B&W algorithm to a set of recorded data when $Stg$ is presented to learners. F&B algorithm is then used to evaluate each obtained model. Given a model $\lambda$ and an observed sequence $Obs_j$, this algorithm is used to find the probability of the observed sequence given the model, $P(Obs_j|\lambda)$. Our approach of evaluation is to run different

models ($\lambda_A$, $\lambda_R$, $\lambda_C$, and $\lambda_S$) on each new observed sequence $Obs_j$ and attribute $Obs_j$ to the maximum likelihood model $\lambda^* = argmax(P(Obs_j|\lambda_i))$, where $i \in \{A, R, C, S\}$. The classification process using different HMMs as well as B&W and F&B algorithms is presented in Figure 1.
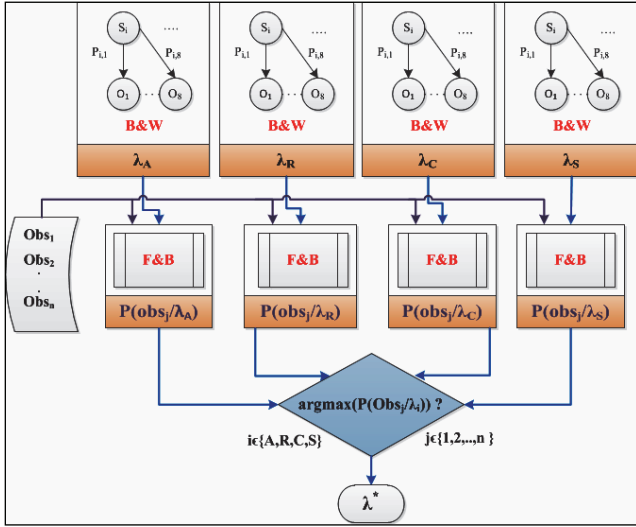


*Figure 1. HMMs based classification*

## Experimental Methodology

**Subjects.** A total of 41 subjects (21 females, 20 males) were invited to play our SG HeapMotiv in return of a fixed compensation. Subjects were randomly attributed to the control group CTR_G (*10 females, 10 males, mean age=23.7, SD=6.8, HeapMotiv without motivational strategies*) or to the experimental group EXP_G (*11 females, 10 males, mean age=25.3, SD=4.5, HeapMotiv with motivational strategies*).

**Data Collection:** Each participant was placed in front of the computer monitor to play our SG, HeapMotiv, using only the mouse. It was then possible to record the skin conductance (SC) and heart rate (HR) data by attaching the appropriate sensors to the fingers of subject's non-dominant hand. An electroencephalogram (EEG) cap was also fitted on participant's head to record the brain's spontaneous electrical activity during the whole of the experiment. We also computed 60s-baseline before the beginning of the game. It should be noted that HR, SC and EEG recordings were managed by the thought Technology Pro-Comp Infiniti Encoder. Two cameras were also used to simultaneously record subject's facial expressions and game progress. Subjects were asked to minimize eye blinks and muscle movements during physiological recordings.

In addition, we used a short motivational measurement instrument called, Instructional Materials Motivational Survey (IMMS) to assess learner motivational state after each mission played in HeapMotiv. This survey derived from the ARCS motivation model and consisted of 16 items (5-point Likert-type). 10 pre-test and 10 post-test quizzes

about general knowledge of binary tree and knowledge presented in HeapMotiv were also administered to compare learners' performance. In the modeling step of different HMMs, the experimental group EXP_G includes a total of 268 sequences of observations with different session window sizes (duration (sec)) distributed over the four motivational strategies: 92 sequences for $Stg_A$, 63 sequences for $Stg_R$, 92 sequences for $Stg_C$, and 21 sequences for $Stg_S$.

## Experimental Results

Since we intend to study several motivational strategies in different missions within HeapMotiv game, we firstly evaluated the effects of these strategies on learners' performance as well as their motivation. We conducted statistical tests and we obtained several results regarding knowledge acquisition (pre-test and post-test) and learners' motivation (ARCS scores).

**Performance:** Scores of pre-test and post-test were used to compare performances of CTR_G and EXP_G. The results of Wilcoxon signed ranks test displayed in Table 2 showed a significant difference between the subjects' scores of the pre- and post-tests in terms of knowledge acquisition (EXP_G: *Z=-3.756, p=0.000*; CTR_G: *Z=-3.348, p=0.001*). Number of correct answers after finishing the game is significantly higher than that of correct answers before start playing within the two groups. However, learners' performance in terms of knowledge acquisition in CTR_G was lower, on the average, than learners' performance in EXP_G. Indeed, the results of Mann-Whitney U test showed a significant difference (*Z=-2.088, p=0.037*) between their average ranks. CTR_G had an average rank of 17.24, while EXP_G had an average rank of 24.95. The obtained results were in the expected direction: motivational strategies experimented by EXP_G's learners significantly participate in improving their performance in comparison with performance of CTR_G's learners.

**Motivation:** Several Mann-Whitney U tests were also conducted to evaluate the hypothesis that the overall motivation and each category of the ARCS model differ between CTR_G and EXP_G. The results were again in the expected direction and significant (Motivation: *Z=-3.281, p=0.001*; Relevance: *Z=-3.209, p=0.001*; Confidence: *Z=-3.745, p=0.000*; Satisfaction: *Z=-2.935, p=0.003*), expect for the attention category (*Z=-1.897, p=0.058*). Reported ARCS scores in EXP_G were higher than those in CTR_G. These results excluded the hypothesis that learners' motivation remains roughly the same between mission periods and with/without motivational strategies. In addition, even non-significant difference was found for the attention category, learners of the two groups reported high scores of attention (*mean=14.11, SD=3.96, ranked second in order after the confidence category*). Thus, we can only say that learners were highly attentive during interactions

with HeapMotiv, but we cannot exclude that the motivational strategies might act on learners' attention. These obtained results opened up opportunities to answer our main research question by studying each of the four motivational strategies during interaction with HeapMotiv.

**HMM.** Our work used HMMs to drive a classification process of motivational strategies. Physiological data from the group EXP_G (learners were experimented HeapMotiv with motivational strategies) were used to train (using 67% of data) and test (using 33% of data) different models. The standard recall, precision and F-score metrics were calculated for the four HMMs. The recall value was the fraction of correctly classified strategy *Stg* as compared to the total number of strategy *Stg* in the test data. The precision value was the fraction of correctly classified strategy *Stg* as compared to the total number of classified strategy *Stg*. F-score was the weighted harmonic mean of precision and recall.

**Table 2.** *Results of different HMMs*

| Model | Recall (%) | | Precision (%) | | F-score (%) | |
|---|---|---|---|---|---|---|
| | *Train* | *Test* | *Train* | *Test* | *Train* | *Test* |
| $\lambda_A$ | 85.25 | 74.19 | 86.67 | 82.14 | 85.95 | 77.97 |
| $\lambda_R$ | 52.38 | 52.38 | 51.16 | 47.83 | 51.76 | 50.00 |
| $\lambda_C$ | 73.77 | 77.42 | 75.00 | 80.00 | 74.38 | 78.69 |
| $\lambda_S$ | 42.86 | 28.57 | 40.91 | 22.22 | 41.86 | 25.00 |

As shown in Table 2, F-score values ranged from 42% to 86% for the training step and indicated a moderately high relationship between the predictors (physiological data) and the dependent variable (motivational strategy). The test step confirms this finding with the application of the trained models on new observations and focuses on the best performance of both models $\lambda_A$ (*F1=78%*) and $\lambda_C$ (*F1=79%*). This confirms our results regarding the confidence category which had the most significant result and shows effectively that some specific physiological patterns were characterized learners dealing with the strategy $Stg_C$. This also emphasizes our last observation regarding the attention category and shows specific physiological patterns correlated with the strategy $Stg_A$.

However, relatively low performances of relevance and satisfaction models ($\lambda_R$ and $\lambda_S$) were found. It is difficult to explain this result, but it may be related to the fact that data were insufficient to consistently estimate model parameters and then test these models on new sequences of observations. For example, only 7 observations have been used in the test step of the relevance strategy $Stg_R$ (an instructional video presenting the relation between the current mission and properties of binary heap). In addition, these strategies were not repeated a lot (exactly one $Stg_R$ before each mission and one $Stg_S$ and the end of game) in comparison with attention and confidence strategies that had large session window sizes and where learners repeated them as many as they liked (with a maximum of six trials).

## Conclusion and future work

In this paper, we have studied different motivational strategies implemented in HeapMotiv. Statistical and physiological analyses using HMM models have given some insights into the assessment of these strategies during gameplay and have shown that physiological parameters can feed these models. The obtained results are very encouraging to study additional intervention strategies within serious games and ITS. One important implication of our work is that it may be possible to enrich intelligent systems with an objective tool to assess tutor intervention strategies and their effects on learners' motivation.

However, one limitation in this work is the assumption that the ARCS categories are independent from each other. One possible extension of the present work would be to consider dependencies between ARCS categories. In addition, we can extend different HMMs to more than two hidden states of motivation and finally build one complex HMM for all different motivational strategies.

## References

Bandura, A. 1986. Social foundations of thought and action: A social cognitive theory. Englewood Cliffs, NJ: Prentice-Hall.

Boyer, K., Phillips, R., Wallis, M., Vouk, M., et Lester, J. 2008. Balancing Cognitive and Motivational Scaffolding in Tutorial Dialogue. *In* the 9th International Conference on Intelligent Tutoring Systems. Pp. 239-249. Montreal, Canada: Springer.

Conati, C. 2002. Probabilistic Assessment of User's Emotions in Educational Games. Applied Artificial Intelligence 16:555-575.

D'Mello, S., Graesser, A., et Picard, R. W. 2007. Toward an Affect-Sensitive AutoTutor. Intelligent Systems, IEEE 22(4):53-61.

Garris, R., Ahlers, R., et Driskell, J. E. 2002. Games, motivation, and learning: a research and practice model. Simulation & Gaming 33(4):441-467.

Keller, J. M. 1987. Development and use of the ARCS model of motivational design. Journal of Instructional Development 10(3):2-10.

Keller, J. M. 2010. Motivational Design for Learning and Performance: The ARCS Model Approach: Springer.

McNamara, D. S., Jackson, G. T., et Graesser, A. C. 2009. Intelligent tutoring and games (iTaG). *In* Workshop on Intelligent Educational Games at the 14th International Conference on Artificial Intelligence in Education. Pp. 1-10. Brighton, UK: IOS Press.

Prendinger, H., et Ishizuka, M. 2005. The Empathic Companion: A character-based interface that addresses users' affective states. Applied Artificial Intelligence 19(3-4):267-285.

Rabiner, L. R. 1990. A tutorial on hidden Markov models and selected applications in speech recognition. *In* Readings in speech recognition. W. Alex et L. Kai-Fu, eds. Pp. 267-296: Morgan Kaufmann Publishers Inc.

Rodrigo, M. M. T., Rebolledo-Mendez, G., Baker, R. S. J. d., du Boulay, B., Sugay, J. O., Lim, S. A. L., Espejo-Lahoz, M. B., et Luckin, R. 2008. The Effects of Motivational Modeling on Affect in an Intelligent Tutoring System. *In* the International Conference on Computers in Education. Pp. 57-64.