

Université de Montréal

Détection et amélioration de l'état cognitif de l'apprenant

par

Ramla Ghali

Département d'Informatique et de Recherche Opérationnelle

Faculté des Arts et des Sciences

Thèse présentée

en vue de l'obtention du grade de Philosophiae Doctor (Ph.D.)

en Informatique

Décembre 2016

© Ramla Ghali, 2016

Université de Montréal
Faculté des Arts et des Sciences

Cette thèse intitulée :

Détection et amélioration de l'état cognitif de l'apprenant

Présentée par :
Ramla Ghali

a été évaluée par un jury composé de :
Jean Meunier, Président et représentant du doyen
Claude Frasson, directeur de recherche
Roger Nkambou, membre du jury
Stefano A. Cerri, examinateur externe

Résumé

Cette thèse vise à détecter et améliorer l'état cognitif de l'apprenant. Cet état est défini par la capacité d'acquérir de nouvelles connaissances et de les stocker dans la mémoire. Nous nous sommes essentiellement intéressés à améliorer le raisonnement des apprenants, et ceci dans trois environnements : environnement purement cognitif Logique, jeu sérieux *LewiSpace* et jeu sérieux intelligent *Inertia*. La détection de cet état se fait essentiellement par des mesures physiologiques (en particulier les électroencéphalogrammes) afin d'avoir une idée sur les interactions des apprenants et l'évolution de leurs états mentaux. L'amélioration des performances des apprenants et de leur raisonnement est une clé pour la réussite de l'apprentissage.

Dans une première partie, nous présentons l'implémentation de l'environnement cognitif logique. Nous décrivons des statistiques faites sur cet environnement. Nous avons collecté durant une étude expérimentale les données sur l'engagement, la charge cognitive et la distraction. Ces trois mesures se sont montrées efficaces pour la classification et la prédiction des performances des apprenants.

Dans une deuxième partie, nous décrivons le jeu *Lewispace* pour l'apprentissage des diagrammes de Lewis. Nous avons mené une étude expérimentale et collecté les données des électroencéphalogrammes, des émotions et des traceurs de regard. Nous avons montré qu'il est possible de prédire le besoin d'aide dans cet environnement grâce à ces mesures physiologiques et des algorithmes d'apprentissage machine.

Dans une troisième partie, nous clôturons la thèse en présentant des stratégies d'aide intégrées dans un jeu virtuel *Inertia* (jeu de physique). Cette dernière s'adapte selon deux mesures extraites des électroencéphalogrammes (l'engagement et la frustration). Nous avons montré que ce jeu permet d'augmenter le taux de réussite dans ses missions, la performance globale et par conséquent améliorer l'état cognitif de l'apprenant.

Mots-clés : Environnement cognitif, jeu éducatif, jeu en réalité virtuelle, électroencéphalogramme, traceur de regard, activité électrodermale, émotions, aide à l'apprentissage, apprentissage machine

Abstract

This thesis aims at detecting and enhancing the cognitive state of a learner. This state is measured by the ability to acquire new knowledge and store it in memory. Focusing on three types of environments to enhance reasoning: environment Logic, serious game LewiSpace and intelligent serious game Inertia. Physiological measures (in particular the electroencephalograms) have been taken in order to measure learners' engagement and mental states. Improving learners' reasoning is key for successful learning process.

In a first part, we present the implementation of logic environment. We present statistics on this environment, with data collected during an experimental study. Three types of data: engagement, workload and distraction, these measures were effective and can predict and classify learner's performance.

In a second part, we describe the LewiSpace game, aimed at teaching Lewis diagrams. We conducted an experimental study and collected data from electroencephalograms, emotions and eye-tracking software. Combined with machine learning algorithms, it is possible to anticipate a learner's need for help using these data.

In a third part, we finish by presenting some assistance strategies in a virtual reality game called Inertia (to teach Physics). The latter adapts according to two measures extracted from electroencephalograms (frustration and engagement). Based on our study, we were able to enhance the learner's success rate on game missions, by improving its cognitive state.

Keywords : cognitive environment, educational game, virtual reality game, electroencephalogram, eye tracker, electrodermal activity, emotions, need of help, machine learning.

Table des matières

Résumé.....	ii
Abstract.....	vi
Table des matières.....	viii
Liste des tableaux.....	xii
Liste des figures.....	xiv
Liste des abréviations.....	xvi
Remerciements.....	xx
Chapitre 1 Introduction.....	2
1.1 Contexte.....	2
1.2 Problématique.....	5
1.3 Objectifs de recherche.....	7
1.4 Organisation du document.....	11
Chapitre 2 État de l’art.....	14
2.1 Introduction.....	14
2.3 État Mental.....	19
2.4 Types de raisonnement.....	24
2.4.1 Raisonnement basé sur le calcul et la logique.....	25
2.4.2 Raisonnement intuitif.....	29
2.5 Des jeux sérieux aux jeux sérieux intelligents.....	30
2.5.1 Jeux Sérieux.....	30
2.5.2 Jeux éducatifs.....	34
2.5.3 Jeux Sérieux Intelligents.....	38
2.6 Senseurs physiologiques.....	39
2.6.1 Électroencéphalogramme.....	39
2.6.2 Activité Électrodermale.....	41
2.6.3 Traceurs de regard.....	42
2.7 Adaptation dans les jeux sérieux intelligents.....	44

2.7.1	Adaptation selon l'aspect narratif	46
2.7.2	Adaptation selon les émotions	48
2.7.3	Adaptation en fonction de la reconnaissance des objectifs.....	49
2.7.4	Adaptation en fonction des connaissances de l'apprenant.....	50
2.7.5	Adaptation selon les senseurs physiologiques	51
2.7.6	Adaptation selon le niveau de difficulté dans le jeu	52
2.8	Conclusion	53
Chapitre 3 Une Étude empirique dans l'environnement purement cognitif logique		
.....		56
3.1	Introduction.....	61
3.2	Previous Works.....	62
3.3	Description of our cognitive tasks	64
3.3.1	Memory.....	64
3.3.2	Concentration.....	65
3.3.3	Reasoning.....	65
3.4	Description of the experiment, gathering the data and our approach	66
3.4.1	Experimental Protocol	66
3.4.2	Data recording and preprocessing.....	67
3.4.3	Our approach.....	68
3.5	Experimental Results	69
3.5.1	EEG and GSR metrics evolution in learning.....	69
3.5.2	Multiple regression for predicting learners' scores	73
3.6	Conclusion	75
Chapitre 4 La classification et la prédiction des performances de l'apprenant		
.....		78
4.1	Introduction.....	83
4.2	Related Works.....	84
4.3	Justification and description of logic environment.....	87
4.3.1	Memory.....	88
4.3.2	Concentration.....	88

4.3.3 Reasoning.....	89
4.4 Experimental Study.....	91
4.4.1 Description of the experiment.....	91
4.4.2 Data Pre-processing for Classification and Regression.....	93
4.5 Algorithms, Results and Discussion.....	94
4.5.1 A brief Description of the Algorithms.....	94
4.5.2 Results and discussion.....	95
4.6 Conclusions.....	100
Chapitre 5 Description et étude générale de l’environnement éducatif 2D LewiSpace	
.....	102
5.1 Introduction.....	107
5.2 Previous Work.....	109
3.3 LewiSpace: an educational puzzle game.....	111
3.3.1 A description of the environment.....	111
5.3.2 A description of the given instructions and rules.....	115
5.4 Experiment and data preprocessing.....	117
5.4.1 Experiment.....	117
5.4.2 Data pre-processing.....	118
5.5 Results.....	119
5.5.1 Descriptive Results.....	119
5.5.2 Selection of the best machine learning model.....	121
5.5.3 Comparison of the importance of each sensor and the Big Five.....	122
5.6 Conclusions.....	124
Chapitre 6 Une approche physiologique pour la prédiction du besoin d’aide dans	
l’environnement éducatif LewiSpace.....	126
6.1 Introduction.....	131
6.2 Related work.....	132
6.3 A brief description of LewiSpace game.....	133
6.4 Our approach: physiological sensing.....	134
6.5 Data preprocessing, results and discussion.....	136

6.6 Conclusion	138
Chapitre 7 Utilisation de stratégies d'aide dans l'environnement virtuel Inertia	140
7.1 Introduction.....	145
7.2 Improving Intuitive Reasoning	146
7.2.1 Intuitive reasoning	146
7.2.2 Virtual Reality Games.....	147
7.2.3 Assistance strategies	148
7.3 Inertia: a Virtual Reality Game for Learning Dynamic Rules	149
7.4 Assistance Strategies in Inertia	151
7.4 Experiment.....	153
7.5 Results and discussions.....	154
7.6 Conclusions.....	157
Chapitre 8 Conclusion.....	158
8.1 Contributions.....	158
8.2 Travaux futurs.....	160
Bibliographie.....	i
Publications.....	xvii

Liste des tableaux

Table 2. 1. Les tâches de classification des états mentaux	23
Table 2. 2 Exemples de jeux sérieux	33
Table 3. 1 Distribution of EEG metrics according to category and type of a task	71
Table 3. 2 Results of Person correlations.....	74
Table 3. 3 Multiple regression model to predict learners' scores	75
Table 4. 1 Classification and regression for different models and features.....	96
Table 4. 2 Confusion matrix for a Random Forest classifier	97
Table 5. 1 Instructions and rules presented in LewiSpace mentionned according to the missions	116
Table 5. 2 The average success time per mission	121
Table 5. 3 Feature selection through classification accuracies	122
Table 5. 4 Confusion matrix for a logistic regression model with Emotiv headset features ..	123
Table 6. 1 Missions' distribution in LewiSpace game.....	133
Table 6. 2 Balanced accuracies depending on feature selection.....	138
Table 6. 3 Confusion matrix for a personalized model.....	138
Table 7. 1 Examples of strategies in bowling 2 game	153
Table 7. 2 Descriptive statistics of scores in Inertia V0	156
Table 7. 3 Descriptive Statistics of scores in Inertia V1	156

Liste des figures

Figure 2. 1 Les différents lobes du cerveau humain	17
Figure 2. 2 Les différentes mémoires (Atkinson et Shiffrin, 1968).....	18
Figure 2. 3 Représentation des émotions selon (Prendinger et Ishizuka, 2005).....	20
Figure 2. 4 Positions des senseurs dans le casque B-Alert X 10	21
Figure 2. 5 Interface graphique de <i>B-Alert X10</i> montrant la variation en temps réel des états mentaux.....	22
Figure 2. 6 Distribution des senseurs dans le casque <i>Emotiv EPOC</i>	23
Figure 2. 7. Exemple de deux matrices progressives de Raven (Vezneva, 2011)	29
Figure 2. 8 Interface du jeu Oxo	32
Figure 2. 9 Les différentes missions dans le jeu <i>HeapMotiv</i>	35
Figure 2. 10 Interface du jeu <i>PrimeClimb</i> (Conati, 2002)	36
Figure 2. 11 Capture d'écran du jeu <i>Physics Palyground</i> (Kai et al., 2015)	37
Figure 2. 12 Environnement virtuel <i>Crystal Island</i> (Min et al., 2016)	38
Figure 2. 13 Les positions des électrodes dans le système 10/20 (Klem et al., 1999).....	40
Figure 2. 14 Illustration de l'état de stress mesuré par l'EDA (Setz et al., 2010)	42
Figure 2. 15 Exemples de saccades et de fixations	44
Figure 2. 16 Critères des techniques d'adaptation dans les JSI (Hocine et al., 2011)	45
Figure 3. 1 Feature Match Task	65
Figure 3. 2 Example of Intuitive Reasoning Task	66
Figure 3. 3 Experimental proces	67
Figure 3. 4 Our proposed approach.....	68
Figure 3. 5 Distraction distribution among learner's	69
Figure 3. 6 Engagement distribution among learners	70
Figure 3. 7 Workload distribution among learners	70
Figure 3. 8 GSR distribution among the participants	73
Figure 4. 1 The six dimensions of Bloom's taxonomy of cognitive tasks.....	87

Figure 4. 2 Example of Rotation task	89
Figure 4. 3 Example of OO task: the odd image is number 6 counting from the left side of the top image.....	90
Figure 4. 4 Example of Intuitive Reasoning task.....	90
Figure 4. 5 Distribution of the tasks of Logic environment according to Bloom’s taxonomy.	91
Figure 4. 6 Experimental process and adopted approach	92
Figure 4. 7 Feature Importance for a Random Forest classifier	99
Figure 4. 8 Feature importance for a Random Forest regressor	100
Figure 5. 1 The beginning of the game	112
Figure 5. 2 Atoms in the environment (left) and using a Lewis diagram tool to produce H ₂ O molecule (right).....	113
Figure 5. 3 Melting a tunnel’s obstruction using a methane torch.....	114
Figure 5. 4 The end of the game	115
Figure 5. 5 The experimental process	118
Figure 5. 6 Statistics of failure per mission	120
Figure 5. 7 ROC curves for the 33 participants	124
Figure 6. 1 Lewis diagram of the sulfuric acid	134
Figure 6. 2 The physiological sensing approach.....	136
Figure 7. 1 Details of Inertia	150
Figure 7. 2 An overview of the distribution of messages and solutions in Inertia	152
Figure 7. 3 The experimental process	154
Figure 7. 4 Descriptive Results in Inertia V0	155
Figure 7. 5 Descriptive Results in Inertia V1	155

Liste des abréviations

EEG	Électroencéphalogramme
IA	Intelligence Artificielle
EDA	Activité électrodermale de la peau
GSR	Galvanic Skin Response
HR	Heart Rate
STI	Système Tutoriel Intelligent
JS	Jeu Sérieux
JSI	Jeu Sérieux Intelligent
JRV	Jeu en réalité virtuelle
RA	Raisonnement Analytique
RI	Raisonnement Intuitif
SA	Stratégie d'aide
EC	État Cognitif
EM	État Mental
ABM	Advanced Brain Monitoring
FDS	Forward Digit Span
BDS	Backward Digit Span
FM	Feature Match
RT	Rotation
AA	Arithmetic Addition
OO	Odd One Out
IR	Intuitive Reasoning
D	Distraction
E	Engagement
W	Workload
TI	Task Information
DL	Difficulty Level

Du Duration
SVM Support Vector machines
RF Random Forest
GB Gradient Boosting
LR Linear Regression
OCC modèle de représentation des émotions proposé par Orthonoy, Clore et Collins (1988)
ANOVA ANalysis of Variance
t-test Student Test
IR Intuitive Reasoning
VR Virtual Reality
RBF Radial Basis Function
ROC Receiver Operator Curves
KNN K Nearest Neighbors
ML Machine Learning
EDSAC Electronic Delay Storage Automatic Calculator
AVC Accident Vasculaire Cérébral

À mon cher père Nejib et ma fille Jasmine

Remerciements

Je remercie au premier lieu mon professeur Claude Frasson qui m'a appris les astuces de recherches durant mes parcours de maîtrise et de doctorat, de m'avoir soutenue et conseillée pour la réussite et la rédaction de plus d'une dizaine d'articles scientifiques. Merci Claude de m'avoir appris les secrets du métier.

Mes remerciements s'adressent aussi aux membres du jury qui ont accepté d'évaluer ce travail, professeur Jean Meunier et professeur Roger Nkambou. En particulier, je tiens à remercier sincèrement le professeur Stefano Cerri pour sa patience et pour le temps qu'il a consacré à la lecture de cette thèse. Vos suggestions m'ont permis d'améliorer considérablement la qualité de cette thèse.

Je tiens à remercier le FQRNT (Fond Québécois de Recherche sur la Nature et technologies) de m'avoir octroyé la bourse d'excellence pour une certaine durée de ma thèse, le CRSH (plus précisément le projet LEADS (*Learning Environment Across Disciplines*)), et le NSERC (*National Science and Engineering Research Council*) pour avoir supporté le projet.

J'exprime ainsi mes sincères remerciements à ma mère Leila qui m'a encouragée à continuer mon parcours académique pendant plusieurs discussions téléphoniques. Merci aussi à ma sœur Asma qui apparaît plus vers la fin de mon chemin. Je remercie aussi mon mari Hatem pour sa patience tout au long de mon parcours de recherche.

Je tiens aussi à remercier l'équipe technique de l'Université de Montréal, en particulier Samy, Maxime, Mohamed et Djamel qui m'ont aidé à réparer mon ordinateur à plusieurs reprises. Je remercie également mes collègues de travail, Samira, Sébastien, Lotfi, Sahbi et Hamdi pour les discussions et le travail en commun.

Je remercie très chaleureusement Monsieur Thierry Trotin et Madame Afifa Monastiri d'avoir eu la patience de relire ce manuscrit et de faire les premières corrections des fautes d'orthographe.

Finalement, je voudrais remercier tous les membres de ma famille que je n'ai pas mentionné (en particulier ma sœur Lina, ma tante Mounira et mon oncle Hatem) et mes collègues d'étude du groupe Héron.

Chapitre 1

Introduction

1.1 Contexte

Notre sujet de recherche se situe dans les travaux menés par le groupe de recherche Héron, un laboratoire multidisciplinaire qui s'intéresse à l'intelligence artificielle, la psychologie, la neuroscience, etc. Dans notre travail, nous nous sommes intéressés plus particulièrement au domaine de l'intelligence artificielle et à ses applications. Parmi ces applications, nous avons porté plus d'attention aux environnements cognitifs et aux jeux sérieux dédiés à l'apprentissage (jeux éducatifs 3D). De nos jours, l'utilisation d'internet comme moyen de recherche d'information et d'acquisition de connaissances devient de plus en plus répandu dans toutes les catégories de population (enfants, adultes, etc.). Il importe donc de porter plus d'attention à la conception des futurs systèmes d'apprentissage et de formation à distance. Parmi ces systèmes, nous trouvons essentiellement deux environnements d'apprentissage qui deviennent de plus en plus répandus : (1) les Systèmes Tutoriels Intelligents (STI) qui sont des systèmes ayant pour objectif principal d'adapter le contenu de l'apprentissage en fonction des caractéristiques individuelles de l'apprenant et de lui fournir une rétroaction continue en fonction de ses besoins (Beck et al., 2008) et (2) les Jeux Sérieux (JS) qui se caractérisent essentiellement par leur capacité de combiner deux aspects fondamentaux (Alvarez et Michaud, 2008) : aspect éducatif (notions cognitives à apprendre, notions à assimiler, etc.) et aspect ludique (plaisir dans le jeu, motivation, attraction, immersion, etc.). Ce dernier aspect fait de cet environnement virtuel un outil de plus en plus efficace car il permet à la fois d'attirer l'apprenant et de le garder dans un état d'immersion et de plaisir. Ces états, qui sont des états mentaux (du cerveau) apparaissant

dans une situation donnée suite à un ensemble de conditions bien déterminées, peuvent être favorables à l'apprentissage et par conséquent ils facilitent la construction¹ de nouvelles connaissances à l'apprenant (Piaget, 1936; Vygotsky, 1934).

Bien que ces deux environnements constituent un moyen d'apprentissage de plus en plus répandu, une meilleure compréhension et étude des caractéristiques de chacun d'eux est utile afin de distinguer leurs lacunes et avantages et de déterminer quand privilégier un environnement par rapport à l'autre. En effet, certains chercheurs (McNamara et Jackson, 2010; Prensky, 2007) pensent que les étudiants ne veulent pas interagir avec les STI et préfèrent par contre interagir avec les JS car ces derniers permettent d'améliorer les performances de l'apprenant (Lester, Ha, et al., 2013; Lester et al., 2014) et par conséquent ils ont des effets positifs sur l'apprentissage (Dempsey et al., 1994; Vogel et al., 2006). D'autres chercheurs pensent que les jeux éducatifs peuvent avoir à la fois des effets positifs et négatifs (Hays, 2005; Mitchell et al., 2011).

Dans cette thèse, nous nous intéressons à ces deux derniers types d'environnements. Dans un premier temps, nous nous intéressons à un type particulier de STI soit les environnements cognitifs ayant pour but d'apprendre différentes catégories de tâches cognitives (mémorisation, concentration et raisonnement analogique) afin d'étudier si la nature et le type d'une tâche cognitive donnée jouent un rôle important sur les performances de l'apprenant. Dans un deuxième temps, nous nous intéressons à un type particulier de JS soit les jeux éducatifs pour utilisation sur un ordinateur et en réalité virtuelle. Les JS sont un outil d'apprentissage très efficace car ils intègrent plusieurs caractéristiques essentielles pour l'acquisition des connaissances. Parmi ces caractéristiques, nous distinguons les critères basés sur les jeux (*Game-Based Features*) qui permettent de susciter chez les apprenants certains états psychiques favorables à l'apprentissage. Plus précisément, ils augmentent les états de motivation et d'engagement (Lester et al., 2014; McNamara et Jackson, 2010). Selon McNamara et ses

¹ Dans cette thèse, nous optons pour une approche constructiviste d'apprentissage où on apprend quand on construit les connaissances, qui est contraire à l'approche behavioriste où l'apprentissage consiste à acquérir un nouveau comportement ou modifier un comportement préexistant.

collègues, il existe cinq caractéristiques fondamentales permettant de distinguer les JS par rapport aux autres environnements d'apprentissage:

- **Commentaires (*Feedback*)** : cette caractéristique consiste à offrir à l'apprenant un commentaire précis, intelligent et motivationnel. Parmi les aspects importants à considérer dans le commentaire : le temps d'action, le contrôle de contenu et la nature du commentaire (visuel ou audio).
- **Incitations (*Incentives*)** : il s'agit de favoriser les aspects de primes et de récompenses qui sont liés à la motivation extrinsèque et qui ont un effet direct sur l'auto-efficacité, l'engagement et l'intérêt. Par exemple, au cours du jeu, l'apprenant peut gagner des points qui peuvent ensuite être dépensés selon ses objectifs (changer un avatar, changer l'arrière-plan, etc.).
- **Difficulté de la tâche (*Task difficulty*)** : cette caractéristique consiste à faire varier la difficulté de la tâche proposée à l'apprenant et l'ajuster selon ses compétences. Les auteurs parlent ici de la sélection de matériel selon la zone adéquate de développement de l'apprenant (*ZPD : Zone of Proximal Development*), c'est-à-dire sélectionner le contenu selon les aptitudes de l'apprenant et les connaissances préalablement acquises.
- **Contrôle (*Control*)** : il s'agit ici de laisser à l'apprenant l'initiative de gérer et contrôler le système tel que le contrôle de l'aspect de l'environnement (changer la couleur de l'arrière-plan, changer l'avatar, etc.), le paramétrage de ses buts et sous-buts (l'apprenant peut viser un certain but ou niveau à atteindre selon les indications affichées), etc. Cet aspect est aussi souligné par (Garris et al., 2002).
- **Environnement (*Environment*)** : les JS se caractérisent par quatre aspects au niveau de leur environnement de développement : esthétique (visualisation en 3D, icônes, images, etc.), avatars (ce sont les agents pédagogiques permettant de capter l'intérêt de l'apprenant), multimédia (disposer de différents modes de présentation) et narrative (raconter des faits).

L'environnement résultant qui combine ces critères est appelé **Jeux Sérieux Intelligent** (JSI) ou encore *Intelligent Tutoring and Games* (ITaG). L'idée consiste ainsi à intégrer les

caractéristiques basées sur les jeux (*Game-Based Features*) dans les systèmes tutoriels intelligents pour créer en quelque sorte des jeux sérieux intelligents. Dans notre cas, nous intéressons à créer des JSI en s'inspirant des critères mentionnés.

1.2 Problématique

Pour garantir le critère d'intelligence dans les environnements cognitifs ou jeux vidéo, il est nécessaire de bien détecter et comprendre l'évolution des états cognitifs de l'apprenant. Nous souhaitons dans cette thèse améliorer l'état cognitif de l'apprenant selon son état mental. Nous définirons plus en détail ces deux états plus loin mais de façon simple **l'état cognitif** de l'apprenant est son niveau d'acquisition de connaissances et **son état mental** est sa capacité mentale à acquérir ou non de nouvelles connaissances. Ainsi, si son état mental montre par exemple qu'il est motivé et concentré il pourra atteindre un état cognitif élevé. De même, si nous détectons que l'apprenant est en train de comprendre les notions présentées, de suivre le contenu pédagogique, nous constatons par conséquent qu'il suit le meilleur raisonnement et que son état cognitif augmente. Si cela n'est pas le cas, il est fortement recommandable de lui offrir l'aide nécessaire et d'intervenir au bon moment afin de corriger les déficits intellectuels. En effet, face à une tâche d'apprentissage, l'apprenant peut réagir différemment. Il peut suivre et assimiler les notions, comme il peut diverger sur d'autres tâches. Par ailleurs, Baker et ses collègues (Baker et Rossi, 2013) ont détecté quatre cas possibles où l'apprenant se trouve dans un état défavorable pour l'apprentissage. Ces cas sont les suivants :

- ***Off-task bahvior*** : ce comportement est défini comme étant « un comportement qui n'implique pas une interaction avec le système ou l'accomplissement d'une tâche d'apprentissage tels que les conversations hors-tâche et l'inactivité (Baker, 2007) ». Ce comportement a été initialement modélisé par l'extraction de certaines caractéristiques et l'utilisation de techniques d'apprentissage machine (modèles à variables latentes). À partir de ce modèle, d'autres alternatives de modélisation ont vu le jour en ajoutant d'autres caractéristiques telles que les données issues de mouvements exercés sur la souris (Cetintas et al., 2010).
- ***Gaming the system*** : il est défini par l'utilisation de toutes les propriétés du système qui servent à aider l'apprenant en cas de difficulté afin d'obtenir toutes

les réponses possibles pour anticiper l'apprentissage de la matière. Ce comportement est étudié par plusieurs chercheurs (Aleven et al., 2006; Baker et al., 2004; Koedinger et al., 2004). Les détecteurs automatiques de ce type de comportement utilisent généralement des règles (par exemple une réponse très rapide ou un clic continu sur l'aide peut renseigner sur ce comportement) ou des techniques de fouille de données et/ou apprentissage machine.

- **Carelessness** : il est défini comme étant « une erreur commise par l'apprenant qui a été déjà faite à l'avance (Clements, 1982) ou une action rapide ou impulsive (Maydeu-Olivares et D'Zurilla, 1995) ». C'est essentiellement le fait d'être inattentif et déconcentré dans une tâche. En se basant sur cette définition, plusieurs modèles automatiques ont été développés (Baker et al., 2008; Hershkovitz et al., 2013; San Pedro et al., 2011) pour détecter ce comportement.
- **Without Thinking Behavior** : il est défini comme étant « un comportement qui n'est pas en relation avec l'apprentissage et le curriculum (Rowe et al., 2009) ». Par exemple, ce comportement peut se manifester par le déplacement vers un endroit inutile dans un jeu (escalader des échafaudages), l'arrêt et le ralentissement sur des objets non nécessaires, etc. Ce comportement a fait aussi l'objet de plusieurs modèles automatiques (Sabourin, Rowe, et al., 2011; Wixon et al., 2012) pour la détection précoce de ce comportement.

Ces états peuvent susciter chez les apprenants un état de désengagement mental où il est tout à fait déconcentré et n'est pas intéressé à apprendre. La tendance négative de cet état de l'apprenant peut avoir une influence directe sur son état cognitif de compréhension et de raisonnement. Dans ce cas, il serait utile d'intervenir immédiatement afin de le changer et le corriger par un état positif plus favorable à l'apprentissage.

Or, intervenir pour corriger une situation n'est pas très évident et facile à entreprendre. Tout d'abord, il faut bien déterminer les moments d'intervention : faut-il intervenir après l'erreur ou avant? Après combien de temps faut-il intervenir? Pour répondre à ces questions, une étude empirique de certains cas serait préférable. Une étude de l'évolution des états mentaux de l'apprenant tels que l'engagement/désengagement, la frustration, le stress, la motivation, etc.

présente aussi une solution possible pour détecter les faiblesses de l'apprenant et par conséquent les corriger. La mesure des états mentaux peut se faire par une approche cérébrale (Chaouachi et al., 2015). Ces auteurs ont créé un système de mesure, nommé Mentor, qui mesure continuellement (en temps réel, online et pas offline) les états mentaux d'engagement et de charge cognitive à partir de traitement des données brutes issues de l'électroencéphalogramme. Ensuite, il faut décider quant aux stratégies d'aide à utiliser et à intégrer selon l'évolution de certaines métriques mesurées sur l'apprenant comme par exemple les états mentaux. De plus, il faut distinguer quels états mentaux sont plus importants à considérer.

Si nous connaissons l'état cognitif de l'apprenant, nous pouvons déterminer quelles stratégies d'aide à appliquer dans chaque cas. Il serait alors essentiel d'analyser d'une façon plus précise l'évolution de l'état cognitif de l'apprenant en fonction de ses états mentaux afin de distinguer et décider quelles stratégies d'aide à prendre en considération. Une étude empirique qui établit la relation entre les états mentaux ainsi que les stratégies d'aide est donc souhaitable.

1.3 Objectifs de recherche

Nos travaux de recherche visent essentiellement à détecter et à améliorer l'état cognitif de l'apprenant dans deux types d'environnements : environnement purement cognitif et jeux sérieux (éducatifs). L'amélioration de l'état cognitif de l'apprenant signifie une amélioration de certains états mentaux de l'apprenant tels que la frustration, l'engagement, la charge cognitive, etc. Dans un premier temps, nous détectons ces états mentaux. Dans un deuxième temps, nous mettons au point une méthode pour améliorer son rendement éducatif et ses performances dans l'apprentissage de nouvelles notions. Plus particulièrement, nous proposons de réaliser les trois objectifs suivants :

- 1. Prédire les performances de l'apprenant dans un environnement purement cognitif :** notre premier objectif consiste à prédire les performances de l'apprenant dans un environnement cognitif constitué de différentes catégories de tâches cognitives. La prédiction se base essentiellement sur des données physiologiques collectées sur l'apprenant et sur l'information sur la tâche cognitive (type, catégorie, difficulté et durée).

2. **Utiliser des JS (en particulier les jeux éducatifs) dans l'apprentissage d'un domaine pour prédire le besoin d'aide** : notre deuxième objectif consiste à prédire le besoin d'aide dans un jeu sérieux. Dans cet environnement, nous avons mesuré le niveau d'échec et de réussite. La prédiction de besoin d'aide a été faite grâce à des algorithmes d'apprentissage machine.

3. **Renforcer le raisonnement intuitif grâce à des stratégies d'aide intégrées dans un jeu de réalité virtuelle (JRV)** : notre troisième objectif consiste à améliorer le raisonnement intuitif en recourant à différents types de stratégies d'aide. Ces stratégies d'aide sont intégrées dans un jeu de réalité virtuelle (JRV). Celle-ci favorise généralement l'immersion et la motivation. Les stratégies d'aide proposées visent essentiellement à diriger l'apprenant dans son raisonnement et le guider vers la solution des problèmes proposés dans ce jeu. C'est un JSI qui repose sur une simulation de quelques notions de physique. Il s'adapte en fonction de deux états mentaux : la frustration et l'engagement pour fournir les stratégies d'aide appropriées.

Pour répondre au premier objectif, nous avons conçu un environnement purement cognitif, nommé 'logique'. Logique est composée de trois différentes catégories de tâches cognitives (mémorisation, concentration, et raisonnement). Chaque catégorie est composée de deux ou trois types de tâches cognitives, soit sept types de tâches au total. Cet environnement est inspiré de la taxonomie de Bloom pour la classification des tâches cognitives (Krathwohl, 2002). Il vise essentiellement à mesurer les connaissances de l'apprenant dans chaque type de tâche cognitive et à déterminer l'influence de la catégorie de la tâche sur la classification des performances de l'apprenant. Grâce à cet environnement purement cognitif qui introduit le rôle de la catégorie de la tâche présentée, nous avons réalisé deux études empiriques. Dans les deux études empiriques, nous avons collecté les données physiologiques issues essentiellement de l'électroencéphalogramme (EEG) et de l'activité électrodermale de la peau (EDA). En ce qui concerne l'EEG, nous nous sommes concentré sur des **états mentaux** définis par (Berka et al.,

2004) pour étudier et prédire les performances de l'apprenant. Ces états sont principalement : l'**engagement** qui est relié à un niveau élevé d'attention et de vigilance pendant une tâche, la **charge cognitive** qui peut être vue comme la charge cognitive et la vigilance mentale dans une tâche particulière (le degré d'effort mental pour la réalisation d'une tâche particulière), et la **distraction** qui reflète le sentiment d'être endormi et léthargique avec une faible énergie (Stevens et al., 2007b). Dans la première étude, nous nous sommes intéressés à étudier la variation des états mentaux et des performances de l'apprenant et leurs différences selon la catégorie et le type de la tâche cognitive. Nous avons utilisé les trois mesures d'états mentaux collectés dans les tâches cognitives de même que la mesure de l'activité électrodermale pour étudier leur variation et évolution. Le profil de l'apprenant est alors construit à partir des trois états mentaux, de l'activité électrodermale, et de la catégorie de la tâche cognitive. Enfin, nous avons développé un modèle de régression multiple pour prédire les performances de l'apprenant dans l'environnement *logique*. Cette question a été approfondie dans une deuxième étude empirique où nous avons entraîné et testé différents algorithmes d'apprentissage machine. Nous avons aussi étudié la possibilité de classifier les apprenants en quatre catégories allant de 'très faibles apprenants' à 'très forts apprenants'.

Pour répondre à notre deuxième objectif, soit de prédire le besoin d'aide dans les JS. Nous avons utilisé un nouvel environnement les jeux éducatifs (JS et JSI) pour l'apprentissage en chimie et en physique. Nous avons développé deux environnements d'apprentissage qui se caractérisent par deux aspects : un aspect ludique et un aspect éducatif. Ces deux aspects se sont montrés efficaces pour des buts éducatifs (Jackson et al., 2012; Prensky, 2007). Pour vérifier ce but, nous avons étudié l'évolution de sept états mentaux extraits des EEG (Ghergulescu et Muntean, 2014), le diamètre de la pupille extraite d'un traceur de regard expérimenté par (Bartels et Marshall, 2012), et sept expressions faciales définies par Ekman (Ekman, 1970). Nous avons développé un jeu éducatif nommé *LewiSpace* pour l'apprentissage des diagrammes de Lewis (Chimie). Le jeu est constitué de cinq missions différentes qui consistent à chaque fois à construire un diagramme de Lewis d'une molécule chimique. Les molécules à construire sont dans un ordre croissant de difficulté (du plus facile au plus difficile). Il intègre explicitement les données d'EEG, traceur de regard, et expressions faciales. Nous avons procédé à une étude empirique où nous collectons toutes ces données physiologiques sur l'apprenant ainsi que la

réponse à un questionnaire auto-évaluatif de besoin d'aide dans chaque mission. Nous avons adopté essentiellement une approche multimodale où nous recourons à différents capteurs physiologiques dans le but de prédire le besoin d'aide (rétroaction non sollicitée par les apprenants). Nous avons prétraité les données pour enlever toute donnée pouvant détériorer la qualité de notre base de données (bruits, artefacts, etc.). Nous avons ensuite entraîné et testé différents algorithmes d'apprentissage machine pour la prédiction du besoin d'aide. Nous avons finalement validé nos modèles et nous avons procédé à la sélection des caractéristiques les plus déterminantes dans ces algorithmes (qui peuvent influencer directement la qualité de prédiction) en étudiant la précision établie par les algorithmes d'apprentissage machine. Pour la prédiction du besoin d'aide, nous avons essentiellement utilisé deux approches : (1) généralisée où les algorithmes d'apprentissage machine sont entraînés sur les précédents participants et testés ensuite sur le participant en cours et (2) individualisée où les algorithmes d'apprentissage machine sont testés et entraînés sur le participant courant en divisant ses tentatives dans chaque mission en fenêtres de temps étiquetées selon deux tâches de calibration (facile et difficile).

Enfin, pour répondre à notre troisième objectif soit de renforcer le raisonnement intuitif grâce aux différents stratégies d'aide et de messages de support, nous avons développé un jeu en réalité virtuelle (qui est un JSI) vue les avantages pour l'apprentissage d'un tel environnement qui augmente les aspects d'immersion et de motivation (Freina et Ott, 2015; Merchant et al., 2014; Pan et al., 2006). Ce jeu nommé *Inertia*, consiste à apprendre intuitivement des règles basiques de physique sur la dynamique du mouvement. Il est formé de quatre catégories de jeux (bowling 1, bowling 2, pivot, et billard). Nous avons mené une étude empirique où nous collectons les données de deux états mentaux extraites de EEG (Ghergulescu et Muntean, 2014) qui sont : la frustration et l'engagement. Le jeu est composé de deux versions : V0 qui ne contient pas de stratégies d'aide et V1, version incluant des stratégies d'aide. Cette dernière version est adaptative (JSI) dans le sens où elle fournit aux apprenants des stratégies d'aide en fonction de l'évolution des deux états mentaux (la frustration et l'engagement) et des états d'échec/ réussite dans chaque étape du jeu.

1.4 Organisation du document

Ce document est organisé comme suit : le chapitre 2 est consacré à l'analyse de la littérature sur les différents types de raisonnement et des jeux sérieux intelligents. Nous présentons dans ce chapitre le cadre général de nos recherches, soit la possibilité de détection et d'amélioration de l'état cognitif de l'apprenant. Dans un premier temps, nous décrivons le lien entre l'état cognitif de l'apprenant et l'amélioration de ses performances. Dans un deuxième temps, nous examinons les états mentaux de l'apprenant qui peuvent induire un état cognitif. Dans un troisième temps, nous décrivons et définissons les deux types de raisonnement, soit le raisonnement basé sur le calcul et la logique et le raisonnement intuitif. Dans un quatrième temps, nous définissons et discutons des jeux sérieux (JS) et en particulier les jeux éducatifs (un type des JS qui s'intéresse à l'apprentissage). Et finalement, nous discutons de l'adaptation dans les jeux sérieux en intégrant des techniques d'intelligence artificielle plus précisément des algorithmes d'apprentissage machine (JSI).

Les chapitres 3, 4, 5, 6 et 7 sont consacrés à nos contributions. Ces chapitres sont présentés sous la forme de cinq articles de recherche : le premier article a été publié dans le journal *International Journal of Information and Education Technology* en 2016. Une version de cet article a été acceptée aussi dans la conférence *International Conference on Educational and Information Technology* et a gagné le meilleur prix de présentation (2015). Les deux autres articles sont acceptés dans le journal *Journal of Education and Training Studies* (2016 et 2015). Le quatrième article est accepté dans la conférence internationale *Florida Artificial Intelligence Research & Society (Flairs 2016)*. Enfin, le dernier article a été accepté à la conférence *Flairs 2017*.

Le chapitre 3 décrit l'étude empirique réalisée dans l'environnement purement cognitif logique pour examiner la variation des facteurs émotionnels et cognitifs de l'apprenant en fonction de la catégorie et de type de la tâche proposée. Le chapitre 4 présente la possibilité de classifier et de prédire les performances de l'apprenant dans l'environnement *logique* grâce à l'utilisation de différents algorithmes d'apprentissage machine. Le chapitre 5 introduit le jeu *Lewispace* (JS) sur l'apprentissage des diagrammes de Lewis (environnement expérimental) et présente une approche multimodale pour la prédiction du besoin d'aide dans chaque mission du

jeu. Le chapitre 6 présente la possibilité d'utiliser deux approches d'apprentissage machine pour la prédiction du besoin d'aide de l'apprenant en temps réel. Le chapitre 7 décrit la possibilité à recourir à des stratégies d'aide pour améliorer le raisonnement intuitif et par conséquent les performances de l'apprenant et son état cognitif. Il présente également le jeu de réalité virtuelle *Inertia* (JSI : deuxième environnement expérimental) développé pour cette fin.

En conclusion, le chapitre 8 résume les contributions de cette thèse, les limites de nos travaux et les perspectives de nos recherches.

Chapitre 2

État de l'art

2.1 Introduction

Ce chapitre décrit l'état de la littérature reliée à nos travaux de recherche. Étant donné que le but de cette thèse est de détecter et d'améliorer l'état cognitif de l'apprenant, nous commençons ce chapitre par des définitions sur la cognition et l'état cognitif. Ensuite, nous distinguons les deux catégories de raisonnement : le raisonnement basé sur le calcul et la logique et le raisonnement intuitif (RI). Nous avons utilisé ces deux types de raisonnement dans la construction de différents environnements d'apprentissage décrits plus tard dans cette thèse. Le raisonnement basé sur le calcul s'intéresse à tout processus de raisonnement qui recourt aux mathématiques, au calcul et aux statistiques. Le raisonnement intuitif recourt plus à l'expérience et au jugement. Puis, nous examinons différents types d'environnements d'apprentissage, plus spécifiquement les jeux éducatifs qui sont un type particulier des jeux sérieux (JS) pour le but d'apprentissage. Nous discutons ainsi de la possibilité d'augmenter ces derniers par une dimension d'intelligence artificielle pour aboutir à des jeux sérieux intelligents (JSI). De plus, dans la dernière partie, nous discutons des techniques d'adaptation dans les JSI selon différentes approches. Nous clôturons le chapitre en mentionnant six types d'adaptation possibles dans les JSI: adaptation selon les émotions ressenties, selon l'aspect narratif, selon les objectifs de l'apprenant, selon les connaissances de l'apprenant, selon le niveau de difficulté du jeu et selon les mesures physiologiques et les différentes approches multimodales de combinaison de ces mesures.

2.2 Cognition et état cognitif

En psychologie, le terme cognition est un terme scientifique qui désigne « une fonction complexe multiple regroupant l'ensemble des activités mentales (pensée, perception, action, volonté, mémorisation, rappel, apprentissage) impliquées dans la relation de l'être humain avec son environnement et qui lui permettent d'acquérir et de manipuler des connaissances (associations, rétroaction, traitement de l'information, résolution de problèmes, prise de décision, etc.) »². Ce terme renvoie ainsi à l'ensemble des processus psychiques liés à l'esprit ou au cerveau. Dans la littérature, il existe des approches qui ne distinguent pas entre le terme cerveau et esprit tandis que d'autres considèrent plutôt l'existence de la composante organique uniquement (cerveau) ou une différence entre les deux composantes. Par exemple, selon l'approche organiciste, le cerveau est un organe physique situé dans nos crânes alors que l'esprit n'existe pas ou est une conséquence de ce dernier. Krishnamurti (Krishnamurti et Bohm, 1987) fait une différence entre le cerveau et l'esprit. Il considère que l'esprit est temporel alors que le cerveau est hors temps. L'esprit regroupe les fonctions mentales et leurs évolutions temporelles (intuition, jugement, perception, pensée, etc.). Ce qui est contraire au cerveau qui est un organe. Tandis que Hobbes opte pour la théorie de l'identité entre l'esprit et le cerveau (Jaume, 1983). Il se base sur le fait que la causalité mentale est identique à des causes physiques (ce qui entraîne que le cerveau et l'esprit sont équivalents).

La cognition désigne différents processus mentaux et fonctions cognitives qui s'impliquent dans le traitement et la manipulation des données pour réaliser un but donné. Parmi ces processus, nous citons à titre d'exemple : la mémoire, le langage, le raisonnement, l'apprentissage, le jugement, l'intelligence, la résolution de problèmes, la prise de décisions, la perception des événements et l'attention.

Les fonctions cognitives réfèrent à la capacité de coordonner les actions et le raisonnement et de le diriger vers un but particulier, ce qui implique de planifier et orchestrer des séquences complexes de comportements et de définir des priorités à ces buts et sous-buts

² Définition prise du Grand Dictionnaire Terminologique, <http://www.granddictionnaire.com/>

(Miller et Wallis, 2009). Ces fonctions regroupent généralement l'attention, les fonctions langagières, les fonctions intellectuelles et les fonctions exécutives. Elles seront détaillées dans ce qui suit.

L'attention est une fonction cognitive complexe qui fait référence à la capacité à être alerte à son environnement et à maintenir son attention sur une durée de temps appropriée pour son âge³. L'attention fait également référence à la capacité à se concentrer sur une tâche donnée et à partager son attention entre plusieurs tâches simultanément.

Les fonctions langagières comprennent un ensemble d'habiletés qui sont généralement divisées en deux catégories, soit les habiletés réceptives (comprendre le langage parlé et écrit) et les habiletés expressives (parler et écrire).

Les fonctions intellectuelles sont un ensemble d'habiletés qui comprennent certaines compétences verbales, le raisonnement non-verbal, des aspects de la mémoire de travail (voir figure 2.2) et la vitesse de traitement de l'information et d'exécution qui réfère au rythme auquel les différentes opérations mentales sont déclenchées et exécutées.

Les fonctions exécutives sont les fonctions cognitives les plus élaborées et font référence à l'acte de prendre une décision et de les exécuter (Godefroy, 2008; Miller et Wallis, 2009). Le traitement de l'information touche des lobes bien précis du cerveau humain plus particulièrement les lobes frontaux (voir figure 2.1) connus sous le nom de cortex préfrontal (Bunge et al., 2002; Fuster, 2002; Miller et Wallis, 2009).

³ <https://aqnp.ca/la-neuropsychologie/les-fonctions-cognitives/>

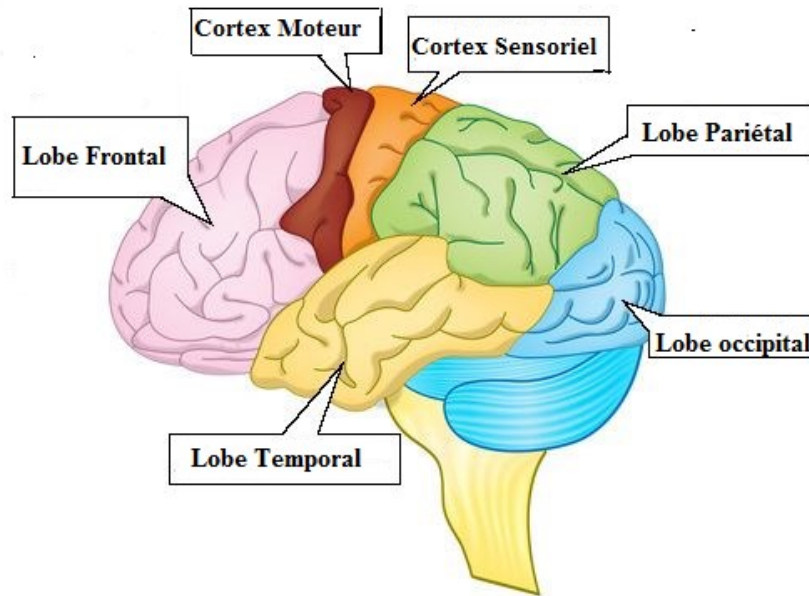


Figure 2. 1 Les différents lobes du cerveau humain

La figure 2.1 illustre les différents lobes du cerveau : le lobe frontal qui est composé des fonctions exécutives, les émotions et le contrôle de comportements. Le lobe temporal qui contrôle la mémorisation, la compréhension et le langage. Le lobe occipital qui est responsable pour la vision, et le lobe pariétal qui dirige la perception, le calcul et le sens. Il est aussi composé du cortex moteur qui est responsable du mouvement et du cortex sensoriel qui contrôle le traitement des sens des informations (Teplan, 2002; Vuilleumier, 2005).

Selon Fuster 2002, les fonctions exécutives regroupent deux catégories (Fuster, 2002) :

- L'intégration temporelle est le processus qui consiste à l'organisation des actions pour l'atteinte d'un but déterminé. C'est un rôle fondamental du cortex préfrontal qui consiste à l'exécution de différentes formes d'actions (mouvements des yeux, comportements émotionnels, performance intellectuelle, langage, etc.) et au choix par la suite des alternatives de prise de décision et d'élaboration des actions structurées temporairement,
- La mémoire de travail (Baddeley, 1986) est la rétention temporaire d'un élément d'une information. C'est la mémoire à court terme qui dure moins d'une minute et qui peut influencer la mémoire à long terme (voir figure 2.2). La mémoire à court

terme permet de retenir et de réutiliser une quantité limitée d'informations pendant un temps relativement court.

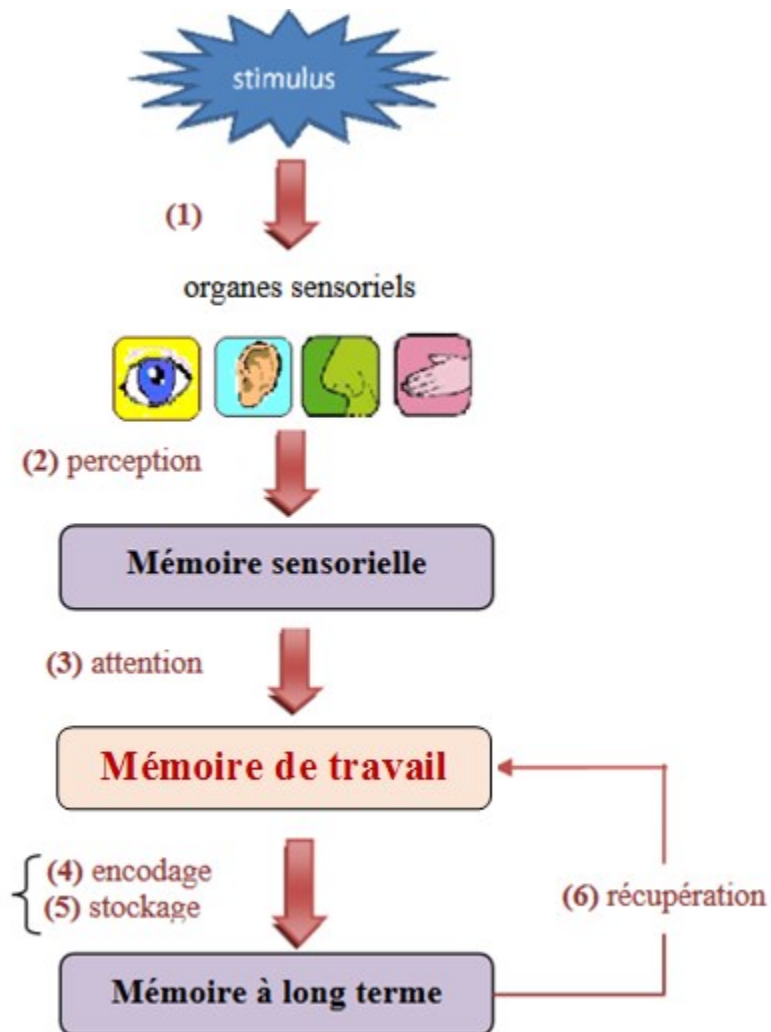


Figure 2. 2 Les différentes mémoires (Atkinson et Shiffrin, 1968)

La figure 2.2 illustre les étapes de mémorisation en commençant par la mémoire sensorielle et en arrivant à la mémoire à long terme. À partir de ce schéma, nous remarquons que la mémoire de travail (mémoire à court terme) est une étape intermédiaire pour la mémorisation à long terme.

Après avoir défini le terme cognition et donné quelques exemples de fonctions cognitives, nous définissons maintenant le terme 'état cognitif (EC)' que nous utilisons tout au long de cette thèse. Nous définissons l'EC comme « **étant le niveau d'acquisition de**

connaissances de l'apprenant. Ce niveau de connaissance peut être mesuré essentiellement par l'évolution des scores de l'apprenant (ou ses résultats obtenus) pour une tâche d'apprentissage bien déterminée ».

2.3 État Mental

L'état mental se distingue de l'état cognitif par l'ajout d'une dimension affective. C'est un état du cerveau incluant les émotions ressenties. Cet état est une évaluation de degré de conscience, de vigilance et d'alerte des individus qui peut être affecté par plusieurs conditions externes ou internes (Conn, 1977). Ainsi, le fait de ressentir des émotions peut avoir une influence directe sur l'état mental dans le sens où cet état peut atteindre une valence positive haute comme l'excitation ou une valence négative faible comme la frustration. Nous pouvons alors définir l'état mental (EM) comme **« étant la capacité mentale à acquérir ou non de nouvelles connaissances. Cette capacité est influencé par l'état émotionnel de la personne ».**

Les émotions sont définies généralement par deux axes : **valence** (positive ou négative) et **activation** (faible ou élevée). Par exemple, Predinger et Ishizuka (2005) utilisent le modèle de Lang (1995) (Lang, 1995) pour classer six émotions (Predinger et Ishizuka, 2005) : la peur, l'excitation, la joie, la tristesse, la frustration et l'état calme. Dans ce modèle, la joie est représentée par une activation élevée et une valence positive, l'état calme est représenté par une activation neutre et valence positive et la frustration est représentée par une activation élevée et une valence négative (voir figure 2.3).

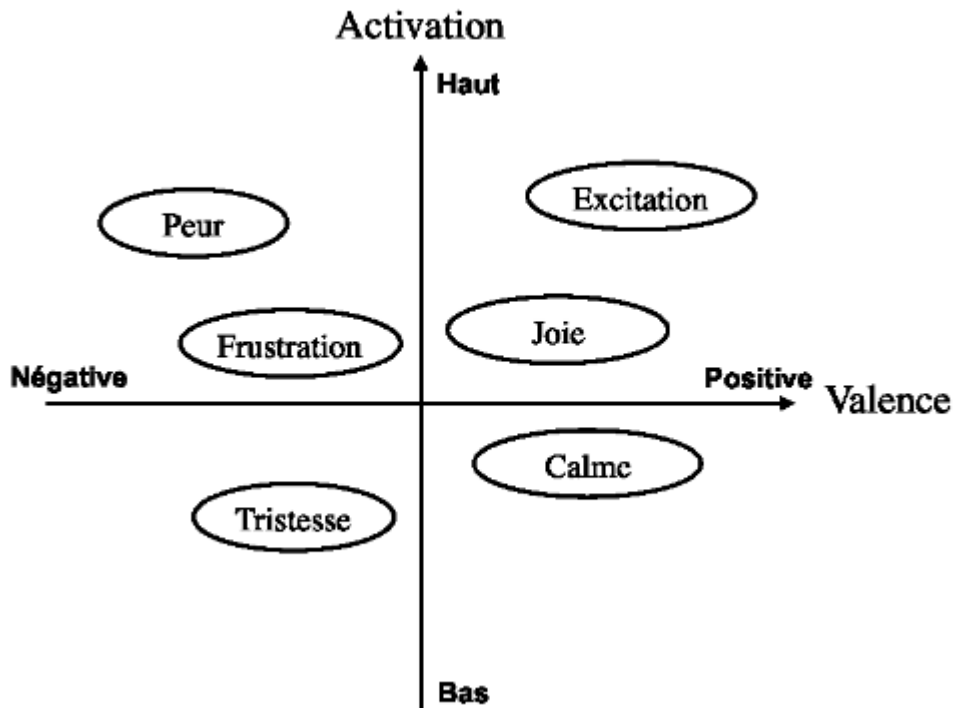


Figure 2. 3 Représentation des émotions selon (Prendinger et Ishizuka, 2005)

Dans les travaux actuels, les chercheurs ne font pas de distinction entre les états mentaux et les émotions (Calvo et D'mello, 2010; D'Mello et Kory, 2015). Ces derniers considèrent par exemple l'engagement comme une émotion. Alors que d'autres considèrent l'engagement comme un état mental (Chaouachi et al., 2010).

Dans cette thèse, nous n'avons pas approfondi cette question. Nous avons plutôt considéré comme 'un état mental' un état qui peut apparaître suite à plusieurs facteurs externes de l'environnement et internes (entre autres les émotions ressenties) et qui peut être observé en temps réel (voir interface graphique à la figure 2.5) et extrait à partir des signaux cérébraux du cerveau (EEG : électroencéphalogramme) (Berka et al., 2004; Stevens et al., 2007a; Stevens et al., 2007b). La mesure d'électroencéphalogramme a été faite dans un premier temps (les deux premières études empiriques présentés plus tard dans le chapitre 3 et chapitre 4) par le casque B-Alert X10⁴ (conçu par *Advanced Brain Monitoring* (ABM) et illustré par la figure 2.3). Ce

⁴ Le casque B-Alert est utilisé dans notre environnement purement cognitif Logique

casque capte les signaux issus de neuf senseurs (F3, Fz, F4, C3, Cz, C4, P3, Poz et P4) avec deux signaux de référence attachés aux oreilles.



Figure 2. 4 Positions des senseurs dans le casque B-Alert X 10

L'interface graphique de B-Alert permet d'obtenir la mesure des ondes cérébrales (alpha, beta, theta et delta décrites plus tard dans ce chapitre). Cette interface montre aussi la variation des signaux et certaines mesures des états mentaux. Les états mentaux extraits sont de l'ordre de trois: engagement, charge cognitive et distraction (Berka et al., 2004; Stevens et al., 2007a; Stevens et al., 2007b). Ces états sont définis comme suit et sont présentés dans la figure 2.5 :

- **Engagement** : le degré d'alerte mentale, d'attention et de vigilance,
- **Charge Cognitive** : la charge mentale d'une tâche donnée en termes d'occupation de mémoire, de processus d'apprentissage et de raisonnement. Selon (Eggemeier et al., 1991), La charge de travail mentale se réfère à « la partie de la capacité de traitement de l'information des ressources qui est réellement nécessaire pour répondre aux exigences du système »,
- **Distraction** : le fait d'être fatigué et léthargique. C'est pratiquement l'inverse de l'engagement.

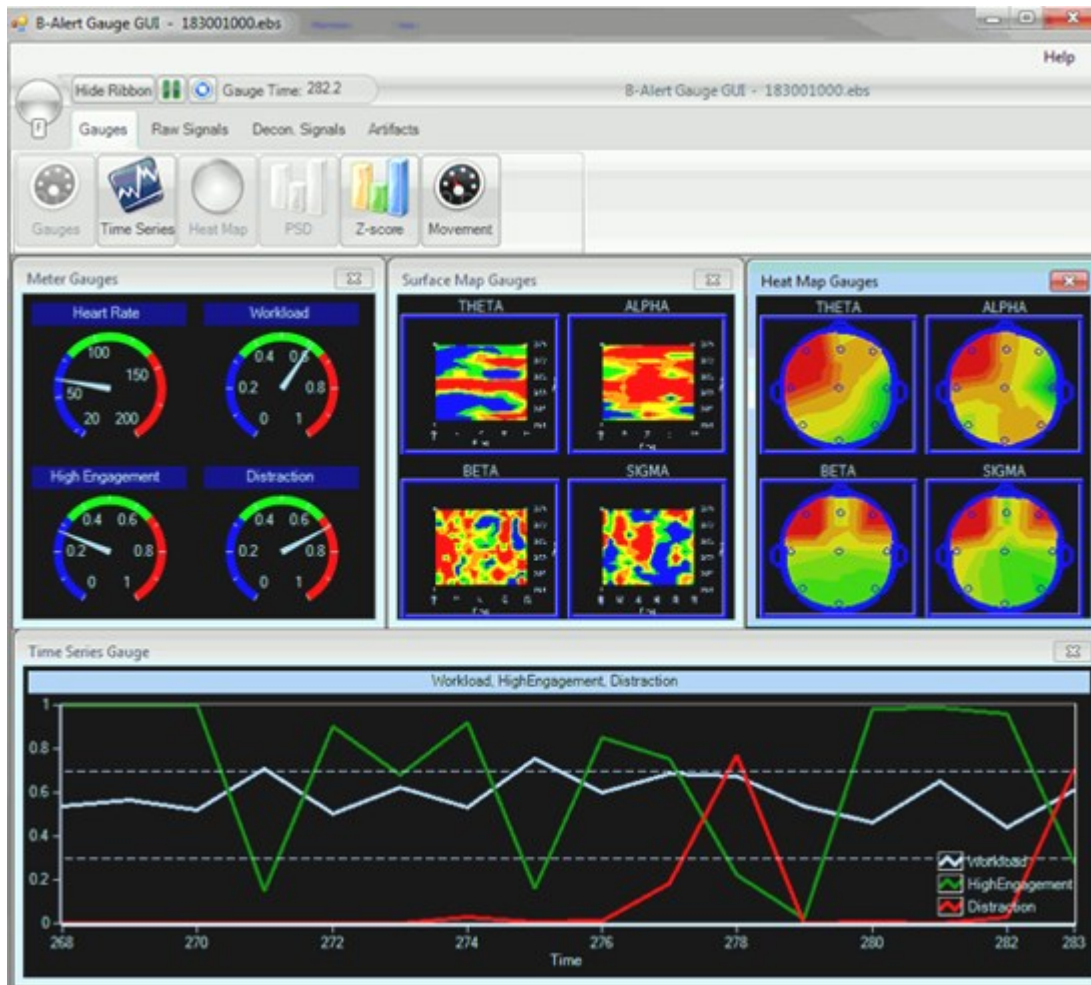


Figure 2. 5 Interface graphique de *B-Alert X10* montrant la variation en temps réel des états mentaux

Il est à noter que ces états sont calculés à partir de trois tâches essentielles définies par ABM et sont validées dans les études faites par Chris Berka et ses collègues (Berka et al., 2004; Stevens et al., 2007a; Stevens et al., 2007b) . Ces tâches sont : 3CVT (*3 Choice Vigilance Task*) où le participant doit choisir une tâche parmi trois tâches principales, EO (*Eyes Open*) qui reflètent la réponse du participant à des taches visuelles (comme le fait de voir un cercle se déplacer dans l'écran) et EC (*Eyes Closed*) où le participant doit fermer ses yeux et répondre à des sons en appuyant à chaque fois sur le bouton. Les détails de ces tâches sont présentés dans la Table 2.1 ci-après.

Table 2. 1. Les tâches de classification des états mentaux⁵

Baseline task	Action	B-Alert Class probabilities
3-choice vigilance task (~7-min; optional 20-min)	Choose between primary vs. secondary or tertiary task every 1.5 to 3-seconds	High Engagement
Eyes open (5-min)	Respond to visual probe every 2-seconds	Low Engagement
Eyes closed (5-min)	Respond to audio tone every 2-seconds	Distraction if episodic Drowsy if sequential
None	Derived by regression from other three tasks	Sleep Onset

Dans le reste de nos travaux (chapitre 5, 6 et 7), nous avons utilisé comme casque de mesure d'EEG, le casque *Emotiv EPOC*⁶ vu qu'il est léger, pratique et plus facile à intégrer avec les jeux vidéo (voir figure 2.6). Il permet de capter 14 senseurs (AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8 et AF4) et deux références liées aux oreilles.

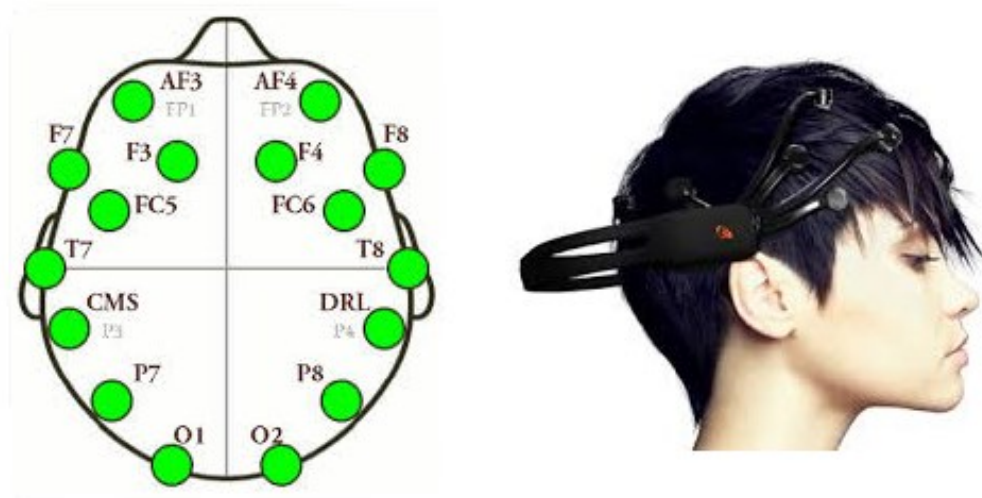


Figure 2. 6 Distribution des senseurs dans le casque *Emotiv EPOC*

Le casque *Emotiv EPOC* mesure déjà un ensemble d'états mentaux fournis et calculés en temps réel. Nous considérons les cinq états mentaux suivants : l'excitation à court terme, l'excitation à long terme, la méditation, la frustration et l'ennui. Ces états sont décrits

⁵ <https://www.biopac.com/product/b-alert-wireless-eeeg-with-acqknowledge-plus-cognitive-state-software/>

⁶ Le casque Emotiv EPOC est utilisé dans ce qui suit pour le jeu éducatif *Lewispace* et *Inertia*

brièvement dans ce qui suit. Ils sont aussi utilisés par (Ghergulescu et Muntean, 2014) pour mesurer la motivation des joueurs dans des jeux vidéo et se sont révélés efficaces :

- **Excitation à court terme** : « C'est un état temporaire ou latent se manifestant par des réactions spontanées ou provoquées d'un système physique, psychique de toute la personne »⁷. La durée est à court terme comme son nom l'indique.
- **Excitation à long terme** : C'est la même définition de l'excitation à court terme sauf que celle-ci dure plus longtemps.
- **Méditation** : En psychologie, la méditation consiste à un état mental de grande concentration d'esprit pour approfondir sa réflexion. Il s'agit d'une pensée réfléchie et concentrée sur un sujet particulier. (Cardoso et al., 2004) définit la méditation comme un état mental qui nécessite les éléments suivants : l'utilisation d'une technique spécifiée, la relaxation des muscles, la relaxation logique, l'auto-induction d'un état et l'utilisation de la compétence optique.
- **Frustration** : C'est un état mental d'insatisfaction provoqué par le sentiment de n'avoir pu réaliser un désir. Elle réveille des émotions telles que l'impatience, la colère ou la tristesse. Elle peut même mener à la dépression et dans certains cas à l'agression (Dollard et al., 1939).
- **Engagement** : C'est un état d'attention, de vigilance et d'attention. Il reflète si l'apprenant est concentré sur la tâche ou s'il est distrait.

2.4 Types de raisonnement

Après avoir décrit l'état cognitif et l'état mental, nous décrirons brièvement dans cette section deux types de raisonnement répandus soit le raisonnement basé sur le calcul et le raisonnement intuitif. Le raisonnement est un processus cognitif qui permet de vérifier ou d'obtenir de nouveaux résultats. Il peut être utilisé dans différents domaines : médecine, mathématiques, informatique, pédagogie, physique, etc.

⁷ Définition extraite de Grand Dictionnaire Terminologique

2.4.1 Raisonnement basé sur le calcul et la logique

Le raisonnement basé sur le calcul est un raisonnement qui utilise la logique et se base sur des calculs et non sur des affects ou émotions. Le raisonnement basé sur le calcul, appelé raisonnement a priori, recourt souvent à une formalisation logique pour établir une preuve. Il existe plusieurs formes de raisonnement basé sur le calcul et plusieurs définitions selon les domaines (Bergqvist, 2007; Boutilier, 1992; Pittalis et Christou, 2010). Nous illustrons dans ce qui suit quelques exemples de raisonnement logique⁸. Les exemples suivants sont tirés de ce site web.

- **Raisonnement inductif :**

Le raisonnement inductif consiste à généraliser des cas particuliers. À partir des phénomènes observés, nous allons induire une loi générale sans vérifier tous les exemples. En d'autres termes, nous allons tirer une nouvelle règle à partir des observations faites sur un ensemble des exemples. Ce type de raisonnement n'est pas toujours correct et peut induire de l'erreur. C'est pourquoi il n'est pas utilisé dans le domaine de mathématiques. Par exemple, 'les arbres ont été mouillées **de nombreuses fois** quand il a plu', donc nous pouvons conclure la règle suivante : 'les arbres se mouillent **toujours** quand il pleut'.

- **Raisonnement déductif :**

Le raisonnement déductif consiste à tirer une nouvelle règle à partir d'une ou plusieurs propositions. Cette nouvelle règle ou assertion est la conséquence des propositions observées. Un exemple concret de ce type de raisonnement est le suivant : 'Quand il pleut, les éléments à l'extérieur se mouillent. Étant donné que les routes se trouvent à l'extérieur, donc nous pouvons tirer la nouvelle règle suivante : quand il pleut, les routes se mouillent'. Ce type de raisonnement est utilisé surtout dans le domaine de mathématiques. Par exemple, si nous avons deux nombres rationnels a et b , nous voulons montrer que leur somme est un rationnel. La démonstration se fait comme suit :

$$a \in \mathbb{Q}, \exists p, q \text{ tel que } a = \frac{p}{q} \quad (2.1)$$

⁸ http://exo7.emath.fr/cours/ch_logique.pdf

$$b \in Q, \exists p', q' \text{ tel que } b = \frac{p'}{q'} \quad (2.2)$$

$$\Rightarrow a + b = \frac{p}{q} + \frac{p'}{q'} = \frac{pq' + p'q}{qq'} \in Q \quad (2.3)$$

Ainsi, nous pouvons déduire à partir de l'équation (2.3) que la somme de a et b est un rationnel.

- **Raisonnement par abduction**

Selon le dictionnaire terminologique, le raisonnement par abduction est un mode de raisonnement qui permet d'augmenter la vraisemblance d'une hypothèse par l'ajout de nouveaux faits⁹. C'est un processus permettant de créer et de sélectionner des hypothèses à partir d'un ensemble de faits, événements ou lois. La création des hypothèses permet d'expliquer un phénomène ou une observation (Aliseda, 2006). Par exemple, pendant la phase de recherche d'un problème mathématique, les étudiants formulent des hypothèses afin de résoudre le problème.

- **Raisonnement cas par cas :**

Ce raisonnement consiste à vérifier en premier temps une assertion pour un ensemble de données à partir d'un sous ensemble de cet ensemble. Ensuite, nous vérifions l'assertion pour le reste des éléments de cet ensemble. Par exemple, pour montrer que $|x-1|$ est supérieure à $(x^2 - x + 1)$ pour tout $x \in R$ (*l'ensemble des réels*), il faut considérer deux cas : $(x-1) \geq 0$ et $(x-1) \leq 0$. Ensuite, nous étudions les deux cas et faisons les calculs :

1^{er} cas : $(x-1) \geq 0 \Rightarrow |x-1| = x-1$

$$x^2 - x + 1 - |x-1| = x^2 - x + 1 - x + 1 = x^2 - 2x + 2 = (x-1)^2 + 1 \geq 0$$

2^{ème} cas : $(x-1) \leq 0 \Rightarrow |x-1| = 1-x$

$$x^2 - x + 1 - |x-1| = x^2 - x + 1 - 1 + x = x^2 \geq 0$$

Donc nous pouvons conclure que pour tout $x \in R$, l'assertion est vraie.

⁹ http://www.granddictionnaire.com/ficheOqlf.aspx?Id_Fiche=17016061

▪ **Raisonnement par contraposition :**

Le raisonnement par contraposition est basé sur l'équivalence suivante qui existe entre les assertions :

$$Si P \Rightarrow Q \Leftrightarrow non(Q) \Rightarrow non(P) \quad (2.4)$$

C'est-à-dire, si nous voulons montrer que $P \Rightarrow Q$ est vraie, nous supposons que $non(Q)$ est vraie et nous montrons alors que $non(P)$ est vraie.

Par exemple, nous voulons monter que pour tout entier n , si n^2 est pair alors n est pair. Pour ce faire, nous supposons que n n'est pas pair. C'est-à-dire $\exists k \in \mathbb{N}$ tel que $n = 2k + 1$. Dans ce cas, $n^2 = (2k+1)^2 = 4k^2 + 4k + 1$ qui est un nombre impair. D'où, par contraposition, nous concluons que si n^2 est pair, n est pair.

▪ **Raisonnement par l'absurde :**

Il imagine les conséquences absurdes d'une idée pour la réfuter. Le raisonnement par l'absurde (ou apagogie) est un raisonnement qui permet de démontrer qu'une affirmation est vraie en montrant que son contraire est faux. Il s'appuie sur la règle logique que : Si "non P" est faux, alors P est vrai.

Le raisonnement consiste à « supposer que l'affirmation contraire est vraie et à en tirer les conséquences que cela pourrait avoir. Une seule conséquence absurde, manifestement fausse ou une contradiction permet d'affirmer que l'affirmation contraire est fausse et donc d'en conclure que l'affirmation initiale est vraie »¹⁰.

Avec ce type de raisonnement, pour montrer que l'assertion $P \Rightarrow Q$ est vraie, nous supposons à la fois que P est vraie et que Q est fausse et nous cherchons une contradiction.

Par exemple, soient deux nombres $a, b \geq 0$, Montrons que

$$Si \frac{a}{b-1} = \frac{b}{1+a} \Rightarrow a = b \quad (2.5) .$$

Pour ce faire, nous supposons que $\frac{a}{b-1} = \frac{b}{1+a}$ et que $a \neq b$. Comme $\frac{a}{b-1} = \frac{b}{1+a}$ alors $a(1+a) = b(b-1)$, c'est-à-dire $a+a^2 = b+b^2$ ou encore $a^2 - b^2 = -(a-b) \Rightarrow (a-b)(a+b) = -(a-b)$. En divisant

¹⁰ Tiré de dictionnaire du politique la Toupie: http://www.toupie.org/Dictionnaire/Raisonnement_par_absurde.htm

par $(a-b)$ car $a \neq b$, nous obtenons $a+b=-1$. C'est une contradiction car la somme de deux nombres positifs ne peut pas être négative. D'où, par absurde, nous avons montré la validité de la proposition (2.5).

▪ **Raisonnement par contre-exemple :**

Ce raisonnement repose sur le fait que pour prouver qu'une assertion est fautive pour tous les éléments d'un ensemble, il suffit de trouver un élément de cet ensemble (E) qui ne vérifie pas l'assertion. Ceci se traduit par l'équation suivante :

Pour montrer que : $\forall x \in E, P(x)$ est vraie (2.6)

Il suffit de montrer que : $\exists x \in E, P(x)$ est fautive (2.7)

Par exemple, considérant la proposition suivante : 'Tout entier positif est la somme de trois carrés'. Cette proposition est vraie pour le nombre 6 ($6=2^2+1^2+1^2$). Mais elle n'est pas vraie pour les nombres 7 et 15 par exemple. L'existence d'un nombre qui contredit la proposition prouve qu'elle est alors fautive.

▪ **Raisonnement par récurrence :**

Le raisonnement par récurrence repose sur trois étapes pour montrer qu'une proposition $(P(n))$ est vraie pour tout nombre n : (1) initialisation, (2) hérédité et (3) conclusion. Lors de l'initialisation, nous vérifions la proposition pour le nombre 0 ($P(0)$). Pendant la phase d'hérédité, nous admettons la proposition à l'ordre n et nous la prouvons à l'ordre suivant $(n+1)$. Dans la conclusion, nous rappelons que par le principe de récurrence que la proposition $P(n)$ est vraie pour tout nombre n .

Par exemple, nous voulons montrer la proposition suivante :

$$\forall n \in \mathbb{N}, 2^n > n \quad (2.8)$$

Nous procédons alors par trois étapes :

Étape 1 : Pour $n=0, 2^0=1 > 0$: vraie,

Étape 2 : Supposons que la proposition (2.8) est vraie pour tout n et la montrons pour $(n+1)$.

$$2^{n+1} = 2^n + 2^n, \text{ Or } 2^n > n, \quad 2^{n+1} > n + 2^n > n + 1 \quad (\text{car } 2^n > 1) \quad (2.9)$$

D'après (2.9), la proposition est vraie pour $(n+1)$,

Étape 3 : Par le principe de récurrence, $P(n)$ est vraie pour tout $n \geq 0$, c'est-à-dire la proposition (2.8) est validée.

▪ **Raisonnement analogique :**

Le raisonnement analogique procède à une comparaison entre deux éléments avant d'aboutir à une conclusion. Il se base sur l'analogie qui existe entre deux éléments. Le point de départ de ce raisonnement peut être singulier, particulier ou universel. Cela importe peu, à condition toutefois qu'on ait au terme du raisonnement le même degré d'universalité qu'au début. Il y a raisonnement analogique lorsqu'on s'appuie sur le fait que deux éléments se ressemblent sur un point pour montrer qu'ils doivent se ressembler sur un autre. Selon Holyoack et Thagard, l'analogie est « ce qui rend possible de comprendre une situation dans les termes d'une autre » (Holyoak et Thagard, 1989). Les matrices progressives de Raven présentent un exemple de raisonnement analogique à choix multiple.

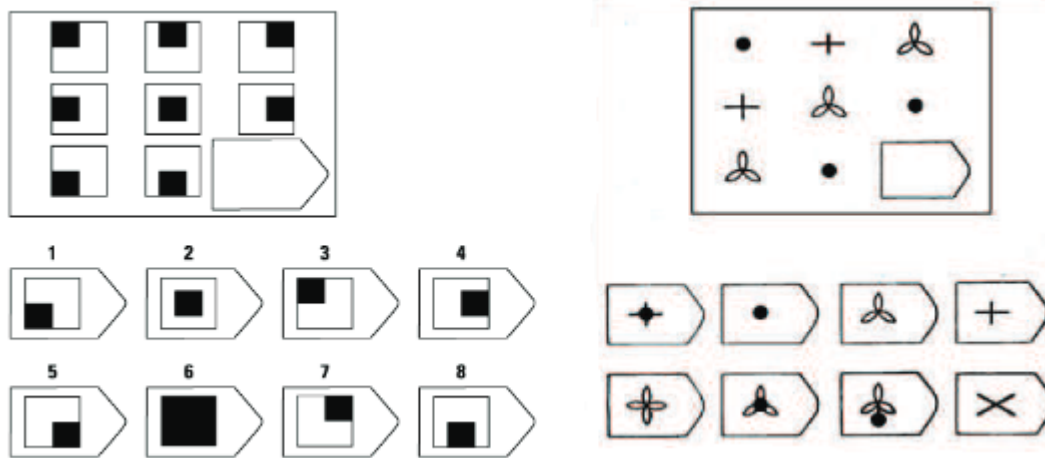


Figure 2. 7. Exemple de deux matrices progressives de Raven (Vezneva, 2011)

2.4.2 Raisonnement intuitif

Le raisonnement intuitif repose sur l'intuition, l'expérience et les décisions affectives pour résoudre un problème. Il s'agit d'une forme d'une intelligence inconsciente (Gigerenzer, 2007; Kahneman, 2002; Kruglanski et Gigerenzer, 2011). Cette forme de raisonnement est très différente du raisonnement analytique qui repose sur le calcul. Il est plutôt basé sur des

comportements affectifs et inconscients. Nous pouvons voir ce type d'intuition tout en pratiquant des sports, des courses, en choisissant des amis, etc.

Selon Betsch, l'intuition elle-même est définie comme le pouvoir d'obtenir des connaissances qui ne peuvent être acquises ni par inférence, ni par observation, ni par le raisonnement logique ni par le calcul. Ces connaissances pourraient être acquises par l'expérience, l'apprentissage associatif et l'information stockée dans la mémoire à long terme (Betsch, 2008). Les gens qui utilisent le raisonnement intuitif s'appuient sur des réponses instantanées, y compris les valences affectives, qui ne nécessitent pas beaucoup d'effort. Nous pouvons donc en déduire que le raisonnement intuitif est basé sur un jugement spontané ou une prise de décision qui repose sur la pratique et l'expérience stockées dans la mémoire à long terme plutôt que sur le raisonnement logique et analytique (Kolb, 1984). Le raisonnement intuitif présente de nombreux avantages. Par exemple, l'intuition peut améliorer d'une façon considérable l'efficacité de la prise de décision. Le raisonnement intuitif est aussi rapide et théorique. Il ne repose que sur la pratique. Cependant, nous devons noter que l'intuition ne fournit pas de prévisions exactes du problème comme le raisonnement analytique.

2.5 Des jeux sérieux aux jeux sérieux intelligents

2.5.1 Jeux Sérieux

Dans la littérature, il existe plusieurs définitions du jeu sérieux (JS). Nous allons commencer cette section par donner quelques définitions des JS, ensuite nous allons distinguer les bienfaits et les caractéristiques des JS. Finalement, nous présentons quelques exemples des JS dans différentes applications.

Parmi les définitions des jeux sérieux, nous avons choisi de mentionner celles données par le grand dictionnaire terminologique de la langue Française, (Michael Zyda, Ben Sawyer et Julian Alvarez) car elles figurent parmi les définitions les plus populaires. Selon ce dictionnaire terminologique, les jeux sérieux sont définis comme étant « une application informatique qui combine une intention sérieuse de type pédagogique, informatif, communicationnel, idéologique ou autre, avec un environnement d'apprentissage prenant la forme d'un jeu vidéo,

afin de transmettre un savoir pratique ou de sensibiliser à un enjeu social ¹¹». Michael Zyda définit les jeux sérieux comme étant « un défi cérébral, joué par un ordinateur selon des règles spécifiques, qui utilisent le divertissement en tant que valeur ajoutée pour la formation et l'entraînement dans les milieux institutionnels ou privés, dans les domaines d'éducation, de la santé, de la sécurité civile, ainsi qu'à des fins de stratégies de communication » (Zyda, 2005). Ben Sawyer définit les jeux sérieux comme étant « des applications informatiques réalisées par des développeurs, des chercheurs, des industriels, qui regardent comment utiliser les technologies associées en dehors du divertissement » (Sawyer et Rejeski, 2002). En s'inspirant et combinant ces deux dernières définitions, Julien Alvarez définit ces jeux comme étant « une application informatique dont l'objectif est de combiner à la fois des aspects sérieux (*Serious*) tel que de manière non exhaustive l'enseignement, l'apprentissage, la communication, ou encore l'information, avec des ressorts ludiques issus du jeu vidéo (*game*). Une telle association a donc pour but de s'écarter du simple divertissement » (Alvarez, 2007). Selon la dernière définition donnée par Alvarez, nous pouvons désigner par jeu sérieux tout jeu vidéo ayant pour objectif principal d'informer, d'expérimenter et de s'entraîner tout en jouant (aspect ludique). Ils peuvent être appliqués dans plusieurs domaines : militaires, gouvernementaux, éducatifs, entreprises, soins de santé, publicité, réseaux sociaux, etc.

Les jeux sérieux (JS) présentent de nombreux avantages : ils augmentent les états mentaux d'engagement, de motivation et d'immersion chez les apprenants à travers leur aspect ludique (Derbali et al., 2013; Ghali et al., 2014; McNamara et Jackson, 2010; Prensky, 2007). En effet, en communiquant avec des interfaces graphiques du jeu vidéo l'apprenant se sent de plus en plus dans un monde virtuel où il communique avec des objets virtuels en 3 dimensions. Ces objets peuvent être manipulés et gérés à sa façon. Le fait d'interagir, de manipuler des objets 3D augmente chez l'apprenant son estime de soi, sa motivation intrinsèque et extrinsèque en gagnant des points de récompenses et en exécutant quelques tâches particulières du jeu pour des buts de jeu. Le fait de jouer en étant concentré et alerté augmente ainsi l'état d'engagement chez l'apprenant. Et finalement, l'excès de motivation peut conduire le joueur à un état d'immersion ou de flot où il sera complètement concentré avec le jeu virtuel et déconnecté du

¹¹ http://www.granddictionnaire.com/ficheOqlf.aspx?Id_Fiche=26502971

monde réel. Ces états sont favorables à l'apprentissage du contenu pédagogique. Les JS peuvent être utilisés et bénéfiques dans différents domaines (éducation, santé, publicité, etc.). Tandis qu'ils présentent certaines limites (Prensky, 2007). Par exemple, un grand nombre de JS se concentrent sur l'aspect ludique qui encourage les apprenants à passer plus de temps à jouer et s'amuser mais ne contribue pas essentiellement à l'acquisition des connaissances. De plus, la plupart des JS présentent un problème sérieux lors de la transition entre les aspects sérieux et non sérieux. La transition entre ces deux aspects doit être bien faite et maîtrisée afin de renforcer l'apprentissage du matériel pédagogique et non des aspects liés aux jeux. Les joueurs ont aussi tendance à passer plus de temps sur les aspects ludiques du jeu et ignorer le contenu du cours à apprendre. Enfin, l'implémentation des JS nécessitent des experts en différents domaines (interfaces hommes machines, informatique, éducation, etc.) pour bien réussir le design et la présentation du contenu pédagogique du jeu.

La première apparition des JS remonte à l'année 1952 où Douglas a créé le premier jeu sérieux. Ce jeu est nommé Oxo et a été conçu pour des fins sérieuses. C'est un jeu vidéo de Tic-Tac-Toe qu'il a conçu dans le cadre de sa thèse sur les interactions hommes machines sur l'ordinateur EDSAC (*Electronic Delay Storage Automatic Calculator*) à l'Université de Cambridge (Yang-Ting, 2011).

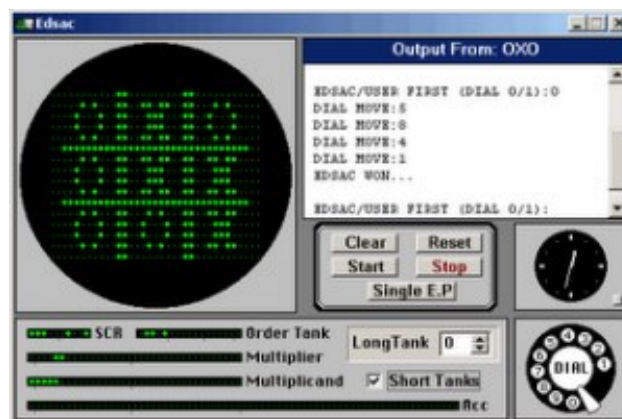


Figure 2. 8 Interface du jeu Oxo¹²

¹² <http://serious.gameclassification.com/FR/games/14983-Oxo/index.html>

Après l'apparition du jeu Oxo, les JS n'ont cessé de se développer pour différentes catégories de population, allant des enfants aux adultes. Ils s'intéressent à plusieurs domaines : éducation, santé, marketing, médecine, sport, etc. Le tableau 2.2 illustre quelques exemples.

Table 2. 2 Exemples de jeux sérieux

	<p>FoodForce : Ce jeu est conçu par le programme alimentaire mondial de l'ONU. Il est constitué de missions afin de sensibiliser les joueurs au problème de la faim dans le monde.</p>
	<p>McDonald : la chaîne de restauration McDonald a réalisé un jeu de restauration qui offre la possibilité pour chaque joueur de simuler et découvrir le processus complexe pour la gestion de tous les détails d'un restaurant.</p>



NFS Underground : C'est un jeu de publicité pour la marque Ferrari. C'est un jeu de course où les joueurs doivent choisir un des modèles de voiture et jouer le jeu. La marque Ferrari est placée dans le jeu pour expliquer ses avantages.



Pulse : Le jeu Pulse est l'un des jeux sérieux les plus célèbres et les plus chers (10 millions de dollars). Ce jeu permet de former les médecins sur des patients virtuels. Grâce à ce jeu virtuel, les médecins peuvent s'entraîner sur des procédures complexes en médecine, avant de s'attaquer au patient lui-même.

2.5.2 Jeux éducatifs

Les jeux éducatifs sont un type particulier de JS qui a pour but principal l'**apprentissage** ou encore l'**éducation**. L'éducation peut toucher à différents domaines : physique, mathématique, médecine, chimie, informatique, etc. Dans ce qui suit, nous décrivons à titre d'exemple quelques travaux de recherche reliés aux jeux éducatifs.

Dans le domaine informatique, Derbali et ses collègues ont développé le jeu *HeapMotiv* pour l'apprentissage d'un cours de structure de données : les tas binaires (Derbali et al., 2013;

Ghali et al., 2014). *HeapMotiv* (illustré dans la figure 2.9) est conçu sous la forme d'un labyrinthe 3D qui comporte plusieurs routes et dont un seul chemin mène à la destination finale (sortie). Sur les chemins de ce labyrinthe, plusieurs panneaux d'informations sont placés pour aider les apprenants à trouver la bonne destination. À chaque fois et avant d'obtenir la bonne destination, l'apprenant doit jouer une mission qui vise à éduquer sur certains aspects de base de tas binaire. Le jeu comporte au total trois missions qui sont présentées dans l'ordre : les deux premières missions (*Tetris* et *Shoot*) sont conçues pour construire un tas binaire et maintenir la propriété de tas, tandis que la troisième mission (*Sort*) est conçue pour montrer les opérations de base pour un tas binaire (insertion et suppression) et l'algorithme de tri.

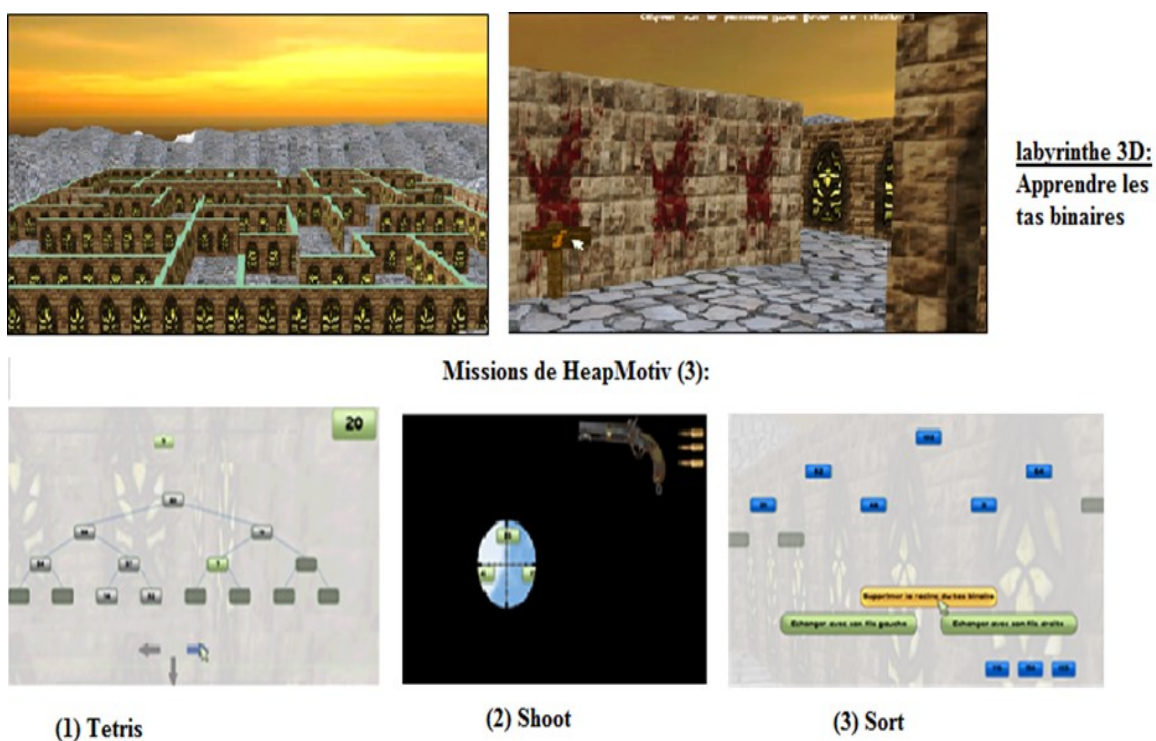


Figure 2. 9 Les différentes missions dans le jeu *HeapMotiv*

Dans le domaine de l'apprentissage des mathématiques, l'équipe EGEGMS (*Electronic Games for Education in Math and Sciences*) de l'Université British Columbia a développé le fameux jeu *PrimeClimb* qui sert à apprendre aux élèves la factorisation des nombres (voir figure 2.10). Dans ce jeu, deux joueurs sont destinés à grimper sur les surfaces de glace de la montagne où sont indiqués des nombres tels qu'illustrés par la figure. Les deux joueurs ne doivent pas se déplacer sur des nombres qui partagent un facteur en commun. Si c'est le cas, le joueur tombe

et perd des point, sinon, il gagne des points et son score augmente. Le jeu dispose aussi d'un outil d'aide de factorisation du nombre qui est présenté à droite de l'écran (Conati, 2002).

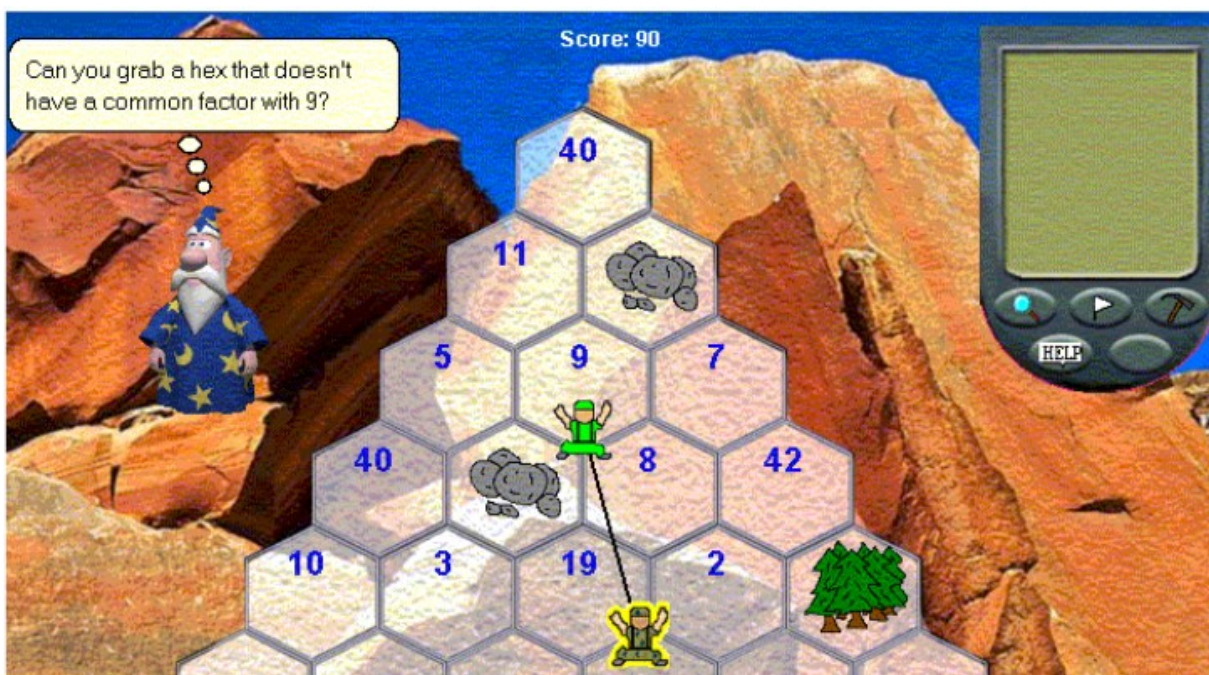


Figure 2. 10 Interface du jeu *PrimeClimb* (Conati, 2002)

Dans le domaine de la physique, Shute et ses collègues ont conçu le jeu *Newton's Playground* (Kai et al., 2015; Shute et al., 2013) destiné à des élèves du secondaire pour apprendre les principes de Newton tout en jouant. *Newton's Playground* ou *Physics Playground* est un jeu à 2 dimensions où les élèves appliquent plusieurs principes et règles de Newton lorsqu'ils dirigent et guident une balle verte vers un ballon rouge placé sur l'écran (voir figure 2.11). Ce jeu permet à l'apprenant d'explorer l'environnement de son choix où il peut se déplacer à volonté. Le jeu favorise ainsi les aspects d'exploration, d'expérience et d'apprentissage. Tous les objets présents dans le jeu obéissent aux lois de Physique, par exemple la loi de gravité et de mouvement de Newton. Dans le jeu, l'environnement compte sept aires de jeu et le joueur peut choisir de jouer dans l'aire qu'il veut. Chaque aire du jeu comporte 10 niveaux de difficultés où chaque niveau consiste en plusieurs obstacles dispersés dans l'espace et en plusieurs positions de la balle dans l'environnement. Chaque niveau présente plusieurs solutions possibles. Les élèves peuvent déplacer la balle à droite ou à gauche mais devront créer des machines simples appelés 'agents de force et de mouvement' pour résoudre les problèmes

du jeu. Dans ce jeu, il existe 4 types d'agents : rampes, pendules, leviers et tremplins. Les élèves utilisent la souris pour dessiner et créer des agents. Les élèves contrôlent le poids et la densité des objets à travers leurs dessins, rendant un objet plus dense, par exemple en le remplissant de plus de lignes (Kai et al., 2015).

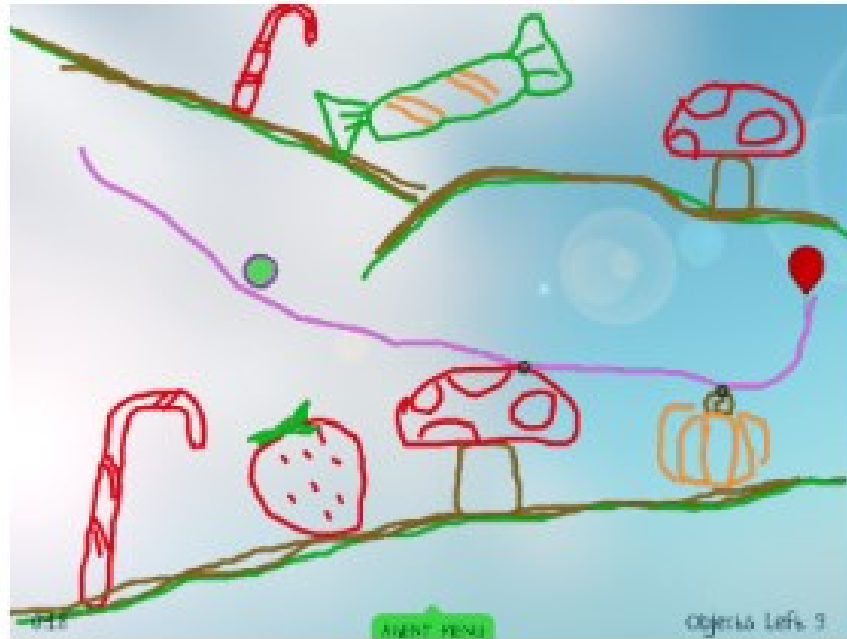


Figure 2. 11 Capture d'écran du jeu *Physics Palyground* (Kai et al., 2015)

Dans le domaine de la médecine et afin de diagnostiquer des patients virtuels, Lester et son équipe ont développé le fameux jeu JSI Crystal Island (Min et al., 2016). Ce jeu est destiné aux élèves du niveau secondaire. Dans ce jeu, le joueur doit identifier la nature de la maladie en collectant un ensemble d'indices sur des patients virtuels. Le joueur doit explorer un monde virtuel qui se présente sous forme d'une île où il doit examiner des patients virtuels. Ce jeu est sous forme d'un environnement narratif qui contient sept buts principaux. Les cinq premiers buts encouragent le joueur à parler avec un autre avatar non joueur pour déterminer les causes de la maladie. Les deux autres buts consistent à tester la nourriture dans le laboratoire et envoyer un diagnostic complet à l'infirmière (Min et al., 2016). Ce jeu s'est montré efficace en termes d'augmentation de la motivation et de l'apprentissage (Rowe et al., 2011). La figure 2.12 présente une capture d'écran de ce jeu.



Figure 2. 12 Environnement virtuel *Crystal Island* (Min et al., 2016)

2.5.3 Jeux Sérieux Intelligents

Les futurs travaux de recherches autour des JS s'orientent vers l'implémentation d'une nouvelle génération des jeux sérieux ou éducatifs à savoir les jeux sérieux intelligents (JSI). Tel que leur nom l'indique, ces jeux doivent être dotés d'une certaine intelligence pour pouvoir répondre adéquatement aux besoins de l'utilisateur. Nous entendons par le terme 'intelligence', le fait d'intégrer des techniques d'intelligence artificielle afin de développer des jeux plus souples et plus adaptatifs selon les caractéristiques individuelles de l'utilisateur.

Dans le but de concevoir des jeux plus efficaces et plus robustes, les chercheurs pensent désormais à développer des jeux sérieux intelligents après le succès de cet environnement d'apprentissage souligné par plusieurs auteurs (Derbali et Frasson, 2010; McNamara et Jackson, 2010; McNamara et al., 2007). Par ailleurs, quelques travaux ont déjà commencé à intégrer des techniques d'intelligence artificielle, et plus particulièrement d'apprentissage machine, pour développer des jeux sérieux plus intelligents (Bernardini et al., 2014; Kai et al., 2015; Lester, Ha, et al., 2013; Lester et al., 2014; Min et al., 2016).

Les travaux actuels sur les jeux sérieux intelligents seront décrits dans la section 2.7 qui s'intéresse à tout ce qui est adaptation. Selon nos connaissances et en se concentrant sur la modélisation de l'apprenant, l'adaptation, que ce soit pour les systèmes tutoriels ou les jeux sérieux, peut être faite essentiellement à six niveaux :

- **Émotions** : il s'agit des travaux autour de la détection et la prédiction des émotions de l'apprenant afin de connaître ses émotions et d'intervenir par conséquent,
- **Reconnaissance des objectifs de l'apprenant** : il s'agit de reconnaître les prochains objectifs de l'apprenant en fonction de ses actions précédentes,
- **Modélisation des connaissances de l'apprenant** : elle s'agit de traçage des connaissances de sa forme et son manière de raisonnement,
- **Aspect narratif** : il s'agit d'organiser l'aspect de narration par les personnages virtuels de l'histoire,
- **Difficulté du jeu** : il s'agit de modifier et d'ajuster (augmenter ou diminuer) le niveau de difficulté du jeu en fonction des performances et des capacités du joueur,
- **Senseurs physiologiques** : il s'agit d'offrir à l'apprenant une adaptation selon des mesures physiologiques qui sont décrites dans ce qui suit.

2.6 Senseurs physiologiques

Après avoir décrit les JS et les JSI, nous décrivons dans cette section un outil parmi d'autres permettant de doter ces plateformes de critère d'intelligence, soit les senseurs physiologiques. Nous portons plus d'attention à expliquer et décrire les concepts liés à l'électroencéphalogramme, l'activité électrodermale et les traceurs de regard.

2.6.1 Electroencéphalogramme

L'électroencéphalogramme est un tracé de l'activité cérébrale du cerveau. On enregistre à l'aide d'électrodes placées sur le cuir chevelu, l'activité électrique produite par la décharge des neurones (Niedermeyer et al., 2011). Les électrodes sont positionnées selon le système international 10/20 (Klem et al., 1999) où chaque position est référencée par une lettre et un chiffre permettant de situer les électrodes sur le cortex cérébral. Les lettres identifient le lobe

correspondant à savoir : « F » pour frontal, « T » pour temporal, « C » pour central, « P » pour pariétal et « O » pour occipital. Les chiffres pairs font référence à l'hémisphère droit et les chiffres impairs font référence à l'hémisphère gauche. La lettre « z » fait référence à une électrode placée au milieu. Les hémisphères cérébraux droit et gauche ont chacun leurs spécialités. Ils sont par ailleurs divisibles en plusieurs régions (postérieure, latérale, antérieure, supérieure dorsale, ventrale, inférieure) correspondant à différents traitements cognitifs (Demos, 2005) ou encore lobes tel que décrit plus tôt dans ce chapitre (lobes frontaux, temporaux, pariétaux et occipitaux). Le signal EEG contient en général des artéfacts et des bruits qui peuvent être filtrés et traités par des filtres passe-bas et passe-haut. Après être amplifié et filtré, le signal EEG peut être divisé en ce que nous appelons des bandes de fréquences.

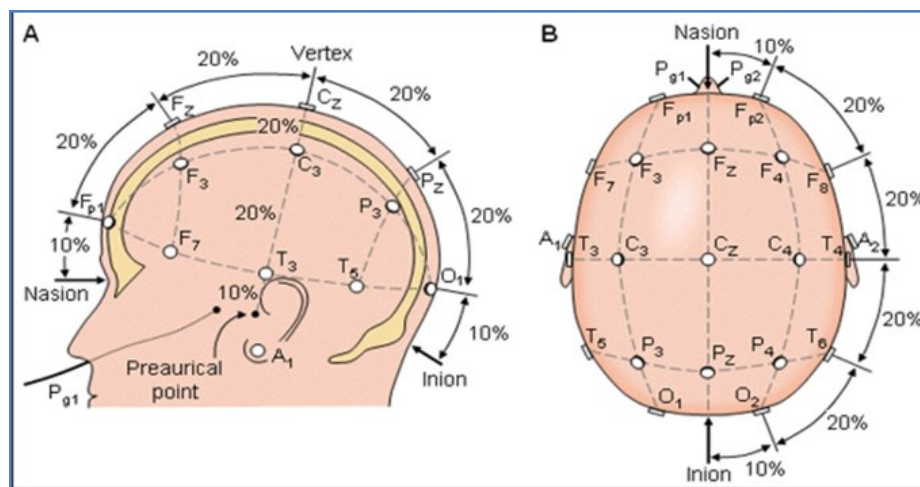


Figure 2. 13 Les positions des électrodes dans le système 10/20 (Klem et al., 1999)

Selon Demos, chaque fréquence est caractérisée par un ensemble de propriétés. Par exemple, les ondes de fréquence delta (1-4Hz) correspondent au sommeil et la résolution des problèmes complexes. Tandis que les ondes Alpha (8-12Hz) correspondent à l'activité d'alerte et de méditation. Le tableau suivant illustre les différentes bandes divisées par des intervalles de fréquence et expliquant les différentes caractéristiques de chacune. Le tableau est extrait de (Demos, 2005) et traduit en français. Il est à noter que cette classification est donnée à titre d'exemple mais elle n'est pas unique. Il existe d'autres auteurs qui regroupent deux bandes en une seule comme les bandes beta ou ils ignorent aussi certaines bandes comme la bande delta

car elle est de fréquence très faible. Pour la mesure de l'activité cérébrale, nous avons utilisé dans cette thèse les casques B-Alert X10¹³ et Emotiv EPOC¹⁴.

Nom de la bande	Fréquence (Hz)	Description des caractéristiques
Delta	1-4	Sommeil, résolution de problèmes complexes
Thêta	4-8	Créativité, vision, états profonds
Alpha	8-12	Alerte et tranquillité, rapidité, méditation
Beta	13-21	Pensée, concentration, attention prolongée
SMR	12-15	Alerte mentale, relaxation physique
High Beta	20-32	Intensité, alerte hyperactif, anxiété
Gamma	38-42	Processus cognitif, apprentissage

Table 2. 3 Distribution des bandes de fréquences (Demos, 2005)

2.6.2 Activité Électrodermale

L'activité électrodermale de la peau (EDA) permet de mesurer le degré de sudation d'un individu (le facteur d'excitation : faible ou fort). Elle représente des changements des propriétés électriques de la peau humaine qui ont lieu lors des interactions avec l'environnement et les états psychologiques d'un individu. Des travaux de recherche ont montré que cette mesure est corrélée avec la charge cognitive, le stress (Mohammad et Nishida, 2010; Setz et al., 2010) et la frustration (Wislow, 2006). En effet, le stress se traduit par une augmentation des glandes

¹³ <http://www.advancedbrainmonitoring.com/xseries/x10/>

¹⁴ <https://www.emotiv.com/epoc/>

sudoripares où l'humidité de la peau augmente et sa résistance électrique diminue. Cette situation est à l'inverse de la relaxation. Dans la figure 2.14 extraite de (Setz et al., 2010), nous pouvons remarquer des pics dans le signal EDA qui apparaissent après un état de repos (*baseline*). Ces pics correspondent à un niveau élevé de stress au contraire de la '*baseline*' qui se manifeste par des mesures d'EDA faibles et par la suite un état de relaxation. Elle est mesurée généralement en microsiemens. Il est à noter que l'EDA varie d'un individu à un autre. Pour la mesure de cette activité, nous avons utilisé dans cette thèse le senseur *Affectiva Q-sensor*¹⁵.

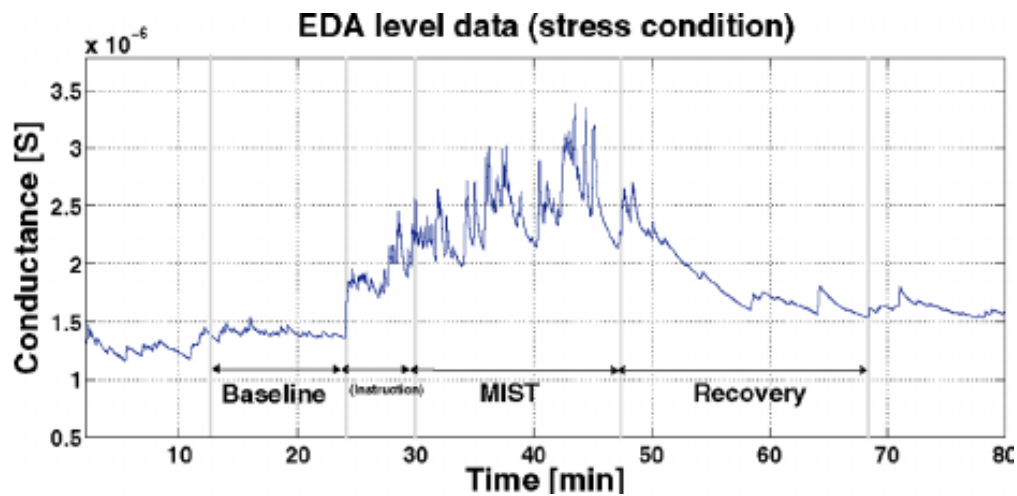


Figure 2. 14 Illustration de l'état de stress mesuré par l'EDA (Setz et al., 2010)

2.6.3 Traceurs de regard

Les traceurs de regard permettent de mesurer la position et les mouvements de l'œil grâce à un champ de lumière infrarouge qui éclaire les yeux et dont le reflet est capturé par des caméras. Dans cette thèse, nous avons utilisé le Tobii Tx 300. Ce traceur de regard se caractérise par une fréquence d'échantillonnage très élevée (300Hz) et par sa robustesse aux mouvements de la tête et aux variations lumineuses. Les données collectées des mouvements oculaires se regroupent principalement en deux grandes catégories :

- **Les mouvements saccadiques** : ce sont les mouvements qui correspondent à un déplacement brusque (très rapide) du regard ou à un saut (*saccade*) dans le but d'explorer

¹⁵ <http://qsensor-support.affectiva.com/>

un environnement ou une scène. Ils peuvent donner un indice sur la compréhension ou non du contenu présenté.

- **Les mouvements de fixation** : ce sont les mouvements qui correspondent à une immobilité ou pause de courte durée de l'œil. Ces mouvements peuvent donner une indication sur la charge cognitive et l'intérêt de la personne à partir de l'extraction de certaines caractéristiques de fixation tels que le nombre, la longueur des fixations et la durée de fixations.

La figure 2.15 présente un exemple de mouvements oculaires face à une page web. Les cercles présentent les fixations : un diamètre de cercle plus grand montre une fixation de durée plus longue, ce qui explique un intérêt plus intense de l'utilisateur sur cette zone et un diamètre plus petit montre moins d'intérêt. Cependant, les saccades correspondent aux déplacements effectués, pour se déplacer d'un point de fixation à un autre. Un déplacement rapide avec un angle de déplacement plus grand peut nous informer que l'utilisateur est en train de sauter dans la page et ne comprends pas bien le contenu.

Le Tobii Tx 300 permet aussi d'extraire d'autres paramètres et, entre autres, les positions des yeux sur l'écran (les abscisses et coordonnées des yeux droite et gauche), les informations sur le regard de données (*gaze data*), la distance entre le traceur de regard et les yeux (droite et gauche) et le diamètre de la pupille (droite et gauche) qui mesure la charge cognitive de travail. Plus ce diamètre est large, plus la charge de travail est importante, ce qui indique que l'apprenant met plus d'effort sur la tâche présentée.

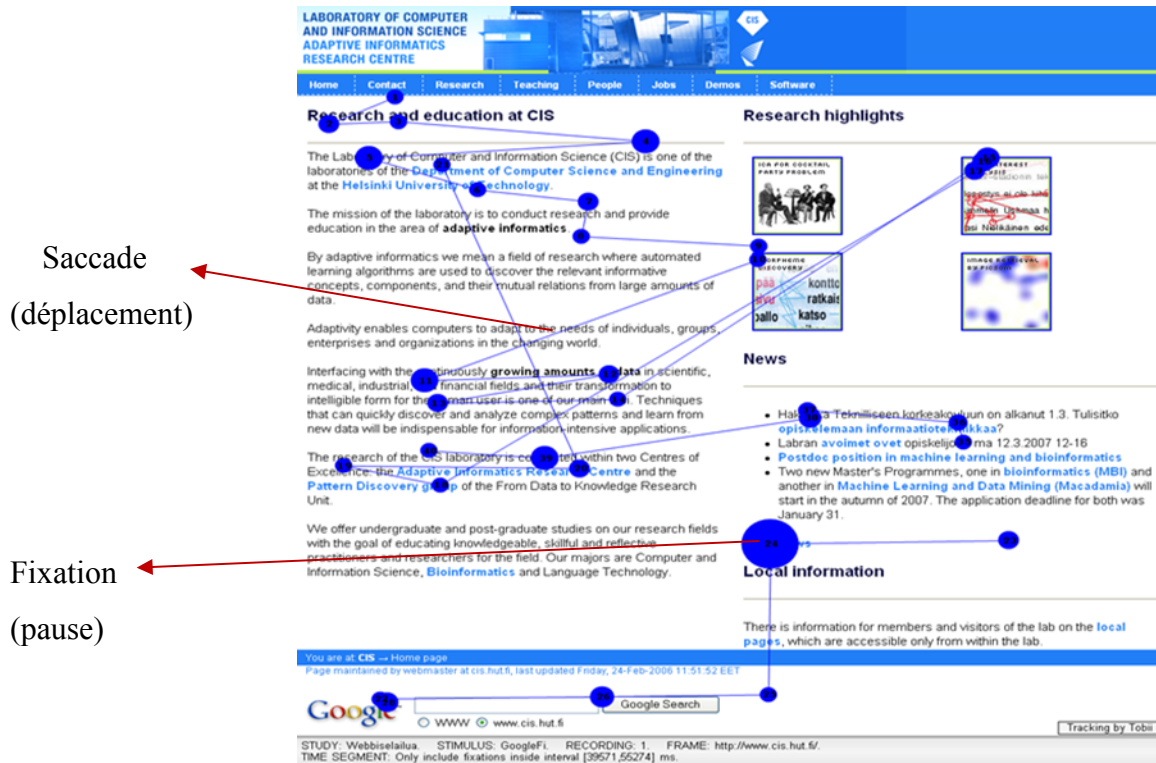


Figure 2. 15 Exemples de saccades et de fixations¹⁶

L'utilisation des traceurs de regard nécessite une phase préliminaire qui s'appelle la calibration. Il s'agit d'ajuster la position assise de la personne en fonction de la position de ses yeux devant les caméras captant les mouvements des yeux. Ceci est fait en suivant quelques points disposés dans différentes positions sur l'écran.

2.7 Adaptation dans les jeux sérieux intelligents

L'adaptation dans les jeux sérieux intelligents (JSI) est une fonctionnalité importante qui permet d'individualiser le jeu selon les caractéristiques individuelles des joueurs. Elle est définie comme une caractéristique exprimée au niveau d'un système informatique qui reflète sa capacité à se modifier structurellement en fonction de certains événements bien identifiés (Andresen et Gronau, 2005). L'adaptation a trois objectifs essentiels : (1) l'efficacité qui évalue le degré du succès avec lequel les utilisateurs réalisent ses objectifs, (2) l'efficience qui évalue les moyens

¹⁶ <http://www.cis.hut.fi/cis/projects/mi/GazePlotCis.png>

mis en œuvre par les utilisateurs pour accomplir les objectifs, et enfin (3) la satisfaction qui évalue le niveau d'acceptation par les utilisateurs (Hocine et al., 2011).

Dans les JSI, il existe plusieurs critères d'adaptation : adaptation du contenu du jeu, adaptation du contrôle, adaptation de la présentation. Il existe aussi des modèles d'adaptation implicite ou explicite (réseaux de neurones, moteur de règles, algorithmes d'apprentissage machine, etc.). Le paramètre d'adaptation peut être le modèle utilisateur ou non. L'adaptation peut se faire sur un seul joueur à la fois ou plusieurs joueurs. La figure ci-après extraite de (Hocine et al., 2011) résume ces différents variables d'adaptation dans les JSI.

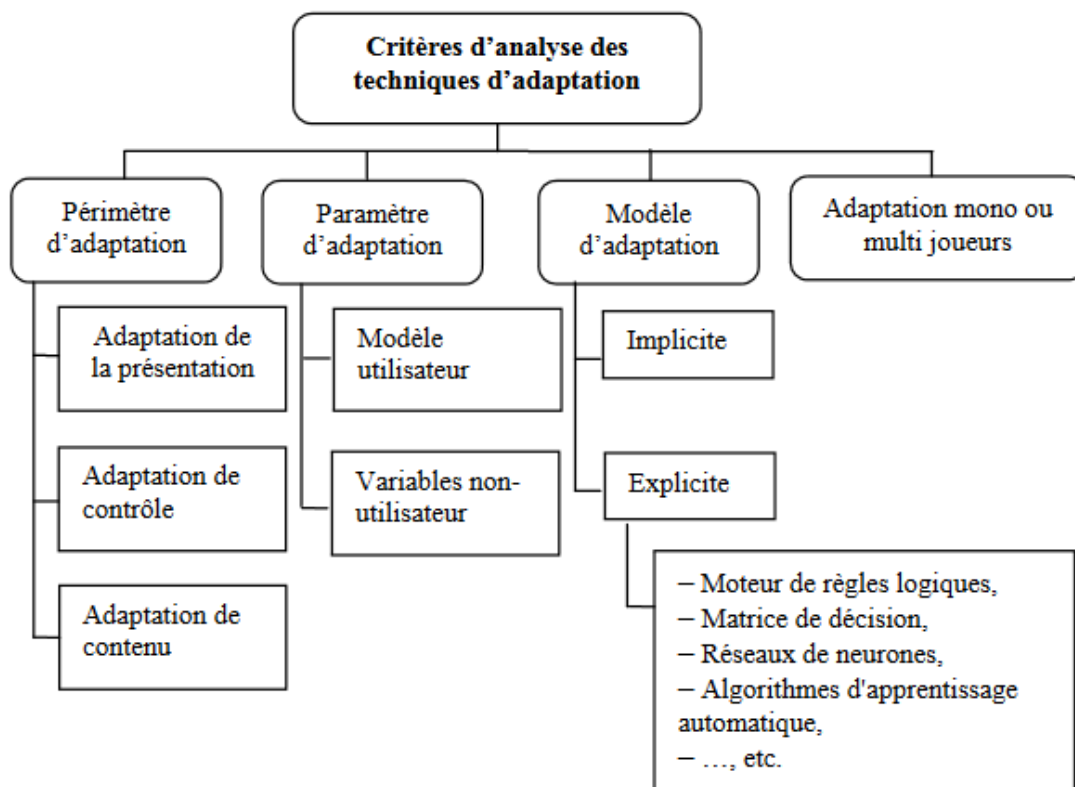


Figure 2. 16 Critères des techniques d'adaptation dans les JSI (Hocine et al., 2011)

Dans ce qui suit, nous nous intéressons seulement à l'adaptation au modèle de joueur (et pas du contenu de jeu ou des événements) en fonction de ses caractéristiques individuelles. Ceci est fait grâce à l'utilisation de modèles explicites d'apprentissage machine. Parmi ces caractéristiques individuelles, nous pouvons citer à titre d'exemple les émotions ressenties (joie, colère, tristesse, etc.), la motivation (est-il motivé à interagir avec le jeu?), l'engagement, la charge cognitive et les objectifs. Dans ce qui suit, nous citons quelques travaux d'adaptation du

modèle utilisateur dans les JSI. Cette revue de littérature expose essentiellement les questions d'adaptation selon les émotions, les états mentaux, les objectifs, les connaissances, l'aspect narratif et la difficulté du jeu. Nous nous sommes essentiellement inspiré des travaux de (Lester, Ha, et al., 2013) qui ont discuté de la possibilité de doter les JS (plus précisément le jeu *Crystal Island*) de critères d'intelligence (pour former des JSI) en se concentrant sur quatre problèmes fondamentaux : (1) la planification de l'aspect narratif du tutoriel (*narrative-centered tutoring planning*), (2) la reconnaissance des émotions de l'apprenant (*student affect recognition*), (3) la modélisation des connaissances de l'apprenant (*student knowledge modeling*) et (4) la reconnaissance des buts de l'apprenant (*student goal recognition*). Nous avons aussi considéré les travaux de (Conati et al., 2013) qui ont discuté de la possibilité de doter le modèle apprenant d'une composante de profil oculaire dans leur jeu *Prime Climb*. De plus, nous avons considéré les travaux de (Derbali et al., 2013) qui ont discuté de la possibilité d'utiliser les senseurs physiologiques pour la prédiction des stratégies motivationnelles à présenter au joueur dans leur JS *HeapMotiv*. Finalement, nous nous sommes inspirés des travaux de (Hocine et al., 2015) qui ont discuté de l'avantage d'utiliser des JS pour améliorer les stratégies de rééducation dans les processus thérapeutiques. L'adaptation se fait essentiellement en variant le niveau de difficulté dans le jeu selon les performances des joueurs.

2.7.1 Adaptation selon l'aspect narratif

Quelques travaux ont cerné la problématique de narration dans les JS. La narration consiste à faire raconter des histoires ou du contenu narratif par des personnages virtuels intégrés dans le jeu. Dans la planification de l'aspect narratif, les travaux intègrent un type particulier d'agent appelé « *Director agent* ». Cet agent orchestre tout le contenu des événements présentés au joueur afin d'améliorer l'expérience de l'utilisateur qui est aussi présenté sous forme d'un personnage virtuel dans l'histoire. Pour créer des histoires riches, l'agent doit lutter contre l'incertitude. L'incertitude peut se révéler dans les interactions du joueur et dans l'absence d'une théorie complète structurée de la narration (Mott et Lester, 2006). Parmi les travaux qui s'intéressent essentiellement à l'aspect de narration, nous décrirons brièvement ceux de Lester dans le JSI *Crystal Island* (Lee et al., 2012; Mott et Lester, 2006; Rowe et al., 2014).

Dans (Mott et Lester, 2006), les auteurs ont présenté une plateforme de narration appelée *U-Director*. Cette plateforme a une architecture de planification narrative basée sur la théorie

afin de modéliser les objectifs narratifs. Elle modélise dynamiquement les objectifs de narration (par exemple la progression et le flot de l'histoire) et les états de l'utilisateur (ses objectifs et ses croyances). Les auteurs ont utilisé les réseaux de décision dynamiques qui sélectionnent dynamiquement les prochaines actions à entreprendre afin de maximiser l'aspect de narration. Des résultats préliminaires ont montré que l'architecture de cette plateforme satisfait les contraintes de narration en temps réel et offrent à l'utilisateur un environnement qui stimule son engagement (Mott et Lester, 2006). L'agent qui dirige cette plateforme est constitué de trois composantes : les états de l'utilisateur, les objectives de narration et l'état de narration.

Dans (Lee et al., 2012), les auteurs ont évalué empiriquement l'utilisation de modèles d'apprentissage machine dans la planification du tutoriel narratif du jeu *Crystal Island*. Ces modèles ont été expérimentés par 183 participants qui sont divisés essentiellement en trois groupes : des étudiants qui reçoivent de l'aide minimale, des étudiants qui reçoivent une aide intermédiaire et des étudiants qui reçoivent une aide complète. En considérant ces trois groupes, les auteurs ont étudié la différence dans le gain de l'apprentissage et les performances des joueurs dans le jeu. Les résultats statistiques (t-test et ANOVA) ont montré que les étudiants qui reçoivent l'aide complète ont des performances meilleures et résolvent les problèmes plus efficacement et ceci en comparant les résultats du pré-test et post-test faites avant et après le jeu avec le groupe des étudiants qui ont reçu l'aide minimale et intermédiaire. Ils ont aussi constaté que les modèles d'apprentissage machine peuvent améliorer les performances liées à la fois à l'apprentissage et l'efficacité du jeu (Lee et al., 2012).

Dans (Rowe et al., 2014), les auteurs ont présenté une plateforme interactive de planification narrative qui utilise l'apprentissage par renforcement pour induire la planification à partir des données de l'apprenant. Le problème est décomposé en différents sous problèmes indépendants qui forment l'aspect de narration. Dans l'apprentissage par renforcement, un agent doit apprendre un ensemble de règles pour sélectionner les actions dans un environnement incertain. Les études empiriques ont montré que les règles de narration interactive avec apprentissage par renforcement peuvent améliorer les performances des joueurs en comparaison avec les méthodes de règles à base d'heuristiques (Rowe et al., 2014).

2.7.2 Adaptation selon les émotions

Quelques travaux de modélisation des émotions dans les JS ont cherché à prédire les émotions du joueur afin d'améliorer son apprentissage et son engagement et d'aboutir à des JSI. La prédiction des émotions peut se faire suivant des approches non intrusives qui n'utilisent pas les senseurs physiologiques (essentiellement des questionnaires auto-évaluatifs) ou en utilisant les senseurs physiologiques (principalement l'électro-encéphalogramme et l'activité électrodermale). Parmi ces travaux, nous citons à titre d'exemples ceux de (Conati et Maclaren, 2009; Kapoor et Picard, 2005; Kim et al., 2009; Lester, Mott, et al., 2013; Sabourin, Mott, et al., 2011; Sabourin et Lester, 2014).

Dans (Sabourin, Mott, et al., 2011; Sabourin et Lester, 2014), les auteurs ont discuté la possibilité de doter le jeu Crystal Island d'une plateforme de prédiction et de reconnaissance des émotions ainsi que leurs valences. Les émotions prédites sont au nombre de sept : anxiété, ennui, confusion, curiosité, excitation, concentration et frustration. La prédiction des émotions s'est faite en utilisant essentiellement les réseaux bayésiens et les réseaux bayésiens dynamiques pour inclure la composante temporelle et augmenter la précision. Pour la reconnaissance des émotions des apprenants, les auteurs ont utilisé le modèle de Elliot et Pekrun (Elliot et Pekrun, 2007). Ce modèle a été empiriquement appris par des données d'interaction des apprenants avec le système. Ces données sont principalement issues d'une méthode subjective d'auto-évaluation des émotions où il est demandé à l'apprenant d'inscrire son émotion sur un appareil portable (*game smartphone device*), et ceci à chaque intervalle de sept secondes. Le joueur peut choisir ainsi une émotion parmi une liste de sept émotions (anxieux, ennuyé, confus, curieux, excité, concentré et frustré). Les réseaux bayésiens sont basés sur trois types de variables : les attributs personnels (personnalité et orientation du but), les variables observables de l'environnement (des captures écrans de l'activité de l'étudiant dans l'environnement) et les attributs d'évaluation (l'évaluation cognitive de l'étudiant sur plusieurs facteurs). Même si ces réseaux bayésiens se sont montrés performants pour la prédiction des émotions et de leur valence (précision de 26,3 % des étiquettes des émotions et de 70,1% pour la valence des émotions (positive ou négative) (Sabourin et Lester, 2014)), ces derniers ne permettent pas de capturer l'évolution des émotions des apprenants dans le temps. De ce fait, les auteurs ont pensé à intégrer les réseaux bayésiens

dynamiques qui se sont montrés plus efficaces où les émotions positives sont reconnues avec une précision de 81,7% vs 61,1% des émotions négatives (Sabourin, Mott, et al., 2011).

Dans (Kim et al., 2009), les auteurs ont incorporé des modèles robustes des émotions dans le JS BiLat destiné aux applications militaires. L'environnement vise à apprendre des personnels du militaire des compétences pour des négociations multiculturelles. Ils utilisent aussi les informations affectives pour diriger les interactions et les réponses des personnages dans le jeu (Lester, Mott, et al., 2013).

Dans (Conati et Maclaren, 2009), les auteurs ont présenté un modèle basé sur les réseaux bayésiens dynamiques pour permettre à un agent intelligent de reconnaître différentes émotions pendant les interactions avec le jeu éducatif *Prime Climb*. Ils modélisent les émotions à partir des causes en se basant sur le modèle OCC (Ortony, 1988). Les émotions modélisées sont : la frustration, l'ennui, la confusion et le flot. Les émotions sont annotées selon un dialogue présenté aux étudiants où il est demandé d'auto-reporter leurs émotions selon une échelle de Likert allant de 1 à 5. Les auteurs discutent aussi de la possibilité d'ajouter des senseurs physiologiques qui permettent d'évaluer la valence et l'activation des émotions afin d'améliorer la précision des prédictions des émotions.

Dans (Kapoor et Picard, 2005), les auteurs ont proposé une approche bayésienne basée sur des classifieurs bayésiens afin de détecter différents niveaux d'intérêts pour les enfants dans un jeu éducatif. Cette approche intègre les senseurs physiologiques tels que la reconnaissance faciale, la reconnaissance de la posture et l'information sur l'état du jeu. Cette approche atteint une précision importante de 86,55%.

2.7.3 Adaptation en fonction de la reconnaissance des objectifs

Lester et son équipe ont étudié la possibilité de modéliser les objectifs du joueur dans le jeu *Crystal Island* (Baikadi et al., 2014; Ha et al., 2011; Min et al., 2014; Min et al., 2016) à partir des observations faites sur les actions du joueur. (Min et al., 2016) ont utilisé essentiellement des techniques d'apprentissage machine avancés tel que l'apprentissage profond¹⁷ ou encore

¹⁷ L'apprentissage profond est une modélisation de haut niveau qui regroupe un ensemble d'apprentissage machine permettant des transformations non linéaires des données.

‘*deep learning*’. En particulier, ils ont intégré les réseaux de neurones récurrents de type ‘*long Short Term Memory*’. Ils ont ainsi développé une plateforme de reconnaissance des objectifs des participants (137) à partir de ces réseaux. Leur plateforme reconnaît précisément les objectifs du joueur en se basant sur les interactions présentes dans les fichiers journaux tout en augmentant la précision par rapport aux autres travaux de littérature (Min et al., 2016). Ils formalisent le problème en une classification multi classe où il s’agit d’entraîner un classifieur sur les données afin de prédire l’objectif correspondant le plus à la séquence observée d’actions et les précédents objectifs.

La reconnaissance des objectifs (*Goal recognition*) consiste à « identifier les objectifs de haut niveau de l’apprenant étant donné une série d’actions de bas niveaux » (Ha et al., 2011; Harrison et al., 2015; Min et al., 2014). La reconnaissance des objectifs des apprenants a fait l’objet de plusieurs travaux, et entre autres ceux de (Bisson et al., 2015; Ha et al., 2011). Dans (Ha et al., 2011) les auteurs ont utilisé les réseaux logiques de Markov pour reconnaître les objectifs des apprenants dans un jeu éducatif et ils ont prouvé une amélioration significative en terme de précision comparé aux précédents n-grammes et aux approches bayésiennes. Dans (Bisson et al., 2015), les auteurs ont présenté des modèles de réseaux de neurones récurrents. Leur modèle apprend la représentation des vecteurs caractéristiques des hypothèses et des résultats. Ensuite, il procède à la classification de la reconnaissance des objectifs. Cette approche est évaluée sur trois domaines incluant le jeu *StarCraft* (Bisson et al., 2015).

2.7.4 Adaptation en fonction des connaissances de l’apprenant

Quelques travaux se sont intéressés à la modélisation des connaissances de l’apprenant. Les connaissances de l’apprenant comprennent ses connaissances sur les scénarios, l’environnement de narration et les stratégies d’interaction. Ces connaissances sont fondamentales dans le contexte de l’aspect narratif interactif et les scénarios de narration pour l’apprentissage et l’éducation. Parmi ces travaux, nous citons à titre d’exemple ceux de (Baikadi et al., 2014; Rowe et Lester, 2010)

Dans (Rowe et Lester, 2010), les auteurs ont proposé d’utiliser une approche basée sur les réseaux bayésiens dynamiques afin de modéliser les connaissances de l’apprenant dans l’environnement narratif incertain. Les réseaux bayésiens dynamiques offrent un formalisme de

représentation unifiée pour représenter les processus temporels stochastiques tels que les processus relatifs à la modélisation des connaissances de l'apprenant. Cette plateforme met à jour dynamiquement un ensemble probabiliste des connaissances de l'apprenant qui sont relatifs à sa compréhension ou non des aspects narratifs présentés, des solutions à des stratégies et des informations sur le curriculum. Les résultats de l'évaluation empirique de ce modèle met l'accent sur différents travaux futurs pour la conception et l'évaluation des modèles des connaissances des apprenants afin de guider la narration et l'adapter au besoin (Rowe et Lester, 2010).

Dans (Baikadi et al., 2014), les auteurs ont exploré les aspects narratifs pour la représentation de la progression des connaissances de l'apprenant avec les réseaux logiques de type Markov et ont montré une amélioration des performances par rapport aux travaux de (Ha et al., 2011).

2.7.5 Adaptation selon les senseurs physiologiques

Quelques chercheurs se sont intéressés à l'utilisation des senseurs électro-physiologiques pour mesurer et inférer certains états chez l'apprenant (tels que son engagement, son attention, sa charge cognitive, sa performance, etc.). Par exemple, (Ghali et al., 2014) ont proposé d'évaluer et d'analyser la variation de l'index d'engagement proposé par (Pope et al., 1995) dans un JS pour l'apprentissage des tas binaires (Derbali et al., 2013). De plus, les auteurs ont étudié l'impact des stratégies motivationnelles sur l'engagement et la motivation des apprenants. Ils ont ainsi pu obtenir une corrélation positive entre l'auto-évaluation d'engagement (choisir une note parmi une échelle allant de 0 à 100) et l'index d'engagement mesuré par cette étude. De même, ils ont obtenu une corrélation positive entre les scores d'engagement (obtenus par auto-évaluation) et les scores de motivation obtenus par le modèle de Keller (Keller, 1987). Cependant, ils n'ont pas pu obtenir une corrélation entre les scores de motivation et les mesures physiologiques d'engagement (basés sur EEG). Ce dernier montre que nous ne pouvons pas estimer une corrélation entre les mesures de motivation et les mesures physiologiques collectés par l'électroencéphalogramme (qui est une tâche difficile).

Dans (Derbali et al., 2013), les auteurs ont discuté de la possibilité de modéliser les stratégies motivationnelles proposées par (Keller, 1987) à partir des données issues de l'activité

électrodermale, le rythme cardiaque et les données EEG. Leur approche se base essentiellement sur les chaînes de Markov cachées. Ils ont montré qu'il est possible d'établir une classification des stratégies motivationnelles en se basant sur le modèle de Keller et les données physiologiques collectées.

Cristina Conati et son équipe (Conati et al., 2013) ont discuté de la possibilité de développer des stratégies qui permettent de décider quand et comment intégrer les conseils dans leur jeu éducatif *Prime Climb* décrit ci-avant, et ceci en utilisant un modèle oculaire (principalement un traceur de regard) dans le modèle de joueur. Elle a essentiellement utilisé les données de suivi du regard qui ont permis de capter l'attention des apprenants à partir d'un certain nombre de statistiques faites sur les données oculaires (durée moyenne de fixation, longueur de message de conseil, type de conseil et nombre total de fixations pour chaque type de conseil).

2.7.6 Adaptation selon le niveau de difficulté dans le jeu

Les JS semble prendre d'autres aspects pour parfois servir à l'abstraction d'autres moyens de guérir les gens et/ou faciliter la rééducation. Dans ce cadre, Nadia El Hocine et ses collègues (Hocine et al., 2015) ont développé un JS destiné aux patients qui ont subi un accident vasculaire cérébral (AVC) pour améliorer les traitements thérapeutiques après cet accident. Leur jeu augmente la motivation et l'immersion chez les patients. Ce JS intitulé Prehab (*Platform games for Rehabilitation*). Il est composé d'un ensemble de jeux pour des fins thérapeutiques lors des séances de rééducation des membres supérieurs. Prehab se caractérise par la capacité d'adaptation de son niveau de difficulté en fonction de profil du patient (ses performances et ses capacités). Le but de ce jeu thérapeutique est d'augmenter l'amplitude de mouvement et le nombre de tâches effectuées dans une séance de rééducation. Dans ce JS, les auteurs ont testé trois stratégies d'adaptation de niveau de difficulté en fonction des capacités des patients: (1) stratégie de difficulté dynamique où le niveau de difficulté peut augmenter ou diminuer en fonction des performances des joueurs, (2) stratégie de difficulté ascendante où le niveau de difficulté augmente à chaque fois et (3) stratégie de difficulté aléatoire (peut augmenter ou diminuer aléatoirement sans tenir en compte des performances des patients). Ils ont fait une expérience dans laquelle ont participé 7 patients qui ont subi un AVC et 3 thérapeutes. Ces participants ont expérimenté les trois versions du JS sur une tablette. Les résultats ont montré

que la stratégie de difficulté dynamique (technique d'adaptation dynamique) augmente l'amplitude du mouvement pendant une séance thérapeutique et par conséquent améliore la guérison des patients (Hocine et al., 2015).

2.8 Conclusion

Dans ce chapitre, nous avons exposé un état de l'art sur les différents types de raisonnement, les jeux sérieux et les méthodes d'adaptation dans les jeux sérieux afin de créer des environnements plus intelligents et plus adaptatifs selon les besoins et l'évolution du module de l'apprenant (JSI).

Dans cette thèse, nous avons utilisé les deux types de raisonnement, à la fois dans notre environnement purement cognitif logique, et dans les deux jeux sérieux *LewiSpace* et *Inertia* qui sont présentés dans les chapitres suivants. Nous avons choisi d'utiliser ces deux types de raisonnement pour explorer la différence qui existe entre ces deux types, de même pour voir quel raisonnement est plus adéquat pour les apprenants. De plus, nous avons opté pour une approche de mesure physiologique qui se base sur les données physiologiques pour faire de la classification ou de la prédiction des performances ou du besoin d'aide de l'apprenant. Les mesures physiologiques se sont montrées efficaces puisqu'elles reflètent des états actuels et précis de l'apprenant tels que les états mentaux d'engagement, de frustration et de stress (Bartels et Marshall, 2012; Ghergulescu et Muntean, 2014; Mohammad et Nishida, 2010). De même, nous avons cerné la question d'adaptation selon le besoin d'aide de l'apprenant lors de ses interactions avec les JS qui n'est pas trop prise en considération dans l'état actuel de la littérature. En effet, lors de son interaction avec un environnement informatique, l'apprenant peut avoir besoin de l'aide mais cet état n'est pas détectable immédiatement. Le fait de détecter adéquatement et au bon moment si ce dernier a besoin d'une aide non sollicitée peut améliorer considérablement ses performances et ses interactions avec le système informatique. Cette question a été étudiée en entraînant et testant différents algorithmes d'apprentissage machine et une méthode auto-évaluative du besoin d'aide. De plus, nous voulons élargir la prédiction d'aide par des stratégies d'aide afin d'améliorer le raisonnement intuitif chez les apprenants et par la suite améliorer leurs performances. Le but est d'améliorer l'état cognitif de l'apprenant et ses conditions d'apprentissage.

Cette thèse comporte trois contributions majeures qui sont présentées sous forme de cinq chapitres (articles scientifiques). Les contributions sont les suivantes : (1) classification et régression des scores de l'apprenant dans des environnements cognitifs (environnement logique constitué de trois catégories de tâches cognitives) qui se rapporte à notre premier objectif, (2) prédiction automatique du score de l'apprenant et de son besoin d'aide dans les jeux vidéo éducatifs 3D (JS *LewiSpace* pour l'apprentissage de diagrammes de Lewis) qui se rapporte à notre deuxième objectif et (3) utilisation des jeux virtuels 3D comme moyen pour améliorer le raisonnement intuitif (JSI *Inertia*: un jeu intuitif pour l'apprentissage de notions de physique) qui se rapporte à notre troisième et dernier objectif. La première contribution sera présentée sous forme de deux articles scientifiques décrits respectivement dans le chapitre 3 et le chapitre 4. La deuxième contribution sera aussi présentée sous forme de deux articles décrits dans le chapitre 5 et 6. Et finalement, la dernière contribution sera présentée par un seul article présenté dans le chapitre 7.

Chapitre 3

Une Étude empirique dans l'environnement purement cognitif logique

Dans ce chapitre, nous présentons une première étude réalisée dans le cadre d'un environnement purement cognitif, logique. Cet environnement est constitué de trois catégories de tâches cognitives : la mémorisation, la concentration et le raisonnement. Elles forment au total sept types de tâches différentes : *feature match* (FM), *rotation* (RO), *forward digit span* (FDS), *backward digit span* (BDS), *arithmetic addition* (AA), *odd one out* (OO) et *intuitive reasoning* (IR). Ces tâches sont extraites du site web « Cambridge Brain Sciences¹⁸ ». Notre objectif principal est d'étudier pour chaque catégorie et chaque type de tâche si les performances des apprenants sont significativement différentes. Nous avons enregistré les mesures physiologiques pour étudier principalement l'évolution de trois états mentaux : engagement, distraction et charge cognitive en fonction des performances de l'apprenant, de la catégorie et de type de la tâche présentée. L'utilisation des mesures physiologiques est motivée par le fait que ces senseurs donnent des mesures réelles sur les différents états mentaux de l'apprenant (EEG) ainsi que son niveau de stress (GSR : *Galvanic Skin Response*). Pour ce faire, nous avons conduit une étude expérimentale sur 20 participants volontaires de l'Université de Montréal. L'étude a durée **90**

¹⁸ <http://www.cambridgebrainsciences.com/>

mn pour chaque participant sur une durée totale de **1 mois**. Les participants interagissent avec notre environnement cognitif développé en Java. Ma contribution dans ce travail consiste à étudier l'évolution des mesures physiologiques en fonction de la catégorie et le type de la tâche présentée. Cette contribution est nouvelle et n'a pas été l'objet d'autres travaux dans notre domaine. Nous avons aussi visé à prédire les performances de l'apprenant (ses résultats dans les tâches cognitives) en utilisant le logiciel SPSS et des modèles **linéaires** tels que la régression multiple. Pour résumer, les données collectées lors de l'analyse expérimentale ou pour la prédiction des résultats de l'apprenant sont les suivantes :

- Les états mentaux extraits de l'électroencéphalogramme B-Alert X10 qui sont de 3 types : engagement, charge cognitive et distraction. Ces états mentaux sont calculés et validés par Chris Berka et ses collègues (Berka et al., 2004; Stevens et al., 2007a; Stevens et al., 2007b). Ce sont des variables quantitatives continues dont les valeurs varient entre 0 et 1,
- Les mesures faites par le *Q-sensor* pour l'activité électrodermale : GSR. C'est une variable quantitative continue dont les valeurs varient normalement entre 5 et 10 microsiemens,
- Les résultats de l'apprenant regroupés par chaque type de tâche cognitive et chaque catégorie (variables qualitatives nominales), de même que le résultat global (variable quantitative continue qui varie entre 0 et 100).

Dans cette première étude, nous avons opté pour une approche purement empirique où nous avons analysé la distribution des mesures de EEG et de GSR dans notre environnement cognitif (nous avons illustré sur les axes des abscisses le nombre de participants (18) et sur les axes des ordonnées les mesures physiologiques avec les variations décrites précédemment). Cette approche est faite essentiellement de quatre étapes : la première et la deuxième étape consistent à collecter les données des participants pour les différents senseurs physiologiques (EEG et GSR), la troisième étape consiste à analyser les résultats obtenus et la dernière étape consiste à prédire les performances de l'apprenant avec des modèles linéaires.

Dans un premier temps, nous avons analysé la distribution des trois mesures EEG (engagement, charge cognitive et distraction) et une mesure de GSR, nous avons étudié leur

variation selon les apprenants. Chaque mesure est comparée avec la moyenne de tous les participants. Des résultats préliminaires ont montré que pour la plupart des participants le niveau de distraction est au-dessous de la moyenne, ce qui montre que la plupart des participants sont concentrés sur les tâches. Les mesures de GSR ne sont pas aussi concluantes du fait que nous avons obtenu pour la plupart des participants des mesures autour de 0. Alors, nous avons considéré ces mesures comme non significatives. Nous avons comparé la variation des trois mesures de EEG selon la catégorie et le type de la tâche cognitive. Nous avons constaté que la catégorie de raisonnement qui a le niveau de charge cognitive le plus élevé, est le raisonnement intuitif. Des analyses de variances ont montré aussi que l'engagement et la charge cognitive dépendent de la catégorie de la tâche présentée.

Dans un deuxième temps, nous avons étudié la possibilité de prédire les performances de l'apprenant en fonction de différents paramètres : la distraction, l'engagement, la charge cognitive, la difficulté de la tâche et la durée. Nous avons pu obtenir un modèle de régression multiple où le résultat est exprimé en fonction de ces paramètres. En effet, il existe une corrélation négative entre le score et le niveau de difficulté (variable quantitative discrète). Nous avons aussi noté que l'information sur la tâche (le type de la tâche, la difficulté de la tâche et la durée) joue un rôle important sur la prédiction du résultat de l'apprenant et qu'en éliminant ces variables et utilisant seulement les variables issues de EEG cela diminue la corrélation entre ces derniers et le score.

Le reste de ce chapitre est constitué de l'article intitulé « Using Electrophysiological Features in Cognitive Tasks : An Empirical study » publié dans le journal *International Journal of Information and Education Technology* en 2016 et la conférence *International Conference on Educational and Information Technology 2015*. Nous rappelons que ma contribution essentielle consiste à montrer le rôle de la catégorie de la tâche sur la variation des mesures physiologiques (états mentaux extraits de EEG) et la prédiction des performances de l'apprenant (ses résultats).

Using Electrophysiological Features in Cognitive Tasks: An Empirical Study

Ramla Ghali, Sébastien Ouellet, Claude Frasson

Published in: International Journal of Engineering and Innovative Technology, 6(8), 584-590, 2016

<http://www.ijiet.org/vol6/756-IT320.pdf>

Abstract

Learners' performances in intelligent tutoring systems or e-learning environments depend on various factors such as the nature of the task presented, their cognitive and affective abilities, etc. In this paper, we focus on studying in detail the variation of these different factors and more specifically the electroencephalogram (EEG metrics) and how they differ according to a category and a type of cognitive tasks. We also studied the possibility of predicting a learner's performance using feature selection and multiple regressions. Primarily, results shows that learners' scores could be predicted using in descending order the difficulty level of the task, the type of a task, the duration of a task and the EEG workload metric by building a multiple regression model that fit our data.

Index Items: Cognitive tasks, EEG features, engagement, workload, distraction, multiple regression

3.1 Introduction

To date, several studies have focused on intelligent tutoring systems and e-learning environments in order to offer a more reliable and adjustable learning experience according to the learners' capacities and performances. For this purpose, several approaches have been proposed in order to assess, predict and improve the learners' performance. However, in our knowledge, there is no study that takes into consideration the category or the type of a learning task. Yet, we think that learner's performances depend strongly on the category and/or the type of a proposed task.

From these approaches, we cited as an example some works in the field of artificial intelligence, human computer interaction, cognition and neuroscience (Baker et Rossi, 2013; D'Mello et al., 2007 ; Pope et al., 1995; Wilson, 2004). These works were mainly based on the use of electrophysiological metrics and log files in order to assess more accurately certain learner's states or behaviors (emotions, workload, distraction, engagement, performance, gaming the system, off-task behavior, etc.). In this paper, we focus only on the use of electrophysiological metrics and log files to study mainly the variation of these metrics and the learners' performance according to different types of cognitive tasks that we implemented.

Electroencephalogram (EEG) was processed by applying Fast Fourier Transformation (FFT) on the signals and extracting from the latter five main power spectral densities (Delta:0-4Hz; Theta:4-8Hz; Alpha:8-12Hz; Beta:12-40Hz and Gamma:>40Hz), and using these measures, Stevens and his colleagues has established three main metrics: mental engagement, mental workload, and distraction (Stevens et al., 2007b). Mental engagement is related to the level of mental vigilance and alertness during the task (high or low states of vigilance). Mental workload can also be seen as the mental vigilance and cognitive load in a particular task. However, distraction or drowsiness reflects the feeling of being sleepy and lethargic (Stevens et al., 2007b).

Galvanic Skin Response (GSR) was used to assess the electrodermal activity and the skin temperature. It measures actually the individual degree of sweating, known as the excitation factor of emotions (low or strong). It is also correlated with cognitive load, stress (Mohammad et Nishida, 2010) and frustration (Wislow, 2006).

Other than the use of EEG and GSR sensors, researchers employ more complex sensors (such as eye trackers, kinect2 depth sensor, posture detection, etc.). In this paper, we use only two main sensors (EEG and GSR) and we conducted an empirical study where we collected data from our cognitive environments (composed of 7 tasks in total and 3 main categories: memory, concentration and reasoning). We also studied the variation of these outputs on the participants, and finally built a multiple regression model using the task information (difficulty level, type of a task, duration and workload EEG metric) to further predict the learners' performance in our cognitive environment.

In the following, we present previous works to our study, a description of our cognitive environment, our experiment, some exploratory and statistical results using EEG features (engagement, workload, and distraction) and GSR, and finally a linear model that uses feature selection in order to provide a moderately useful prediction of the learners' performance.

3.2 Previous Works

In intelligent tutoring systems and e-learning environments, the detection and assessment of some learners' states (for example: engagement, distraction, motivation, emotions, etc.) remains of a great importance in order to build more adaptive and effective human computer interaction systems. For instance, an intelligent system that is able to automatically identify if the student is engaged or not in a task and adjust its content according to this variation, would be more effective for the student and could significantly improve his performance. Moreover, an intelligent system that detects if a student is sleepy and distracted while accomplishing a task using a machine learning algorithm for a classification of this state, may be used in order to adjust its learning strategy and provide a learner with more suitable content and/or feedback that allows him to avoid these situations of disengagement.

In order to detect and regulate the disengagement behavior, some researchers have used log files, data mining and machine learning techniques. Among these works, we cite as an example those of Baker and his colleagues (Baker et Rossi, 2013; Baker, 2007). In (Baker et Rossi, 2013), Baker has established a classification of a student disengagement behavior into four main categories: (1) off-task behavior, (2) gaming the system, (3) carelessness and (4) without thinking behavior. He also discussed the conditions that promoted the occurrence of

each type of these behaviors and proposed some methods to detect them automatically. Whereas, in ref. [8], Baker was interested more specifically to off-task behavior that occur when a student is inactive or doing some actions which are not related to the learning content (speak to other students, stay inactive, etc.). This behavior was initially modeled by extracting certain features and using latent variable models as a machine learning technique.

Other studies have adopted a different approach by using mainly the outputs of electrophysiological sensors, such as EEG, GSR, and HR (heart rate). For example, Pope and his colleagues [1] at NASA have developed an EEG engagement index based on brainwave band power spectral densities and applied in a closed-loop system to modulate task allocation. They proved that performance in a vigilance task improved when this index was used as a criterion for switching between manual and automated piloting. Berka and her colleagues (Berka et al., 2004) explored the feasibility of monitoring EEG indices of engagement and workload acquired and measured during cognitive tests. Results showed an increase of these two indices during the encoding period of verbal and image learning and memory tests compared with the recognition period. She showed also that workload increased linearly with the level of difficulty and that the EEG measures correlated with both subjective and objective performance metrics. Mostow, Chang and Nelson (Mostow et al., 2011) used EEG (single channel) to distinguish the reading of adults and children and to predict the difficulty of the text. Nourbakhch and his colleagues (Nourbakhsh et al., 2012) have used GSR as an indicator of cognitive load and emotions. They have assessed GSR data from two experiments: (1) during text reading tasks and (2) during arithmetic tasks that have different levels of difficulty. From these experiments, they found a strong relationship between GSR features and mental workload. Moreover, GSR data was also used in order to detect emotions (Nakasone et al., 2005; Wislow, 2006) and stress states (Mohammad et Nishida, 2010; Setz et al., 2010). Other studies combined the outputs of different sensors to build more reliable models. For instance, in (Jraidi et al., 2013), a framework based on Bayesian network was built to recognize three trends of the interactions, namely: flow (a total involvement within the task), stuck (a difficulty to maintain focused attention) and off-task (a drop out from the task), as well as the emotional responses occurring subsequently. EEG and physiological devices (GSR and HR) were used and the resulting evaluations showed an accuracy rate of 82% to characterize a positive vs. a negative experience, and an accuracy

ranging from 81% to 90% was achieved to assess four emotions related to the interaction namely stress, confusion, frustration and boredom.

Other than assessing learners' emotions and behaviors, some works have focused on using machine learning technique in order to predict in advance a learner's performance. For example, Beal and Galan (Galan et Beal, 2012) studied the use of certain mental states extracted from the electroencephalogram established by (Stevens et al., 2007b) as two fundamental features that predict the success or failure of the learner in the next task. These two mental states (engagement and workload) have been tested and trained on a support vector machine (SVM) to construct fairly reliable models of learner performance prediction for mathematical problems. Naser (Naser, 2012) used knowledge extracted from learner models to determine the academic performance of a learner in order to offer him the proper difficulty level of problems to solve. Recently, the US Army Research Laboratory (ARL) evaluated and developed concepts for two types of learner models: (1) short-term learners and (2) long-term learners. Sottolare (Sottolare, 2014) also studied the feasibility of using recorded learners' performances in the development of adaptive tutoring learning chains to design future learner models.

3.3 Description of our cognitive tasks

In this section, we describe the cognitive environment we used to study the evolution of a learner's profile which is mainly based on three parameters: his EEG metrics, his GSR and the nature of a cognitive task presented. This environment is mainly divided in the three categories mentioned below:

3.3.1 Memory

This category contains two subcategories of the famous task of Digit Span (DS) where the learner is shown a series of numbers according to an ascending difficulty level and asked to remember and type them in the appropriate place. (1) Forward Digit Span (FDS) consists of typing the numbers in the same order that they appeared on the screen and (2) Backward Digit Span (BDS) where numbers should be entered in reverse.

3.3.2 Concentration

This category has two different tasks: (1) Feature Match and (2) Rotation. Feature Match (FM) consists in identifying whether the two images appearing on the screen are identical or not according to their forms, numbers and colors (see figure below). It has also six difficulty levels (ranging from 1 to 6) which vary in their geometrical number and forms. However, Rotation task (RT) consists in identifying whether two images are identical or not if any rotation is allowed. It has also five difficulty levels.



Figure 3. 1 Feature Match Task

3.3.3 Reasoning

This category has three subcategories of tasks: (1) Arithmetic Addition, (2) Odd One Out and (3) Intuitive Reasoning. In the Arithmetic Addition task (AA), the learner is asked to add two numbers and the difficulty level (3 difficulty levels were implemented) increases according to the numbers of digits to add. In Odd One Out (OO), the learner has to identify the odd image from a series of nine images appearing on the screen, and has four difficulty levels. Finally, for Intuitive Reasoning (IR), it has three levels of difficulty (varying according a time constraint: unlimited, 1 minute and 30 seconds), and 15 series in total; every level contains 5 series of exercises. Unlike other tasks, this task is based on intuitive or analogical reasoning (see fig 3. 2).



Figure 3. 2 Example of Intuitive Reasoning Task

3.4 Description of the experiment, gathering the data and our approach

3.4.1 Experimental Protocol

In order to study a learner's profile's evolution during the completion of our cognitive tasks, we conducted an experiment to gather several types of data from different subjects. In this study, all the participants (20 in total) are from the Université de Montreal. They are wearing an EEG helmet (B-Alert X10) and a bracelet (Affectiva Q-sensor). Participants are from different disciplines, and they are distributed as follows: 9 women and 11 men, a mean age of 28, and a standard deviation of 4.67. The study is located in our research laboratory, where we explained the objectives of the study and asked the participants to sign a consent form (ethics agreement) before starting the experiment. The study lasted about 2 hours, during which they performed the following four steps: (1) installation of the B-Alert X10 headset on the participant (mean duration of 20-25 minutes), (2) installation of Q-sensor bracelet (about 2 minutes), (3) the participant is invited to do 3 baseline tasks defined by the manufacturer of the headset to establish a classification of mental states of engagement, workload and distraction (mean duration \approx 20 minutes), and (4) the participant is finally invited to use our cognitive

environment, which includes different sets of cognitive tasks as described in the previous section (mean duration \approx 80 minutes). Fig. 3.3 illustrates the whole process.

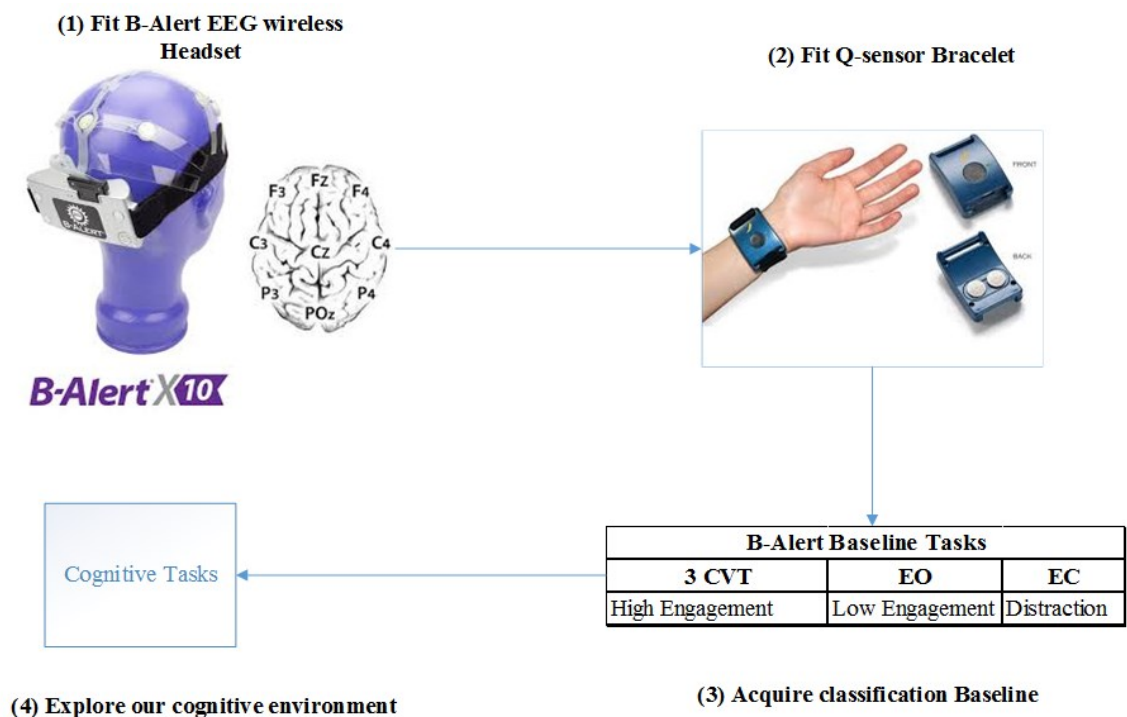


Figure 3.3 Experimental proces

3.4.2 Data recording and preprocessing

The electroencephalogram (EEG) was recorded from 9 sensors (F3, Fz, F4, C3,Cz, C4, P3, Poz and P4) integrated into a cap covering the upper half of the head, with 2 reference signals attached to the mastoid bones (behind the ears). The location of each sensor was determined by the international 10-20 system. This cap was equipped with a small USB dongle which received the wireless transmissions connected to a PC computer. Each second, 256 EEG signals were transmitted and converted to power spectral densities (Alpha, Beta, Theta and Sigma) and processed by the *B-Alert* software which is the proprietary of Advanced Brain Monitoring Company¹⁹. This software allows us to obtain a real time classification of certain mental states (Sleep Onset, Distraction, Low Engagement, High Engagement and High Workload). In this

¹⁹ [18] <http://advancedbrainmonitoring.com/xseries/x10/>

study, we selected the mental states of High Engagement, Workload and Distraction. Galvanic Skin Conductance (GSR) also known as ElectroDermal Activity (EDA) was measured using *Affectiva Q-sensor* bracelet 2.0 at a sampling rate of 32 Hz. This bracelet allows us to measure three features: skin conductance (GSR or EDA), skin temperature and an accelerometer that outputs three-dimensional displacements of the sensors. From 20 participants, only 18 data subjects were considered in the data analysis for EEG and only 16 data for GSR due to missing recordings. The EEG mental states and the GSR measures were manually synchronized with all the tasks of the cognitive environment based on the date and time extracted from log files. We calculated for each category and for each level of a task the EEG mental states values and the GSR values in order to analyze their variation depending on category and difficulty level of a task. All the data was preprocessed and labeled according to SPSS rules.

3.4.3 Our approach

In this section, we show a diagram that summarizes our proposed approach. We recall that we are mainly based on the study of the results from the experiment described above and the collected data from two sensors (EEG and GSR). In the first step of this approach, we gathered data from different students based on the sensors and students' interaction with our cognitive tasks (step 2). In the third step, we analyzed and studied the obtained results. Finally (step4), we predicted learners' performances (scores) using mainly multiple regression algorithms.

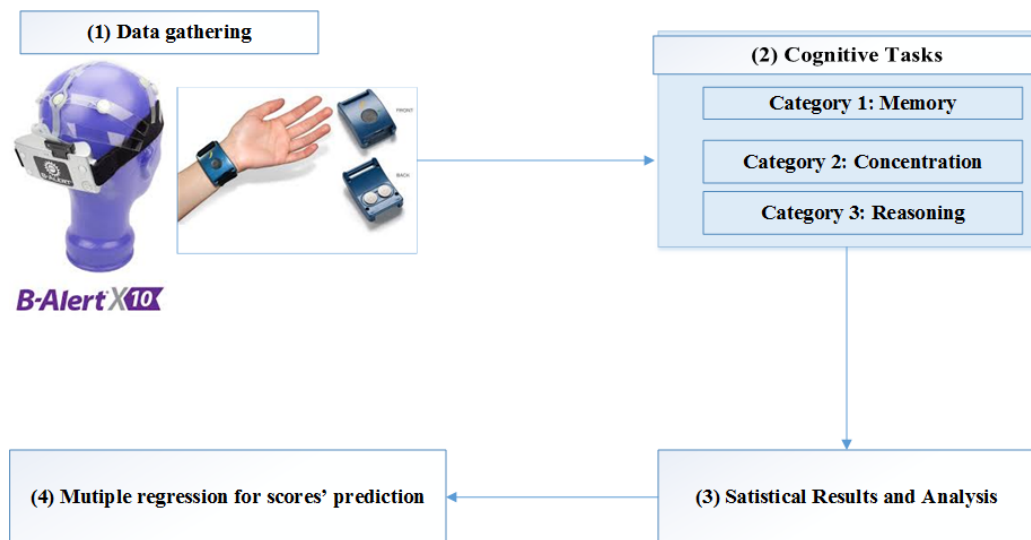


Figure 3. 4 Our proposed approach

3.5 Experimental Results

In this part, we analyze the variation of EEG and GSR metrics among all the participants. We also take into consideration the type and/or the category of a task in our study. It is mainly divided into two parts: (1) the study of EEG and GSR evolution during the cognitive tasks and (2) the possibility of predicting the learner's score using linear models such as regression.

3.5.1 EEG and GSR metrics evolution in learning

In the first step, we studied the distribution of the electrophysiological metrics (EEG indices of engagement, workload and distraction) between the learners during all the cognitive tasks.

For EEG indices (distraction, engagement and workload) and for GSR metrics, we calculated the average of these metrics for each participant during the completion of our cognitive environment and the average for all the participants for each metric. We consider the latter as a threshold that we represent with a red solid line in the following figures.

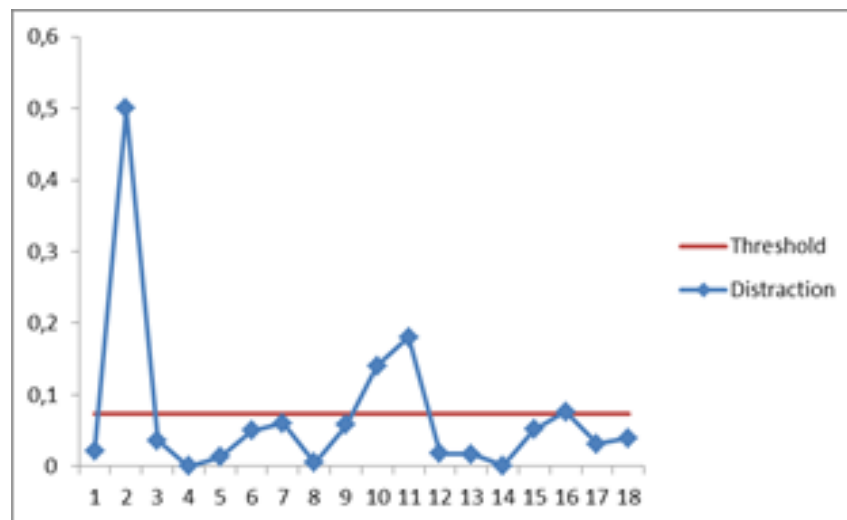


Figure 3. 5 Distraction distribution among learner's

For the distraction (D) indice (fig. 3. 5), we notice that almost all the learners have an indice below the threshold, except the first one that has the highest value of distraction (may be due to a feeling of tiredness during the experiment). That means that almost all the learners are very awake and not sleepy during these cognitive tasks. It also shows that our tasks are interesting for the participants.

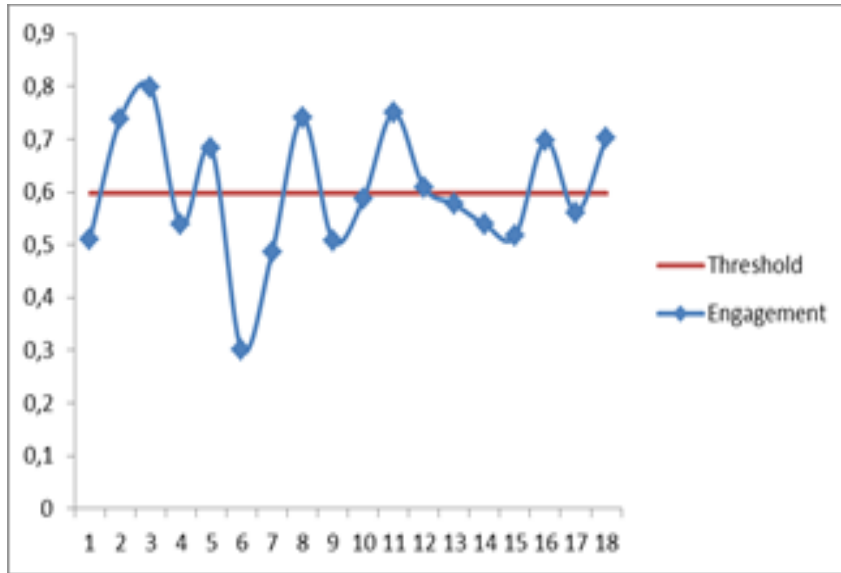


Figure 3. 6 Engagement distribution among learners

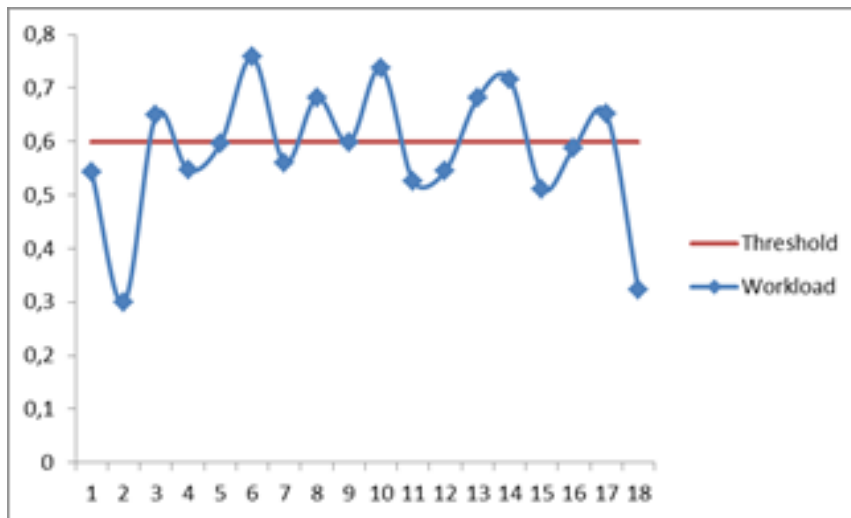


Figure 3. 7 Workload distribution among learners

However, for the engagement (E) and workload (W) indices (see fig. 3. 6 and fig. 3. 7), we notice that these indices vary and fluctuate according to the individual differences and situations. Therefore, some people have a high level of engagement or workload and others have low ones. So, we can say that this variation may depend on each category or type of a task but not on the overall of the cognitive environment.

To verify this hypothesis, we calculate for each task category and each task type the variation of these EEG indices (mean and standard deviation). These results are presented in the following table.

Table 3. 1 Distribution of EEG metrics according to category and type of a task

Category	Mean (SD)			Task Type	Mean (SD)		
	D	E	W		D	E	W
Memory	0.08 (0.17)	0.52 (0.19)	0.65 (0.11)	FDS	0.1 (0.16)	0.5 (0.2)	0.61 (0.16)
				BDS	0.09 (0.18)	0.51 (0.21)	0.67 (0.09)
Concentration	0.08 (0.14)	0.56 (0.18)	0.58 (0.12)	FM	0.09 (0.17)	0.54 (0.02)	0.6 (0.13)
				RT	0.06 (0.07)	0.6 (0.1)	0.6 (0.1)
Reasoning	0.06 (0.13)	0.62 (0.19)	0.64 (0.09)	AA	0.03 (0.04)	0.62 (0.16)	0.62 (0.12)
				OO	0.05 (0.07)	0.62 (0.1)	0.63 (0.06)
				IR	0.08 (0.18)	0.58 (0.23)	0.65 (0.09)

From this table, we can see clearly that the workload indice (W) has almost the highest value comparing to the other EEG metrics (E and D). This means that our cognitive environment requires a minimum effort and level of mental workload to complete the tasks. Moreover, according to a comparison by the category of a task, we can see that reasoning category has the highest values of workload (0.64) and engagement (0.62) and a lower value of a standard deviation, which indicates that the data points tend to be very close to the mean. Conversely, the distraction's values are very small (near to 0) which results in a highest standard deviation for all the categories and types of tasks. Alike, we can notice that for the type of a task, the workload

value has always the highest value except for the Arithmetic Addition (AA) and Rotation (RT) tasks. Moreover, we can see that Intuitive Reasoning (IR) task is the most difficult one because it has the highest value of workload among all types of tasks. These results lead us to accomplish two one way ANOVA tests to study if the Engagement and/or the Workload depend on a category and/or type of cognitive tasks. By realizing a one way ANOVA between the engagement EEG metric and the workload and the three categories of tasks (independent variables), we reported a very significant differences for the engagement : $F(2,401)=10.89$; $p=0.00^{**}$. Three popular post hoc tests were also taking into consideration (Least Significant Level (LSD), Tukey and Scheffe) for multiple mean comparisons and corrections. For all these three post hoc tests and for the engagement metric, we obtained a very significant result by comparing all the categories twice. Whereas for the workload EEG metric, it also depends on a category of a task ($F(2,401)=3.9$; $p=0,02^*$) but the results are not significant when we consider all the three post hoc tests previously mentioned. We performed likewise another ANOVA test which studied the EEG metrics and a type of a task. We find that the engagement depends on a type of a task ($F(6,397)=4.95$; $p=0.00^{**}$) and the same result for the workload ($F(6,397)=2.86$; $p=0.01^*$). However, results for post hoc tests are not significant in the two cases.

After studying the behavior of EEG metrics, we analyze in a next step the GSR metric (Galvanic Skin Response) which gives an indication of emotions valence or a level of stress (Mohammad et Nishida, 2010; Wislow, 2006). The figure below illustrates the variation of this metric accounting the average values for each participant and a threshold computed by considering the average of all the data as we mentioned in the last section.

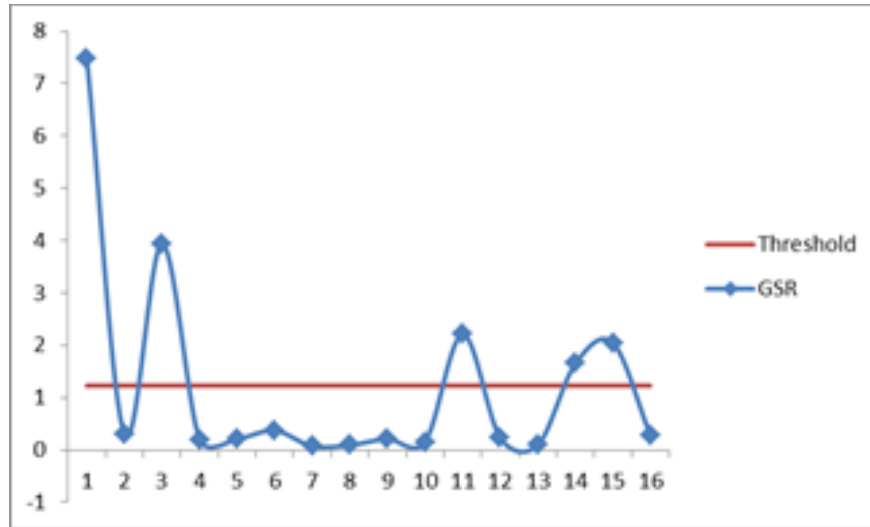


Figure 3.8 GSR distribution among the participants

From this figure, we realize that almost the values are under the threshold which is also very low comparing to other studies (1.22). We can see that we have only five values that we can take into consideration for more statistical analysis, which is a small number of data and we can't apply a technique to change these values according to the data distribution (such as Expectation Maximization technique or Multiple Imputation) due to this small sample of data. This result could be explained by a technical problem occurring during recording the learners' GSR or the individual difference reactions to this metric (Haapalainen et al., 2010; Nourbakhsh et al., 2012). For this reason, we don't focus too much to study the relationship between this metric and the different tasks of our cognitive environment. We think rather that we don't have a reliable dataset for this purpose.

3.5.2 Multiple regression for predicting learners' scores

After studying the electrophysiological metrics and their influence depending on a category and a type of task, we focus next to find a reliable function that connects the electrophysiological variables (EEG indices of workload (W), engagement (E) and distraction (D)) and task information (task duration, task level: ranging from 1 to 6 and task type) in order to predict learners' score.

For this purpose, we performed cross tables that studied the Pearson correlation between the score (the dependent variable) and the others variables (engagement, workload, the

distraction, the type, the duration and the level of the task). Results from this tables show a very low correlation between the score and the distraction ($R=0.036$; $p=0.233>5\%$). That means that there isn't any correlation between these two variables and the distraction logically could be eliminated if we want to predict the score using fewer features. We obtained significant results for the EEG metrics and the task information: type of a task (TT), level of difficulty (DL) and duration of completing the task (D). The following table reports in details the values obtained for the different metrics using mainly Pearson's Correlation.

Table 3. 2 Results of Person correlations

Pearson Correlation	Score	
	R	p
Distraction (D)	0.036	0.233
Engagement (E)	-0.118	0.009**
Workload (W)	-0.122	0.007**
Task Type (TT)	-0.13	0.004**
Difficulty Level (DL)	-0.382	0.000**
Duration (Du)	-0.125	0.006**

Where ** presents a significantly level at 1%, and * at 5%.

This table gives us a brief idea which variables are independent from the score and could be used further to fit a linear regression curve according to these values.

We performed next a linear regression using multiple variables and using a stepwise method to establish the best model that fit our data. Using EEG indices and task information, we obtained the best model with a coefficient of correlation R equal to 0.501 (medium correlation) and a Dubin-Watson value equal to 1.34 which is in the interval of [0, 4] and near to 2. Our model is constructed in descending order by the difficulty level of a task, the task type, the duration of the task and the workload EEG metric (see table 3. 3).

Table 3. 3 Multiple regression model to predict learners' scores

Model Components	Coefficients			
	B	Beta	t	p
Constant	125.89		17.02	0.000**
Difficulty Level (DL)	-8.667	-0.484	-10.4	0.000**
Task Type (TT)	-4.304	-0.293	-6.3	0.000**
Duration (Du)	-0.002	-0,122	-2.81	0.005**
Workload (W)	-23.268	-0,099	-2.27	0.024*

From this table, we can deduce the equation of a predicted score:

$$S = 125.89 - 8.667DL - 4.304TT - 0.002Du - 32,268W \quad (3.1)$$

This equation (3.1) indicates that all our variables have a negative sign indicating an inverse relationship between those independent variables and the dependent one (score).

Finally, in order to study the role of task information, we conducted another linear regression using only the EEG metrics, more precisely the engagement and workload. By eliminating the task information, we noticed that correlation indices R decreased to more than three times (R=0.162). We noticed also that the p-value increases when constructing the model. However, using only this information, the score could be predicted as follows:

$$S = 95.4 - 26.32 W - 15.75 E \quad (3.2)$$

This second equation (3.2) confirms too that the score has a negative relationship with EEG metrics.

3.6 Conclusion

In this paper, we conducted an empirical study of how to include EEG features depending on a category or a type of cognitive task. For GSR feature, we don't have unfortunately correct values from all the participants and the reliable values are not enough to establish a conclusion of how it varies according to a type of cognitive task. We also noticed a very significant result between the engagement and the category of a task. This result is confirmed by applying three post hoc

tests (LSD, Tukey and Scheffe). Moreover, we noticed a strong negative correlation between engagement, workload and task information (difficulty level, task type and duration). The latter leads us to build a linear regression model that fits our dataset, and we successfully obtained an inverse relationship with the score (the variable to predict). As a conclusion, we can moderately predict the score using a multiple regression model that is constructed in descending order of relevance by the difficulty level, the task type, the duration of a task and the EEG workload feature.

Nevertheless, our work has some limitations. For instance, we can improve the prediction of learners' scores by using more complex machine learning techniques (nonlinear models, decision trees, Adaboost, etc.). We also can integrate the best model in real time estimation and regulation of learning content in order to improve the low scores. We can also develop pedagogical strategies in order to deal with learner's performances and the type of a cognitive task.

Acknowledgment

We acknowledge the FQRNT (Font Québécois de Recherche et de Nature et Technologies), SSHRC (Social Science and Human Research Council) through the LEADS project and NSERC (National Science and Engineering Research Council) for funding this research.

Chapitre 4

La classification et la prédiction des performances de l'apprenant

Dans le chapitre précédent, nous avons développé et décrit brièvement l'environnement logique. Nous avons aussi étudié la possibilité de prédire les résultats des apprenants en recourant à des modèles de prédiction linéaires tels que la régression multiple.

Ce chapitre est une extension et continuation des travaux entrepris dans le chapitre précédent. En effet, dans ce chapitre nous présentons et décrivons plus en détail notre environnement purement cognitif logique. Nous avons expliqué aussi la motivation de créer cet environnement qui est inspirée de l'architecture de Bloom (Krathwohl, 2002). Dans ce chapitre, nous proposons d'étendre les méthodes de prédiction des résultats de l'apprenant en utilisant des algorithmes d'apprentissage machine (modèles **linéaires et non linéaires** à la fois). Nous avons alors repris les données collectées lors de notre première expérimentation avec l'environnement logique et nous avons développé en Python des algorithmes d'apprentissage machine en utilisant la bibliothèque *Sickit learn* (Pedregosa et al., 2011). De plus, nous avons réalisé deux tâches essentielles : la classification et la prédiction des scores de l'apprenant. Notre contribution consiste à établir à partir d'un ensemble de caractéristiques collectées (*features*) et un ensemble des algorithmes d'apprentissage machine supervisé une classification et une prédiction des résultats de l'apprenant. Nous rappelons que les données collectées sont les suivantes :

- Engagement issu de l'électroencéphalogramme (variable quantitative continue qui varie entre 0 et 1),

- Charge cognitive issue de l'EEG (variable quantitative continue qui varie entre 0 et 1),
- Distraction issue de EEG (variable quantitative continue qui varie entre 0 et 1),
- Information sur la tâche qui regroupe la durée de la tâche (variable quantitative continue), le niveau de difficulté (de 1 à 6 : variable ordinale), la catégorie de la tâche (3 catégories) et le type de la tâche (7 types). La catégorie et le type de la tâche sont deux variables qualitatives nominales,
- Émotions autoévaluées (7 émotions basiques (Ekman, 1999)) : variables qualitatives nominales,
- Age de l'apprenant : variable quantitative discrète (moyenne est égale à 28 ans),
- Sexe de l'apprenant : variable qualitative nominale (féminin ou masculin).

Ces données sont regroupées dans des vecteurs caractéristiques pour entraîner et tester différents algorithmes d'apprentissage machine. Les questions abordées par cet article sont les suivants :

1. Quels sont les algorithmes d'apprentissage machine les plus adéquats pour la classification et la prédiction des résultats dans les tâches cognitives?
2. Quelles sont les caractéristiques (*features*) les plus importantes qui contribuent à la classification ou prédiction des résultats de l'apprenant?
3. Comment les différences individuelles (principalement l'EEG et l'information sur la tâche) ont une influence sur le résultat de l'apprenant?

La classification et la régression sont étudiées en utilisant quatre algorithmes d'apprentissage machine. Pour la classification, les résultats sont étudiés en termes de précision. Pour la prédiction, nous avons calculé le coefficient de détermination.

Pour répondre aux questions posées, les premiers résultats ont montré que les arbres de forêts (*Random Forest*) sont les meilleurs algorithmes à utiliser tant que pour la classification que pour la prédiction. Nous avons assumé de même que certaines caractéristiques jouent plus d'importance dans la classification ou la prédiction des résultats de l'apprenant. Pour ce faire, nous avons calculé l'indice de Gini (Breiman et al., 1984). En ce qui concerne la classification

selon quatre groupes (allant de très faibles apprenants à de très forts apprenants), les caractéristiques les plus importantes sont : le niveau de difficulté, la durée de la tâche, l'engagement, la distraction et la charge cognitive. Les autres caractéristiques (en particulier le sexe) sont négligeables. Pour la prédiction, deux caractéristiques principales déterminent le résultat de l'apprenant : la difficulté de la tâche et la catégorie de la tâche. Les autres caractéristiques sont très négligeables (le sexe est le dernier à considérer).

Le reste de ce chapitre est constitué de l'article intitulé « *Classification and Regression of Learner's Scores in Logic Environment* » publié dans le journal *International Journal of Information and Education Technology* en 2015. Sébastien Ouellet m'a appris à utiliser la bibliothèque *Sickit Learn* et a contribué dans certains résultats de cet article. Nous rappelons que ma contribution essentielle consiste à utiliser les algorithmes d'apprentissage machine (modèles non linéaires) pour la classification et la prédiction des résultats des apprenants.

Classification and Regression of Learner's Scores in Logic Environment

Ramla Ghali, Sébastien Ouellet, Claude Frasson

Published in: Journal of Education and Training Studies, 3(5), 2015

<http://redfame.com/journal/index.php/jets/article/view/1016>

Abstract:

This paper presents the possibility of classifying and regressing learner's scores according to different cognitive tasks which are grouped with difficulty level, type and category. This environment is namely, Logic environment. It is mainly divided into three main categories: memory, concentration and reasoning. To classify and regress learner's scores according to the category and the type of cognitive task acquired, we trained and tested different machine learning algorithms such as linear regression, support vector machines, random forests and gradient boosting. Primary results shows that a random forest algorithm is the most suitable model for classifying and regressing the learners' scores in cognitive tasks, where the features most important for the model are, in descending order: the task difficulty and the task category in the case of regression, the task difficulty, the time taken by the participant before completing it, and his electroencephalogram mental metrics in the case of classification.

Keywords: Cognitive tasks, task information, engagement, workload, distraction, machine-learning algorithms.

4.1 Introduction

In Intelligent Tutoring Systems (ITS) and Massive Open Online Courses (MOOC), recognition of user affective states, cognitive status and performance evolution during a task remain of great importance (Berka et al., 2004; Pope et al., 1995; Prinzel et al., 2000; Ramesh et al., 2014). In fact, several studies have shown that the emotional state in which a learner is placed has an impact on learning the concept (Damasio, 1995; Isen, 2005). This receptivity is subject to several complex parameters: emotions that are the basis of these affective states, the category of the task, the type of the task, the level of the learner, the individual differences (such as intelligence quotient) and his objectives. To detect and assess users' alertness several studies have been undertaken in the field of artificial intelligence, human computer interaction, cognition and neuroscience (Prinzel et al., 2000; Wilson, 2004). These works focus on using electroencephalogram (EEG) to extract more important features and bands. Three fundamental mental metrics are commonly used from EEG, namely, mental engagement, mental workload and distraction. Mental engagement is related to the level of mental vigilance and alertness during the task (high or low states of vigilance). For instance, highly challenging or difficult tasks involve more engagement. Mental workload can also be seen as the mental vigilance and cognitive load in a particular task. It was calculated according to three electroencephalogram (EEG) channels and two ratios extracted from Power Spectral Densities (Berka et al., 2004). However, distraction or drowsiness reflects the feeling of being sleepy and lethargic (Stevens et al., 2007b). It was calculated mainly from Theta band.

These measures intervene when a learner is involved in a task. They can reflect the degree of a learner's concentration during a task that is necessarily depending on different types of factors such as a learner's situation during the task (if he is relaxed or not), his familiarity with the presented task (the level of the learner), the type of task presented, the difficulty of the task, his motivation and emotions. All these factors lead the learner to reach a skill level that allows him or not to complete the task and acquire some knowledge. Purely calculating his score or performance on each task can assess this skill level. Despite the effort of many researchers to establish a classification and/or regression of a learner's performance in some cognitive tasks (Galan et Beal, 2012), it seems that it is very difficult to establish an accurate estimation of a learner's performance due to all factors that we previously mentioned.

In the same vein and in order to have an overview of a learner's performance before accomplishing a cognitive task, we have developed Logic environment which contains seven types of cognitive tasks. These tasks are grouped into three different categories (memory, concentration and reasoning) and explored different machine learning (ML) algorithms in order to select the most suitable model for the possibility of classifying the learners into different groups according to their levels or numerically predicting directly a learner's score using different regression models. In this paper, we will present a new approach, which is based mainly on gathering different types of data (EEG, task duration, task difficulty, emotion, task category, etc.) from a cognitive environment (Logic environment) and using them to train different machine learning algorithms. In this paper, our purpose is to answer the following questions: (1) Which are the most suitable machine learning algorithms that could be used to classify and/ or regress learner's scores before completing a cognitive task?, (2) Which are the most important features that contributes significantly in classifying or regressing learners' scores? And (3) How individual difference (more precisely EEG metrics) and task information (duration, difficulty, type, etc.) have an influence in learners' scores?

In the following section, we present related works to our study, our methodological approach, and a discussion of our results. An experimental study was also conducted to gather data for training and testing ML algorithms. Finally, we present the results obtained from these algorithms and give some suggestions for future use according to different features used in this study (EEG mental metrics, task information and self-reported emotions, gender and age).

4.2 Related Works

To date several studies were conducted to detect, assess and predict some learner's states in ITS or MOOCs. Among these states, we quote learner's emotions, behaviours, performance, etc. For instance, to assess a learner's emotions, researchers use many methods, which are divided into five main categories: (1) self-judgment where the learner is asked to report his current emotion each period of time. It uses a questionnaire based on a Likert scale, (2) peer-judgment is to record a video of a learner (his face) and then give it to another learner to judge emotions from this video once every slice of time (e.g., a slice of 5 seconds), (3) trained-judgment where humans (teachers or pedagogue) are trained to manually codify learners' emotions from a

recorded video of their face (using the motions of different components of the face and the theory of Actions Units defined by Ekman (Ekman, 1999)). It detects mainly seven types of emotions: disgust, contempt, sadness, happiness, neutrality, surprise and anger, (4) automatic detection by using software that implements algorithms for detecting and tracking the learner's face components such as FaceReader, Facet and Affdex and (5) physiological sensors that can assess posture, Heart Rate (HR), Galvanic Skin Response (GSR), temperature and respiration (Derbali et Frasson, 2011).

To detect and predict a learner's behaviours such as engagement or disengagement on a task, Baker and his colleagues (2013) combined log files; data mining and machine learning techniques in order to automatically establish a classification of a learner's behaviours of disengagement (Baker et Rossi, 2013). Four categories are reported: (1) off-task behaviour that is defined as following: "a behaviour that does not involve interaction with a system or performing a learning task such as conversations or inactivity" (Baker, 2007), (2) Gaming the system has been defined as "exploiting proprieties of the system rather than learning the material and trying to use that knowledge to answer correctly" (Baker, 2007), (3) carelessness has been defined as "an error made on a task that a student already knows (Clements, 1982) or as impulsive and/or hurried actions (Maydeu-Olivares et D'Zurilla, 1995)" and (4) Without Thinking Behaviour is a "behaviour that seems unrelated to a student's learning task" (Rowe et al., 2009).

In addition, some researchers used EEG analysis and feature selection to detect and classify whether the student is engaged or not in a task. For example, Berka (Berka et al., 2004) explored the feasibility of monitoring EEG indices of engagement and workload acquired and measured during the completion of cognitive tests. Results showed an increase in engagement and workload during the encoding period of verbal and image learning and memory tests compared with the recognition period. They also showed that the workload index increased linearly with the level of difficulty. Pope (Pope et al., 1995) has developed an EEG engagement index based on brainwave band power spectral densities and applied in a closed-loop system to modulate task allocation. Performance improvement was reported using this engagement index for task allocation mode (manual or automated). In (Chaouachi et Frasson, 2012) it was experimentally demonstrated that learners' mental workload and engagement are closely related

to specific emotions with regards to different learning phases. A high level of workload indicated the elicitation of negative emotions whereas engagement level was mainly associated with positive emotions. Galan and Beal (Galan et Beal, 2012) evaluated positively the use of EEG for estimating attention and cognitive load during mathematical problems. They could be used for predicting their success or failure by a combination of engagement and workload measures with a support vector machine model. Finally, Jraidi and Frasson (Jraidi et Frasson, 2010) used three sensors (EEG, HR and GSR) in order to evaluate the effect of using a subliminal method to enhance learners' self-esteem during completing three different logic tasks. Their tasks are similar to our categories of reasoning tasks. But, to our knowledge, they did not use machine-learning techniques to classify or regress learners' performance in these tasks.

Some other studies are focused on classifying and/or regressing a learner's performance. For instance, Ramesh (Ramesh et al., 2014) constructs a probabilistic model that connects student's behaviours based on engagement and class performance. Instructors to initiate interventions and assist students could use his model. Recently, the US Army Research Laboratory (ARL) evaluated and developed concepts for two types of learner models: short-term learners and long-term learners (Sottolare, 2014). Sottolare (Sottolare, 2014) also studied the feasibility of using recorded learners' performances in the development of adaptive tutoring learning chains to design future learner models. Naser (Naser, 2012) used knowledge extracted from learner models to determine the academic performance of a learner in order to offer him the proper difficulty level of problems to solve.

Unlike other works, we will show in the next sections that it is possible to use different types of data for real time classification and/or regression of a learner's performance. As features, we will use mainly three indices extracted from EEG and calculated from Power Spectral Densities by B-Alert Software (workload, distraction and engagement), nominal data indicating different types of cognitive tasks (difficulty, category and duration of a task), self-reported emotions, and participant attributes (age and sex). Our evaluation relies on comparing different machine learning models in terms of mean accuracies for classification, and their coefficient of determination in the case of regression. Our main goal is to find the best algorithm

that fit our data collected from subjects for these types of cognitive tasks. We aim also to extract the most relevant features that contribute to learners' performance improvement.

4.3 Justification and description of logic environment

In order to evaluate the importance of introducing the category, the type and the difficulty of a task for the classification or prediction of the learner's performance, we developed Logic environment which contains seven cognitive tasks. These tasks are inspired from Bloom's revised taxonomy classification of cognitive tasks. In fact, Bloom (Krathwohl, 2002) has created a hierarchical taxonomy that classifies cognitive tasks into six main categories: (1) remembering which consists of recall or retrieve previous information, (2) understanding which deals with comprehending the meaning, translate and interpolate objects, (3) applying which consists of using a concept in a new situation by computing or predicting information, (4) analyzing which requires analyzing, comparing, inferring and selecting material, (5) evaluating which consists of making judgments about the existing material by using a self-report questionnaire and finally (6) creating which deals with creating a new model from the previous material. This latter is not addressed in this work. An organization chart summarizes the taxonomy (figure 4. 1). We presented in green color the aspects that we integrated in our cognitive tasks, orange the aspects that are a little bit addressed and red one the aspects not reached here.

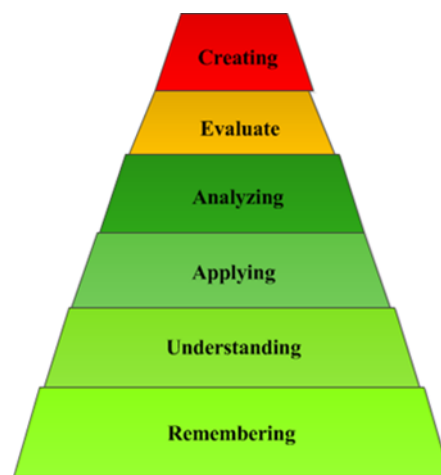


Figure 4. 1 The six dimensions of Bloom's taxonomy of cognitive tasks

In the following, we will describe in detail our Logic environment composed of three categories of cognitive tasks. At the end, we will reproduce Bloom's diagram by adding the corresponding

cognitive tasks in the appropriate cases of this diagram. As we mentioned before, Logic environment (Ghali, Ouellet, et al., 2016b) is divided into three categories: memory (the ability of retain and recall information), concentration (the ability to identify, classify and recognize similar object) and reasoning (the capacity of calculate, analyze, identify the mismatching images, explain and complete some logical series). These tasks are extracted from Cambridge Brain Sciences (CBS).

4.3.1 Memory

This category is mainly based on the famous task of Digit Span (DS). In this task, the learner is shown a series of numbers according to an ascending difficulty level and asked to remember and type them on the appropriate place. In this category, we developed Forward Digit Span (FDS) where the learner has to type the numbers in the same order that they appeared on the screen. We also developed Backward Digit Span (BDS) where numbers should be entered in reverse. This task had six difficulty levels (ranging from 1 to 6) where the learner has to retain a fixed number of digits according to his capacity. The number of sets for each level was fixed in advance. For instance, For FDS task level 1 corresponds to remember 20 sets of 2 digits. Level 2 is to retain 18 sets of three numbers, level 3 is to retain 12 sets of 5 digits, level 4 corresponds to remember 10 sets of 7 digits, etc. In total, learners have to retain 60 sets of varying length of digits where the difficulty level is gradually incremented and the learner cannot move to a higher level if he did not finish the series of lower levels.

4.3.2 Concentration

This category has two different tasks: (1) Feature Match and (2) Rotation. Feature Match (FM) consists in identifying whether the two images appearing on the screen are identical or not according to their forms, numbers and colors. It has also six difficulty levels (ranging from 1 to 6) that vary in their geometrical number and forms. For example, level 1 consists in identifying the identity of 20 sets of images containing each two forms; level 2 is to identify 20 sets of 4 forms on each image, etc. At total, learners have to solve 80 sets. Whereas, Rotation task (RT) consists of identifying, if two images are identical or not and if any rotation was allowed (see figure 1). It has five difficulty levels. For instance, in level 1, the images (12 sets at total) are

constructed by 2 different forms or colors; in level 2, the images (8 sets at total) contain five different forms or colors; in level 3 (8 sets at total), the images have 8 forms with two different colors (see figure 4.2), etc. At total, learners have to identify 44 sets.

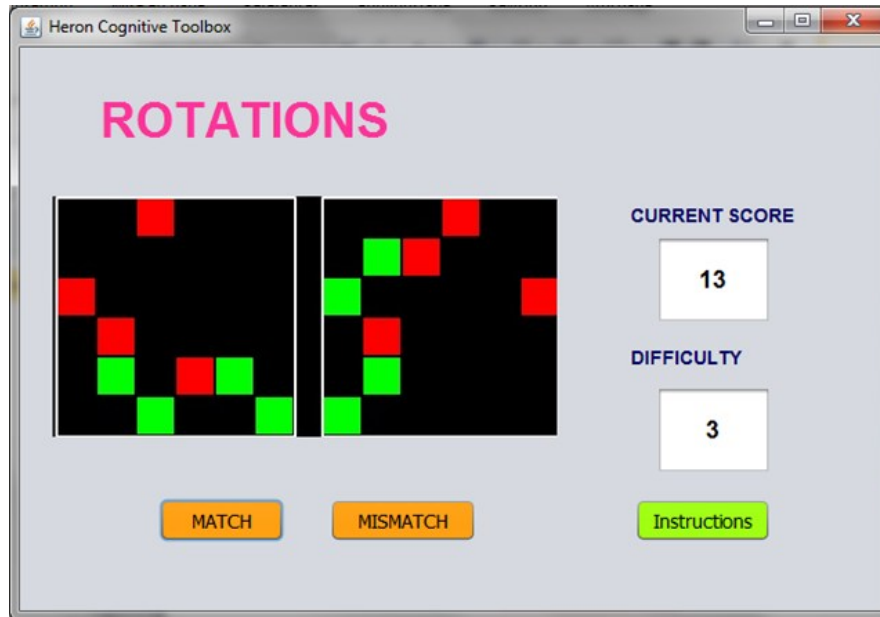


Figure 4. 2 Example of Rotation task

4.3.3 Reasoning

This category has three different tasks: (1) Arithmetic Addition, (2) Odd One Out and (3) Intuitive Reasoning. In the Arithmetic Addition task (AA), we ask the learner to add two numbers, which are varying according to the numbers of digits to add. For instance, level 1 consists of adding twelve sets of two numbers with 2 digits; level 2 consists of adding ten sets of 3 digits' numbers, etc. Following this rule, three main difficulty levels are implemented in ascending order by changing the number of digits each time. In Odd One Out (OO), the learner has to identify the odd image from a series of 9 images appearing on the screen and has four difficulty levels. The difficulty varies according to the number of the different forms and their corresponding colors. For instance, for level 1, the odd image has only a color or a form different from the eight other images (20 sets at total); in level 2, the odd image is a little bit more difficult to find (18 sets at total) and finally in level 3, to find the odd image, the learner has to calculate and infer some rule (12 sets at total). For instance, in the presented example (figure 4. 3), the

odd one is the third (right side) one at the middle of the image because it does not have a corresponding image with three plus form and pink color.

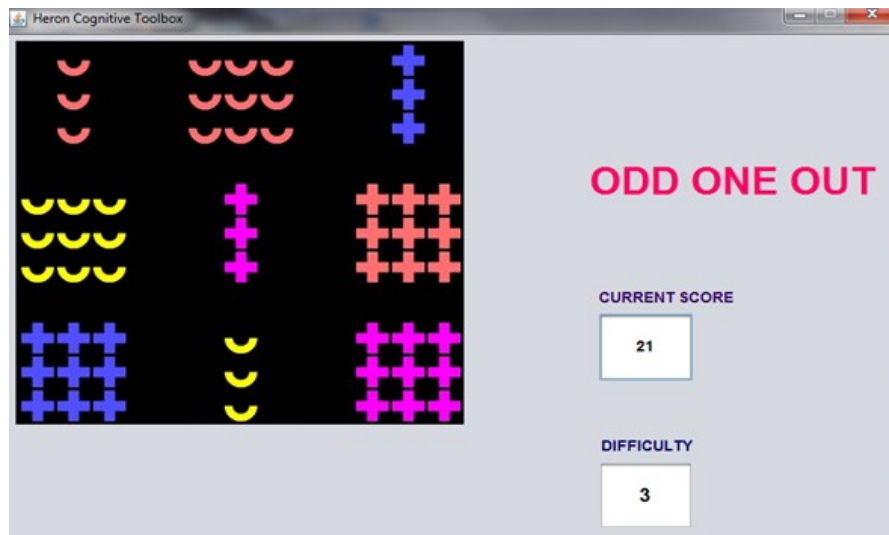


Figure 4. 3 Example of OO task: the odd image is number 6 counting from the left side of the top image

Finally, for Intuitive Reasoning (IR), it has three levels of difficulty (varying according to a time constraint: unlimited, 1 minute and 30 seconds) and 15 series in total; every level contains 5 series of exercises. Unlike other tasks, this task is based on intuitive or analogical reasoning (figure 4) where the learner has to read carefully the instructions, figure out the rule used in the presented example and apply it to resolve another example by selecting the most appropriate choice.

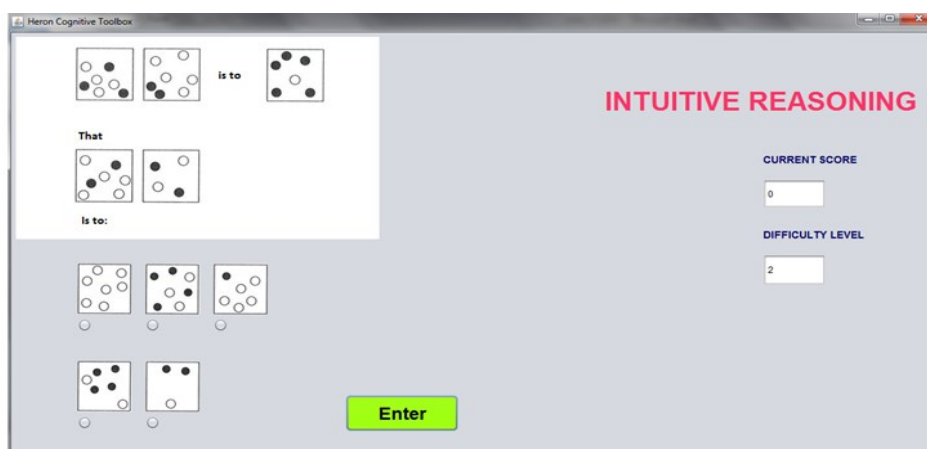


Figure 4. 4 Example of Intuitive Reasoning task

After each type of task, we asked the learner to report his most evident emotion during the task. He had to choose from a list of seven emotions (disgust, contempt, sadness, happiness, neutrality, surprise, anger) defined by Ekman (Ekman, 1999). The use of this set of emotions is based on works in education field as all researchers agree that this small list of emotions is sufficient enough to generate the basic emotions in learning process.

According to the revised Bloom's taxonomy, we think that our cognitive tasks could be classified according to his suggestion as follows:

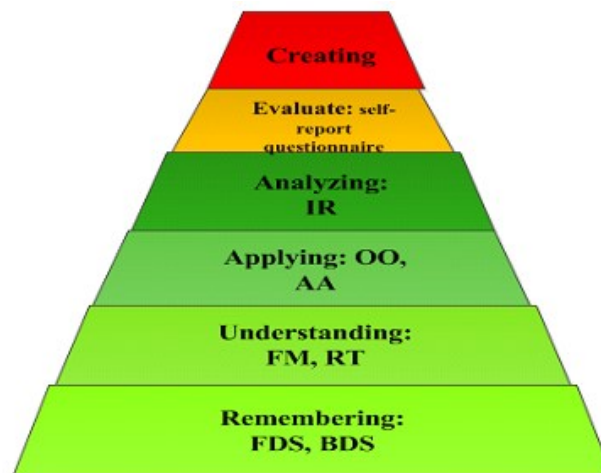


Figure 4. 5 Distribution of the tasks of Logic environment according to Bloom's taxonomy

4.4 Experimental Study

4.4.1 Description of the experiment

In order to test different machine learning algorithms that aim mainly to classify and regress a learner's performance according to different types of data collected from Logic environment described above, we conducted an experiment where the learner is wearing an EEG helmet (The B-Alert X10). In this study, we recruited 20 participants (each received a compensation of 20\$ after the study) from the University of Montreal (9 women and 11 men, mean age=28, standard deviation=4.67). They are from different disciplines and in good mental health. Participants were invited to do the study in our research laboratory, where we explained the objectives of the study and asked them to sign a consent form (an ethics agreement) before starting the experiment. The study lasted about 2 hours, during which they performed the following three

steps: (1) installation of the B-Alert X10 headset on the participant (mean duration \approx 20 minutes), (2) the participant is invited to do 3 baseline tasks defined by the manufacturers of this headset to establish a classification of mental states of engagement, workload and distraction (mean duration \approx 20 minutes) , (3) the participant is finally invited to use our cognitive environment Logic, which includes different sets of cognitive tasks (mean duration \approx 80 minutes). After gathering data from all the participants, the fourth step (4) consists of training and testing different Machine Learning (ML) algorithms in order to early classify or predict learner’s performance. Finally, step (5) allows us to select the most suitable algorithm for the classification and regression of learners’ performances in our cognitive environment, Logic. Figure 4. 6 illustrates the whole process.

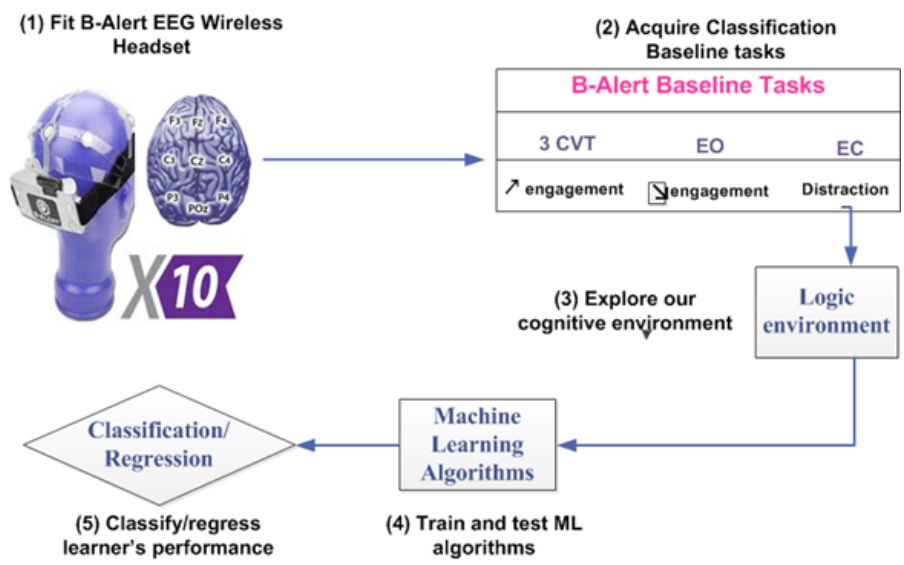


Figure 4. 6 Experimental process and adopted approach

The electroencephalogram (EEG) was recorded by the B-Alert X10 from 9 sensors (F3, Fz, F4, C3,Cz, C4, P3, Poz and P4) integrated into a cap covering the upper half of the head, with 2 reference signals attached to the mastoid bones (behind the ears). The location of each sensor was determined by the international 10-20 system. This cap was equipped with a small USB dongle, which received the wireless transmissions to a PC computer. The EEG was sampled at a rate of 256 Hz, transmitted and converted to Power spectral densities (Alpha, Beta, Theta and Sigma). This process is performed with the B-Alert software that is the proprietary of

Advanced Brain Monitoring Company (ABM). This software allows us to obtain a real time classification of certain mental states (sleep Onset, Distraction, Low engagement, High Engagement and High workload). It also uses some techniques in order to remove noise and artefacts. In this study, from 20 participants only 18 data subjects were considered in the data analysis due to technical problems. Log files were also used in order to synchronize our cognitive tasks with the EEG mental metrics.

4.4.2 Data Pre-processing for Classification and Regression

For each task type, a feature vector was built with the following data sources: sequential data collected by the EEG helmet, a self-report questionnaire inquiring about the emotion felt most strongly for the given task, the gender of the participant, and task information. The EEG features (the Distraction, Workload, and Engagement indices reported by the B-Alert software) were calculated as the means of each index for the duration of each task. Task information included the duration of the task, its difficulty level (from 1 to 6), as well as its category (3 categories) and its type (7 types). The self-report questionnaire asked if the participant felt one of these seven emotions while completing the task: disgust, contempt, sadness, happiness, neutrality, surprise, anger. Four participants omitted to report an emotion for a few tasks, and we filled the missing data with Expectation Maximization (EM) algorithm (Dempster et al., 1977).

Categorical features were transformed into a N number of binary features, where N represented the number of categories, and numerical features were individually scaled by removing the mean and dividing by the standard deviation assuming that the data is Gaussian distributed. The total dimensionality of the samples was then 25, accounting for 3 task categories, 7 task types, 1 number for the task's difficulty, 1 for the time taken by the participant (or duration), 7 self-reported emotions, 3 EEG-derived measures, 1 for the age of the participant, and 2 for the gender of participant.

The target feature was the task score, given as a percentage. For regression, the score was directly used, whereas for classification, the scores were divided in four classes of approximately equal sizes (about 110 samples each): a very strong score above 92% (126 samples), a good score above 80% but below 92% (103 samples), a weak score above 60% but below 80% (123 samples), and a very weak score below 60% (107 samples). The intervals were

selected to produce classes of equal sizes. This is done in order to avoid classification issues related to an unbalanced dataset.

4.5 Algorithms, Results and Discussion

In this section, we provide in the beginning a brief description of the machine learning algorithms used in this work. For more details, the reader could refer to (Hastie et al., 2012). Then, we will describe the results that we obtained from this study.

4.5.1 A brief Description of the Algorithms

a. Support Vector Machine (SVM)

Support vector machine is a generalization of linear decision boundaries where optimal hyperplanes are introduced. The algorithm learns a linear discriminant function by choosing a particular hyperplane that has a maximum margin (the distance between the hyperplane and the closest trained example known as support vectors). It produces then nonlinear boundaries by constructing a linear one in a large and transformed version of feature space. The algorithm can be used with the kernel method, which allows it to change the feature space of the data in a problem to find more adequate representations for the task. A hyper parameter used across all SVMs is the cost parameter C that controls tolerance to errors. Both regression and classification can be done with SVMs.

b. Random Forest (RF)

Random forest is an ensemble of classifiers that uses many decision trees. It combines the bagging idea (use of different models to create a model with a low variance) and the random selection of features to construct different decision trees. It could be used for classification or regression problem. To classify a new point, the algorithm selects the class that has the majority votes from all the decision trees. Whereas, to predict a new point the algorithm calculates the average of the outputs of all the decision trees.

c. Gradient Boosting with Decision Trees (GB)

Gradient boosting is a technique for combining the outputs of different basic models (models could be similar as in our case decision trees or different) called ‘weak learners’. It could be used for classification or regression. The main idea of this algorithm is to add new models to the ensemble sequentially. At each iteration, a new weak learner is trained on data that is weighted with respect to the residual error function of the whole ensemble learnt previously.

d. Linear Regression (LR)

Linear regression is an elementary model used to predict a dependent variable Y by using a linear function of an observable variable X (an independent vector). From the existing examples of training set, it established a link between these both variables (fit a linear straight) in order to estimate the value of Y of observable value of X . The estimation procedure was done by the popular Least Square method, which consists of minimizing the residual (error term). This algorithm is to use only on the case of the regression problems.

4.5.2 Results and discussion

As we mentioned before, four main machine-learning algorithms are reported in this section: support vector machine, random forest, gradient boosting with decision trees (the three highest performance ones), and linear regression (the basic regression model). Tests for other algorithms (k-nearest neighbors, logistic regression, multiple regression, naïve Bayes, etc.) were also done but yielded poor initial accuracies and will therefore not be discussed further. All algorithms discussed in this section are implemented through the python library Scikit-learn (Pedregosa et al., 2011). The evaluation was done with a “leave-one-participant-out” scheme, where 17 participants were used to train the model, which was then tested on the remaining participant. The scores reported in Table 1 below are therefore the mean and the standard deviation of 18 tests. Classification scores are reported in term of accuracy which is defined as the number of correctly classified samples divided by the number of total samples, while the regression models were scored according to the coefficient of determination defined as:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (4.1)$$

Where SS_{res} is the residual sum of squares and SS_{tot} is the total sum of squares. R^2 indicates how well the outcomes data fit the statistical model, with 1 being a perfect fit.

Table 4. 1 Classification and regression for different models and features

Algorithm	Regression M(SD)	Classification M(SD)	Regression without TI* M(SD)	Classification without TI* M(SD)
Random Forest	0.444(0.26)	0.528(0.1)	0.197(0.32)	0.427(0.07)
SupportVector Machine	0.350(0.30)	0.471(0.1)	0.124(0.26)	0.343(0.07)
Gradient Boosting	0.423(0.21)	0.503(0.1)	0.208(0.17)	0.431(0.09)
Linear Regression	0.30(0.21)	x	0.200(0.18)	x

* Where TI indicates task information

Table 4.1 shows clearly that random forest is the most suitable algorithm for both classification and regression tasks when we consider all features (EEG, self-reported emotion, gender, age and task information features). We remember that task information features include the duration of the task, its difficulty level, its category and its type. Taking into consideration all these features, random forest models provide **twice** as good compared to a random baseline for classification (accuracy equal to 0.528 against 0.25). We can see that this algorithm performance is also high if we don't consider the task information for classification (0.427) but becomes very low for regression (coefficient of determination equal to 0.197). We can explain these results by the fact that task information has an important influence when calculating a new case for regression model for random forest algorithm, but it is not too important for the problem of classification. We will confirm this hypothesis by studying further the feature importance for this model (see last section). Moreover, we can notice that gradient boosting combined with decision trees performs also as well as random forest (the results are closed to random forest in the two cases (regression or classification)). It also allows a higher value of classification without including task information (0.431) and a little higher value for prediction (0.208). Linear regression has whereas a lower performance (coefficient of determination of 0.3 and 0.2)

compared to random forest and gradient boosting. We conclude that our first choice for regression and classification in this study would be random forest models followed by gradient boosting models.

For this study, grid searches were also done in order to find good values for the hyper-parameters (mainly for the number of trees as well as their depth for both RFs and GB, and for the C value for the SVMs). SVM models used a radial-basis function kernel, as other kernels were also experimented with but did not perform as well. SVMs performed best with a C between 0.1 and 0.5, Random Forests with about 100 estimators and a maximum depth of 6 to 11, and Gradient Boosting with 50 estimators and a maximum depth of 2 to 5.

In the case of classification into four main classes (very strong scores, strong scores, weak scores and very weak scores), we also calculated a confusion matrix for the random forest algorithm (the best algorithm fitting our data), where true labels are on the vertical axis and predicted labels on the horizontal axis (see Table 2). From this table, we can notice that very weak scores are easiest to classify (accuracy equal to 0.794). However, strong and very strong scores are more difficult to classify (0.563 and 0.429 respectively).

Table 4. 2 Confusion matrix for a Random Forest classifier

True\Predicted	Very strong scores	Strong scores	Weak scores	Very weak scores
Very strong scores (>92%)	0.429	0.286	0.056	0.080
Strong scores (>80% and <92%)	0.252	0.563	0.214	0.165
Weak scores (>60% and < 80%)	0.065	0.195	0.309	0.269
Very weak scores (<60%)	0.047	0.103	0.234	0.794

As we mentioned before, in order to study the difference between the accuracy and the coefficient of determination for the Random forests model, we assumed that this result depend mainly on the relative importance of model's features. So, to verify this hypothesis, we calculated feature importance for the two best models (RF and GB). Feature importance is mainly calculated through the Gini importance (i.e. the mean decrease in impurity with the Gini index), as defined by Breiman and Friedman (Breiman et al., 1984), averaged over all trees in the model (equation 4.2).

$$G = \sum_{i=1}^{n_c} p_i (1 - p_i) \quad (4.2)$$

Where G is the Gini Impurity index, n_c is the number of classes in the target variable and p_i is the ratio of this class at the current node. G needs to be calculated for each node of the trees.

The importance for a feature at a node is then calculated with the following equation:

$$I = G_{parent} - \sum_{i=1}^n G_i \quad (4.3)$$

Where I is the importance for a node, G_{parent} is the Gini Impurity index of the current node, G_i are the Gini Impurity indices for the children of the current node and n is the number of children nodes. The importance of a feature is then the average of I across all nodes that split (i.e. make a decision) based on the feature, and the value of I at each node is weighted by the number of samples that reach that node during training. Intuitively, a feature that is used to make a decision earlier in a tree is more important.

The most important features for classification are (in order of decreasing importance): the difficulty level of the task, the duration of the task, and the indices for engagement, distraction, and workload, as shown in Figure 7. This result could be explained by the fact that not only the difficulty and the type of a task have an impact on learner's scores but also his mental states of engagement and workload during the tasks. These mental states are eventually extracted from the EEG bands (more precisely alpha and beta bands calculated by the B-alert software and that reflect more attention and mental alertness). These metrics are mainly used for calculating the engagement and workload indices, whereas the theta band was more used to reflect distraction. Therefore, we can deduce that EEG is a very important measure to introduce for classifying learners' performance in cognitive tasks in real time applications. We can also see that task type, self-reported emotion, age and task category are about of the similar importance. However the gender is the least one. As a conclusion, for classification of learners we think that their performance depend mainly on the difficulty of the task followed by the time

spent to complete the task (duration) and his EEG mental indices. These EEG metrics reflect in such a way the intellectual differences of individual differences. This result can explain why task information doesn't have much importance in decreasing accuracy in Random Forest model.

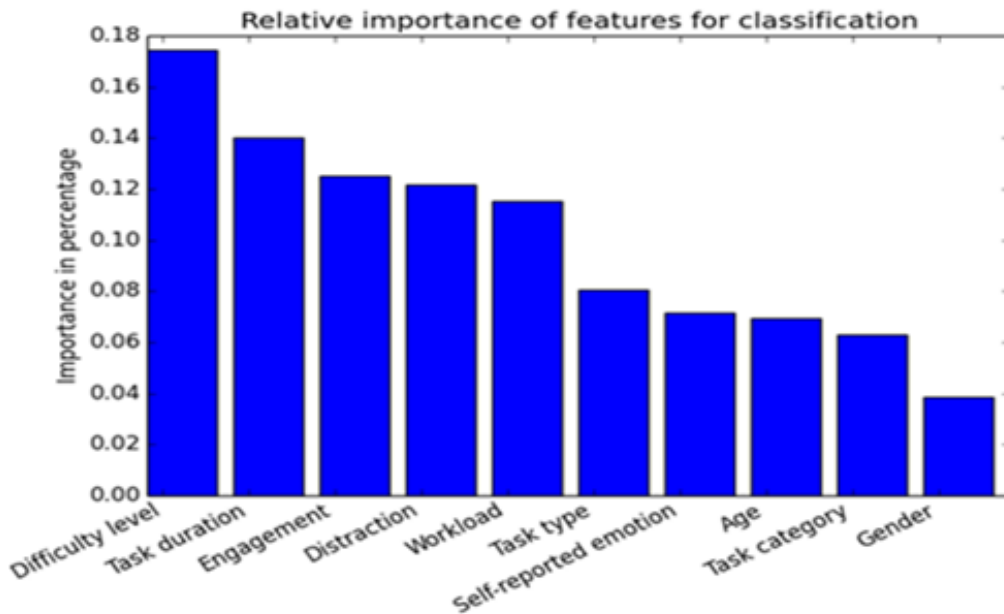


Figure 4. 7 Feature Importance for a Random Forest classifier

Concerning the regression with random forest model, features are considered differently. We noticed more importance is attributed to the task's difficulty level and task category (see figure 8). This result is very surprising for us as we noticed in the inverse of classification, the prediction of learners' scores didn't depend on his intellectual differences (EEG metrics) neither the task duration to complete it. We see also that task information doesn't have a direct influence on learners' scores as we observed in table 1 before. The Gini's index calculation for feature importance showed only that two attributes from task information have higher values (the difficulty level and the task category) but the other attributes (type of task and duration) are insignificant. As a conclusion, we retained that for regression cognitive tasks using random forest models, two significant attributes contributes significantly in calculating a new predicted score in this study: the difficulty level and the task category of the cognitive task. This is could

be explained by the small number of our sample (only 18 subjects) which is the major limitation of this study.

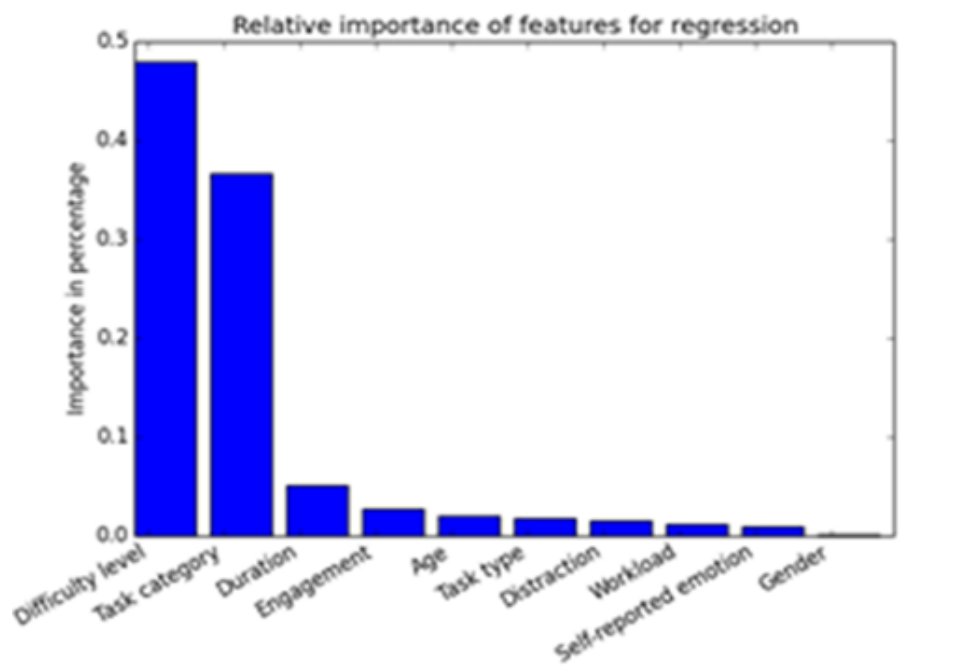


Figure 4. 8 Feature importance for a Random Forest regressor

Finally, we reported that feature importance are not discussed for a Gradient Boosting model as it does not fit the data as well as Random Forest model, but the ranking of top features is similar. We also mentioned that descriptive results issued from Logic environment are described in detail in (Ghali, Ouellet, et al., 2016b).

4.6 Conclusions

In this paper, we studied the possibility of combining different types of a learner's data (EEG indices, self-reported emotions, task duration, task category, task type, age and gender) in order to build real time statistical models of learners' classification and regression. From this study, we have shown that is possible to classify and regress learners' scores before completing a cognitive task. More precisely, we find that a random forest machine learning algorithm is the most suitable to fit our data with a moderate value both for the classification and regression. Using this algorithm, we obtained a classifier that performs twice as well as the random baseline

(0.25) and a regressor with a coefficient of determination of 0.444. We also calculated the weights of the features in order to find the most important features to take into consideration in both of cases. For classification, the most relevant features are the following: the level of difficulty of a task, the task duration and the EEG indices (engagement, distraction and workload). Concerning regression, the two most relevant features are the difficulty level of a task and the task category. This result encourages us to integrate EEG measures into real time applications because it reflects in somewhere learners' intellectual differences and to combine it with different data extracted from log files (task information, self-reported emotion, age and sex). Moreover, we think that is essential in our case to increase the number of participants in order to get more accurate models with higher values.

Future work will focus first on validating the random forest model for classifying and regressing a larger number of learners. Second, we will include this model in real time integration for our cognitive environment in order to enhance learning and adapt it according to learners' individual differences. We will finally try to integrate an automatic detection of emotions instead of using self-reported questionnaires in order to avoid subjective measures.

Acknowledgements

We acknowledge the FQRNT (Font Québécois de Recherche et de Nature et Technologies), CRSH (Conseil de Recherche en Sciences Humaines, more precisely LEADS project) and NSERC for funding this work.

Chapitre 5

Description et étude générale de l'environnement éducatif 2D LewiSpace

Dans les deux chapitres précédents, nous avons présenté notre première contribution qui consiste à prédire et classer les résultats de l'apprenant dans des environnements purement cognitifs.

Ce chapitre et le suivant reflètent notre deuxième contribution dans les jeux éducatifs permettant de s'adapter et s'ajuster en fonction de nombreux critères collectés sur l'apprenant entre autres le nombre d'échec, le nombre de réussite, la performance globale et le besoin d'aide non sollicitée par l'apprenant.

Dans ce chapitre, nous présentons et décrivons notre deuxième environnement expérimental pour l'apprentissage en chimie et plus spécifiquement les diagrammes de Lewis pour des étudiants de collège. Notre environnement est intitulé *LewiSpace* (JS). Dans cet environnement, l'apprenant est simulé par un astronaute qui explore une caverne construite de différents obstacles où il doit à chaque fois accomplir une mission afin de surpasser l'obstacle. Cet environnement exige que l'apprenant construise les diagrammes de Lewis de cinq molécules : H_2O , CH_4 , H_2SO_4 , $\text{C}_2\text{F}_3\text{Cl}$ et $\text{C}_2\text{H}_6\text{O}$. Ces molécules forment les cinq missions du jeu. Cet environnement a été développé par Sébastien Ouellet dans le cadre de sa maîtrise en *Unity 3D* (Ouellet, 2016). Il intègre des mesures en temps réel des états mentaux extraits de l'EEG (l'excitation à court terme, l'excitation à long terme, la méditation, la frustration et l'ennui), le diamètre de la pupille de traceur de regard Tobii TX300 et la mesure de sept expressions faciales (joie, tristesse, surprise, peur, dégoût, colère et neutre). Ces expressions

sont définies par (Ekman, 1970). Cet environnement stimule chez l'apprenant la motivation et l'immersion à travers son aspect 3D et les aspects de recherche des atomes dans différents types d'environnements qui construisent la caverne. L'environnement est doté en tout temps d'un fichier journal où le joueur peut visualiser les instructions à suivre à chaque étape. En particulier, nous soulignons que pour accomplir les différentes missions, nous avons fourni aux participants les différentes instructions et indices facilitant l'accomplissement de chaque tâche. Dans cet environnement, j'ai proposé une étude expérimentale qui sert à améliorer la structure de cet environnement en détectant des faiblesses et en offrant aux joueurs plus d'instructions ou d'aide selon les besoins de chaque utilisateur. C'est pourquoi, j'ai proposé qu'à la fin d'exploration de cet environnement, les participants répondent à un questionnaire sur le besoin d'aide demandé dans chaque mission de *LewiSpace*. Ce questionnaire interroge tout simplement s'ils ont besoin d'aide ou s'ils n'ont pas de tout besoin d'aide. Une échelle de Likert allant de 1 à 3 a été alors utilisée. De même j'ai réalisé différentes statistiques et entraîné des algorithmes d'apprentissage machine, pour la prédiction de besoin d'aide. Ce qui constitue ma contribution principale dans ce travail. En bénéficiant de cet environnement, j'ai réalisé une étude expérimentale où j'ai collecté les données de 40 participants volontaires, de l'Université de Montréal qui n'ont pas de connaissance sur les diagrammes de Lewis. L'expérimentation a duré en moyenne 45 minutes par participant pour un total de **deux mois et une semaine**. Au cours de l'expérimentation, j'ai procédé à l'installation du casque Emotiv EPOC, à la synchronisation de tous les programmes pour la collecte des données (*facereader* pour la reconnaissance des émotions, *Tobii* pour la capture des données de regard et les états mentaux) et au lancement du jeu. Le jeu est précédé et terminé aussi par un pré-test et par un post-test. Nous avons aussi collecté les données sur la personnalité en utilisant le questionnaire Big Five (Barrick et Mount, 1991). Nous avons étudié l'amélioration des performances dans le jeu par rapport au pré-test. De même, nous avons construit des vecteurs caractéristiques (15 caractéristiques en total) :

- 5 états mentaux (variables quantitatives continues dont les valeurs entre 0 et 1),
- Diamètre de la pupille (variable quantitative continue qui varie entre 0,75 et 4 millimètres),
- Valence de l'émotion (variable quantitative continue qui varie entre -1 et 1),

- Activation de l'émotion (variable quantitative continue qui varie entre -1 et 1),
- 7 émotions basiques (variables quantitatives continues dont les valeurs varient entre -1 et 1).

Ces caractéristiques sont étiquetées selon l'échec ou la réussite pour chaque mission et forment ainsi 633 séquences au total où pour chaque vecteur formé de 15 caractéristiques, nous avons calculé quatre statistiques : la moyenne, l'écart type, le maximum et le minimum. Ce qui donne lieu à des vecteurs de 60 dimensions (15*4).

Les résultats descriptifs de cet environnement montrent que la mission 3 (construire la molécule H₂SO₄) est la mission la plus difficile à accomplir et la plupart des participants décident de quitter à cette mission. Notre environnement encourage les erreurs et ne fournit pas beaucoup d'aide (il fournit seulement les instructions nécessaires à l'accomplissement des missions et ne contient pas des exemples et des solutions à des molécules) à l'apprenant. La mission 1 qui consiste à construire la molécule d'eau est la plus facile à accomplir. De même, notre jeu JS s'est montré efficace pour l'amélioration des performances de l'apprenant. Dans ce jeu, le score (en secondes) est calculé en utilisant la durée de chaque mission et le nombre d'échecs pour chaque mission, Il est calculé comme suit :

$$S = T + \alpha * nF \quad (5.1)$$

où T : durée total pour accomplir la session; α : durée moyenne pour chaque essai pour tous les participants; nF : nombre d'échecs pour compléter la mission par l'apprenant.

Nous avons ensuite normalisé ce score entre 0 et 100 en utilisant cette équation :

$$SN = 100 * \left(1 - \frac{M-S}{M-m}\right) \quad (5.2)$$

où m et M sont respectivement les valeurs minimum et maximum des scores collectés pour tous les participants.

Nous avons aussi entraîné des modèles généralisés qui permettent de prédire le besoin d'aide de l'apprenant en fonction des vecteurs caractéristiques collectés. Nous avons essentiellement testé les machines à vecteurs de support (SVM) avec un noyau non linéaire gaussien radial RBF (*Radial Basis Function*) et les régressions logistiques. Les deux algorithmes ont des performances presque similaires. Par exemple, les régressions logistiques nous ont

permis d'obtenir une précision de **56.4%** (qui est supérieur à un random baseline de **33.33%**). Nous avons aussi considéré l'importance des caractéristiques et nous avons constaté que les états mentaux sont les caractéristiques les plus importantes à considérer. Nous avons aussi tracé les courbes ROC (*Receiver Operator Curves*). Nous avons constaté que les mesures des participants fluctuent au-dessous et au-dessus de *baseline*, ce qui nous a suggéré pour un futur travail de considérer les modèles individualisés.

Le reste de ce chapitre est constitué de l'article intitulé « *LewiSpace : an Exploratory Study with a Machine Learning Model in Educational Games* » publié dans le journal *International Journal of Information and Education Technology* en 2016.

LewiSpace: an exploratory study with a machine learning model in an educational game

Ramla Ghali, Sébastien Ouellet, Claude Frasson

Published in: Journal of Education and Training Studies, 4(1), 2016

<http://redfame.com/journal/index.php/jets/article/view/1153>

Abstract:

The use of educational games as a tool for providing learners with a playful and educational aspect is widespread. In this paper, we present an educational game that we developed to teach a chemistry lesson, namely drawing a Lewis diagram. Our game is a 3D environment known as LewiSpace and aims at balancing between playful and educational contents in order to increase engagement and motivation while learning. The game contains mainly five different missions aim at constructing Lewis diagram molecules which are organized in an ascending order of difficulty. We also conducted an experiment to gather data about learners' cognitive and emotional states as well as their behaviours through our game by using three types of sensors (electroencephalography, eye tracking, and facial expression recognition with an optical camera) and a self-report personality questionnaire (the Big Five). Primary results show that a machine learning model namely logistic regression, can predict with some success whether the learner will success or fail in each mission of our game, and paves the way for an adaptive version of the game. This latter will challenge or assist learners based on some features extracted from our data. Feature extraction integrated into a machine learning model aims mainly at

providing learners' with a real-time adaptation according to their performance and skills while progressing in our game.

Keywords: educational game, electroencephalogram, eye tracking, facial expression recognition, logistic regression model, Big Five Questionnaire.

5.1 Introduction

In Human Computer Interaction (HCI), the Intelligent Tutoring Systems (ITS) were among the first sophisticated learning systems. These systems are characterized by their capacity to provide a continuous feedback, hints, helps, etc. to learners. The adaptation was done mainly by using intelligence artificial techniques and more specifically machine learning algorithms. The use of the latter provides these systems with the 'intelligence' criterion. Nowadays, the ITS have progressed and moved to another type of environment since 2002: educational games or serious games (SGs). SGs are video games that aim to inform, test and train people while playing. They can be applied in several fields: military, government, education, business, health care, etc. Recently, they are more used for educational reasons (Conati, 2002; Ghali et al., 2014; Jackson et al., 2012; Rowe et al., 2009), through their playful aspect, known as 'Game based Features' (McNamara et Jackson, 2010).

Although this environment presented a very appropriate and quite moderate way of learning, the problem of user adaptation according to educational aspect remains of great importance. We think that researchers should focus more on this problem in this type of environment due to its difference with ITS. However, to our knowledge, only few works are interested to develop some adaptive user models in SGs and very few works are interested to provide learners' with real time adaptation which reacts while interacting with the game. However, we believe that this issue presents a very important factor to consider in SGs since that their main goal in educational applications is to improve and focus more on the educational and pedagogical content and not the playful aspect which does not contribute to learning improvement process. Therefore, on one hand we think that the adaptation must more focus on the type and nature of help to supplement and complement the educational content of the game. On the other hand, we think that this latter should be instantly and in real time. So, the resulting environments (SGs) should react immediately and/or according to user's needs and learning

capacity. Moreover, many works focus more on the playful aspect which increases motivation and pleasure but not necessary contributes to improving the learning process which depends mainly on the pedagogical contents. In addition, almost games present a serious problem when translating from game situations to non-game contexts. Furthermore, the students generally spend too much time in using entertainment and playful aspects rather than practicing educational contents.

In order to solve this problem and to develop more effective real time adaptive tools which consist of taking learners' differences (a Big Five personality was administrated in our work) and focus more of detecting when users need really more pedagogical help in SGs, we proposed in this paper a first version of a 3D puzzle game called **LewiSpace**. We hypothesize that this game will focus more on learning how to draw Lewis diagrams rather than playful features (which are available when navigating in a 3D environment, changing backgrounds' colors with the different places and gathering the requested atoms to construct complex molecules). Our goal with the study described in the current paper is to investigate whether it is possible to predict a learner's success and his desired level of help based on information gathered through different types of data: electroencephalography (The Affectiv Suite from EPOC Emotiv), eye tracking (the indice of workload extracted from pupil diameter), facial expression recognition (through FaceReader software) and self-report big five personality questionnaire. Since this is part of a larger project that aims to develop a game that will be able to adapt in real-time to learners, we first studied in this paper the descriptive results obtained, the importance of each sensor used and how it improves prediction of learner's success or fail in each mission of our game. We also studied the utility of combining different types of data and the necessity of using them to build the most appropriate real-time users' adaptation.

This paper is structured as follows: in the next section, we describe some related works and mention the disadvantages of the existing works. Next, we describe LewiSpace game that is designed to teach how to construct Lewis diagrams for some complex molecules. Next, we describe the experiment that we conducted in order to gather data and their pre-processing stage. Finally, the last section presents some descriptive results about learners' performance distribution in the different missions of our game. We also provide a comparison and a discussion about the different real time statistical machine learning techniques used in this study,

how we extract the features from our multimodal kinds of data and how we select the best approach, more specifically concerning the real time machine learning algorithm, the features and the hyper parameters to take into consideration for this type of applications.

5.2 Previous Work

Recently, the use of educational games or serious games became widespread. These games are beneficial for learning because they incorporate two fundamental aspects: (1) educational aspect interested to learning content and strategies to present to learners, and (2) playful aspect that allows learners to play, explore, take rewards, control the environment, etc. In fact, researchers believe that this last aspect can increase learners' motivation and engagement (Ghali et al., 2014; Lester et al., 2014; McNamara et Jackson, 2010). This aspect is also known as 'Game-Based Features' (McNamara et Jackson, 2010). Moreover, Prensky, Johnson and Wu (Johnson et Wu, 2008; Prensky, 2007) agree that educational games have not only playful aspects but several criteria and characteristics to increase exploration, immersion and motivation aspects.

According to McNamara and Jackson (McNamara et Jackson, 2010), games based features could be grouped into five main categories: (1) **Feedback** which consists at providing learners with a specific, intelligent and motivational feedback; (2) **Incentives** which aims at promoting the aspects of bonuses and rewards. The latter are related to extrinsic motivation and have a direct effect on learners' self-efficacy, engagement and interest; (3) **Task difficulty** which consist at varying the difficulty of a task and adjust it according to learners' skills; (4) **Control** which allows the learner to monitor and manage the environment such as changing the color of background or avatar and finally (5) **Environment** which focuses at the design and the type of the environment.

Despite the last criteria proposed by (McNamara et Jackson, 2010) to develop more effective educational games, the latter present several problems. Among them, we cite briefly the problem of spending too much time for playing instead of entertainment and learning, the problem of the construction and the order of pedagogical content, the problem of translating between playful and educational aspects, etc. To solve these problems, we suggest that more research will be done in the field of SGs and that this latter should be more intelligent. The intelligence criterion consists of offering to user a real time, continuous, and individualized

adaptation according to learning content. We define this type of game as Intelligent Educational Games (IEG). Whereas, to date, only some works are interested to automatic (but not real time) user modeling and/or adaptation either in tutoring systems or educational games (D'Mello et al., 2012; Gobert et al., 2015; Lester et al., 2014; McQuiggan et Lester, 2006). The adaptation or modeling is usually done using some learners' criteria (such as emotions, engagement, motivation, workload, self-efficacy, performance, etc.). They could also be classified into two kinds of groups: (1) works based on the learner's interactions with the system and (2) works based on the electro-physiological sensors. Among these works, we cite as an example those of Gobert, Baker and his team which are interested to automatically detect learner's disengagement (Gobert et al., 2015). They build an automatic machine learning detector of disengagement behavior. Their model is based on human labels of behaviors from log files and data mining techniques. Lester and colleagues (Lester et al., 2014) used Elliot and Pekrun (Elliot et Pekrun, 2007) model to automatically predict and adapt learners' emotions. This model has been empirically used with learners' interaction data with the system which are derived from a subjective method of self-assessment of emotions. Emotions are recorded from learners using a portable device (smartphone game device) every seven seconds. D'Mello and colleagues (D'Mello et al., 2012) have used eye tracking data to automatically detect emotions of boredom and disengagement among learners in interactions with a tutoring system. Automatic tracking of eye movements was integrated into a tutor that identify when a learner is bored, looking or zooming on the screen.

Recently, Jaques and colleagues (Jaques et al., 2014) used also gaze data features in order to predict two main emotions: boredom and curiosity. These emotions are predicted from several machine learning and feature selection algorithms collected from students' self-reported emotions in Meta tutor system (Azevedo et al., 2010). They obtained an accuracy of 69% for boredom and 73% for curiosity. Finally, (McQuiggan et Lester, 2006) used decision trees and Bayesian networks to generate predictive models of self-efficacy. They obtained two families of models: (1) **static models** that are based on the demographics of the student from pre-test self-efficacy, and (2) **dynamic models** that combine static data model and physiological data (heart rate and skin conductance). The authors have shown that static models predict self-

efficacy of students with an acceptable accuracy rate (73%). However, the dynamic models allow a prediction of self-efficacy with better accuracy rate (83%).

Although these works present a very important way to automatically detect some learners' negative behaviors or emotions which are not effective for learning, we were not interested in this paper to predict learners' emotions because our designed game LewiSpace is not emotionally engaged but focus more on educational aspect. The game also detects automatically learners' emotions through FaceReader software. We use this sensor to extract seven basic emotions (happy, sad, angry, surprised, scared, disgusted, and neutral) in addition to the valence and arousal of each emotion (Lewinski et al., 2014). In (Chaffar et Frasson, 2006) we have also anticipated learners' emotional response using EEG techniques. We also complete the miss detection of some emotions due mainly to mouth occlusion by using the Affectiv Suite provided by Emotiv EEG sensor (Ghergulescu et Muntean, 2014). Whereas, the usage of gaze data (Tobii Tx300 sensor) more precisely pupil diameter is to measure learner's state of workload (Bartels et Marshall, 2012) while interacting with our educational game. The following section describes LewiSpace, a 3D educational puzzle game.

3.3 LewiSpace: an educational puzzle game

3.3.1 A description of the environment

LewiSpace is an educational game which aims mainly to teach learners how to construct chemical structures of molecules using Lewis diagrams (Ghali et al., 2015). The game is mainly designed to be explored by college students who didn't have any knowledge about how to build Lewis' diagrams (a chemistry lesson). It has an exploratory environment (3D) developed using Unity 4.5 for the design of the game, integrating EEG and eye tracking sensors data using the Emotiv SDK v2.0 LITE and the Tobii SDK 3.0.

In the first user's interaction with the game, the player is simulated as an astronaut exploring a planet's surface and communicating with a non-player character known as Commander Arnold (figure 5.1). The player is told that he fell down into a cavern and that he has to explore the underground where he has each time to overcome obstacles (obstructions, lack of a useful resource, etc.) in order to progress in the game and find his lander, allowing him

to return home. The player starts by exploring the environment which is mainly composed of five types of scenes leading to five different missions to accomplish during the game. The missions are constructed in ascending order of difficulty according to the complexity of molecules' structures and the player can't progress to the next mission before completing the latest one. By exploring our educational game, the player accumulates a certain number of atoms (which are hidden somewhere on the environment) that he adds to his inventory and can use them further in order to construct chemical compounds. The latter are used to unlock paths and move to another stage in the game. In the following, we will describe and present some screenshots for the different missions of our game.

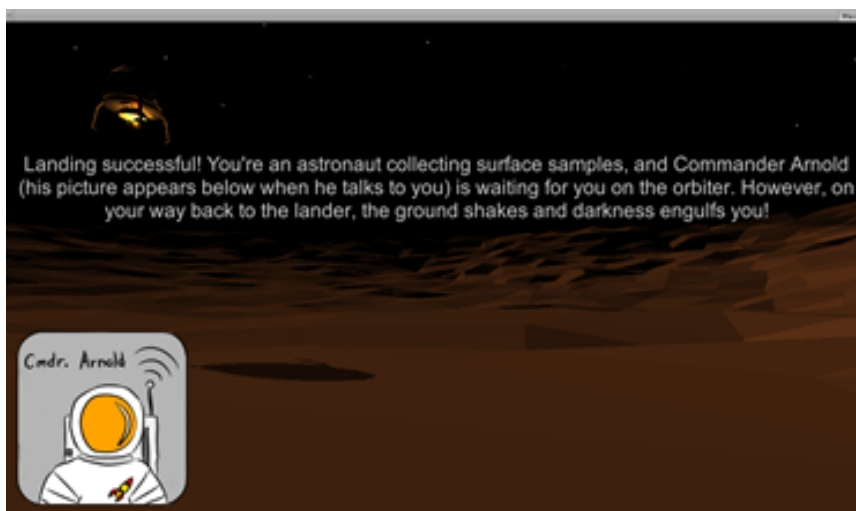


Figure 5. 1 The beginning of the game

a. Mission 1: Produce Water

After starting by exploring the environment and seeing tanks of hydrogen and oxygen, the player is invited to produce H₂O molecule (water) in order to refill his spacesuit's thermal regulation system. He has to place atoms in a grid by clicking on them and deduce the correct Lewis diagram structure (figure 5. 2) according to rules that are presented to him throughout the game (see section 3.2), in the manner of a puzzle.

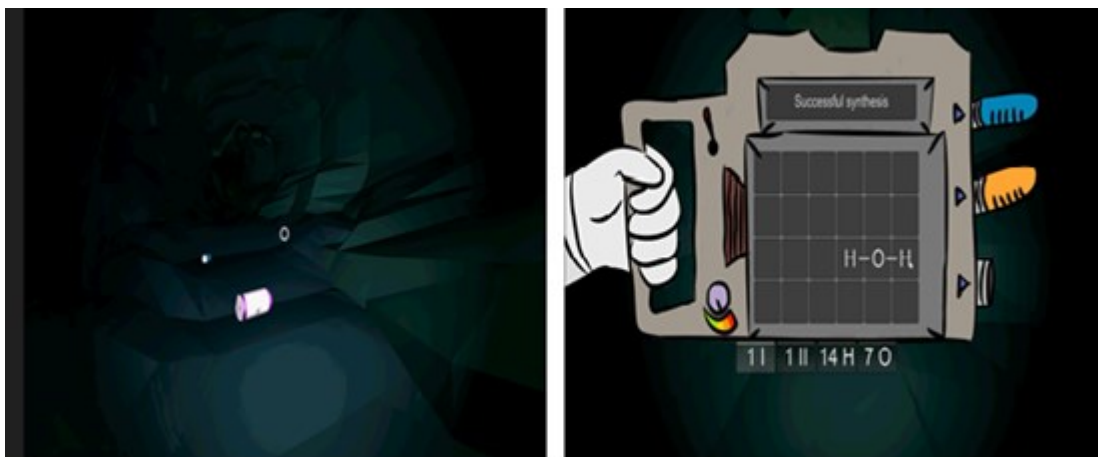


Figure 5. 2 Atoms in the environment (left) and using a Lewis diagram tool to produce H₂O molecule (right)

b. Mission 2: Warm and destruct a tunnel.

In this mission, the player has to gather carbon and hydrogen in order to produce methane gas (CH₄). The player is provided with the information about the group of each atom and also the periodic table that represents the structures of atoms according to their group's number and atomic number of each one during all the missions. He has to find the number of electrons available on the last atom's layer (the valence shell) and to construct the right diagram according to the octet rule and the valence electrons indicated by the periodic table. After producing the methane, an animation is shown to the player, allowing the player to melt the tunnel's obstruction with a methane torch (figure 5. 3).



Figure 5. 3 Melting a tunnel's obstruction using a methane torch

c. Mission 3: Dissolve a metal debris.

The player has to construct a more complex molecule which lets him dissolve metal debris. The compound to produce is sulfuric acid (H_2SO_4). This structure has 32 electrons to distribute between the atoms and the player has to move some electrons in order to construct double bonds. We noticed that until this stage, all the three missions could be presented by a symmetrical diagram Lewis structure.

d. Mission 4: Craft a refrigerant.

Here, the player has to craft a refrigerant ($\text{C}_2\text{F}_3\text{Cl}$) and use it for his spacesuit to regulate his body temperature. This compound could be seen as less complex than the previous one, but it is the first one to present an asymmetrical structure.

e. Mission 5: Fill in the fuel cell with ethanol.

In the last mission, the player is out of the cavern. He finds his lander module on the surface but its fuel cell is empty. As a final task, the player has to gather and construct ethanol ($\text{C}_2\text{H}_6\text{O}$). As soon as this is done, the rocket takes off. At this stage, the game is over (see figure 5. 4).

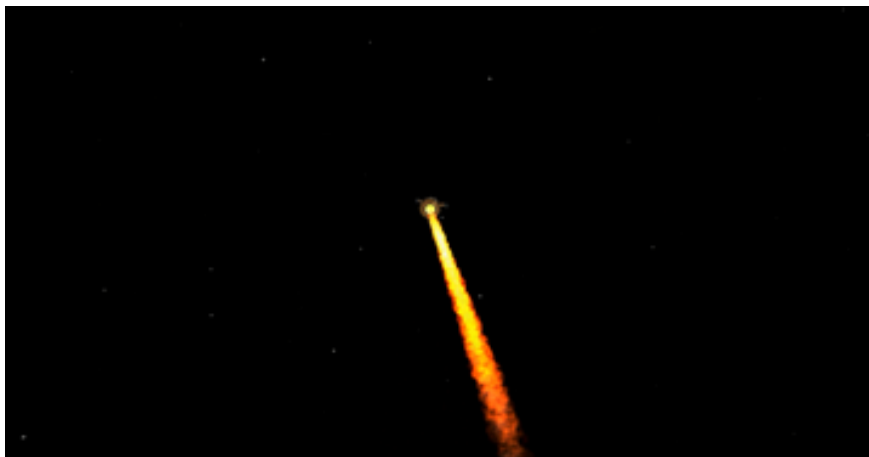


Figure 5. 4 The end of the game

5.3.2 A description of the given instructions and rules

In this section, we will focus more on how we present the educational materials to learners while they progress on the game. Hence, to motivate the students with a playful aspect, we think that this latter is covered when they are navigating on a 3D environment. Every time, they have to look everywhere in the cavern structure to find the appropriate atoms to gather that allow them to construct the requested molecule. The design and the colors of the environments are also attractive and different from each other's. However, the educational aspect is covered by announcing the rules to use on some missions and the directives mentioned in order (see table 1). The learner has also the option to see the periodic table at any time while building a molecule by pressing a shortcut button that enables him to open or close this informative tool.

Table 5. 1 Instructions and rules presented in LewiSpace mentioned according to the missions

Missions	Instructions
Mission 1	<ul style="list-style-type: none"> - Hydrogen atoms can only bond once. This is because a single covalent bond involves one pair of electrons, and hydrogen needs 2 electrons to be full. This is an exception, as other atoms need 8 electrons. This is known as the octet rule, atoms tend to combine to satisfy it.
Mission 2	<ul style="list-style-type: none"> - Double covalent bonds involve 2 pairs of electrons. You can figure out if single or double bonds are needed with the octet rule and with the number of valence electrons the atoms have. - Open your Periodic Table by pressing I. Each column (except the pink-colored ones) group atoms by their number of valence electrons. For example, hydrogen has 1, calcium (Ca) has 2, aluminum (Al) has 3, fluorine (F) has 7. - When crafting a compound, each single bond you add represents 2 electrons, shared between two atoms. If an atom doesn't have 8 electrons after you sum up its lone electrons and those shared through bonds, you might have to add double bonds or to redraw the structure.
Mission 3	<ul style="list-style-type: none"> - It's often important to consider formal charges when drawing structures. You can calculate each atom's formal charge by subtracting each bond (1 for a single, 2 for a double one) and each lone electron from its initial number of valence electrons. - If the formal charge is not zero, you might be able to reconfigure the diagram (change the shape or the type of bonds). The octet rule can be violated in some cases. - Also, keep in mind that elements in the third row of the Periodic Table can sometimes hold more than 8 electrons
Mission 4	<ul style="list-style-type: none"> - No new rules
Mission 5	<ul style="list-style-type: none"> - No new rules

As we mentioned before, the learner is provided at any time he wants by an informative tool, the standard periodic table. This tool describes for each atom the symbol, the atomic number, the mass number or the number of nucleons and the group indicated on the top of each column.

5.4 Experiment and data preprocessing

5.4.1 Experiment

In order to gather data from eye tracking, electroencephalogram (EEG) sensors and learners' emotions, we conducted an experiment where 40 participants (25 males and 15 females aged between 19 and 35 years) from Montreal University participated voluntarily in the study (with a compensation of 20\$ for each participant). As a criterion for admissibility, we requested students who have no prior knowledge about Lewis Diagrams (a chemistry course supposed to be learned to college students). The study was held under strict laboratory conditions. Once we explained the whole process of the study and the participant signed the ethics agreement, the participant is invited to start our experiment. During the experiment, EEG is recorded with the Emotiv headset, which is communicating to the computer through Wi-Fi and only requires a saline solution for conduction. EEG is sampled at a rate of 128 Hz at a second and 14 channels could be measured using this device through TestBench. The headset was also communicating to the Affectiv suite (Ghergulescu et Muntean, 2014), which outputs five high-level features (short-term excitement, long-term excitement, meditation, frustration and boredom). Eye tracking was performed using Tobii Tx300 which is characterized with a high rating sampling frequency of 300Hz per second and its robustness to head movements and light variations. We extract from this sensor the pupil diameter in order to measure learner's workload (Bartels et Marshall, 2012). Facial expression recognition was done using the FaceReader 6.0 software by recording the participant with a webcam. The face tracking process uses the popular Viola and Jones algorithm (Viola et Jones, 2001). This latter allows us to obtain a real time classification of seven basic emotions defined by Ekman (Ekman, 1970): happy, sad, angry, surprised, scared, disgusted, and neutral with their valence and arousal (Lewinski et al., 2014).

The experiment process is composed of 7 steps: (1) installation of Emotiv EPOC headset, (2) calibration of Tobii eye tracker, (3) fill in a personality test (Big Five (Olivier et al. 1999)), (4) a pre-test that consist of constructing 3 molecules (CO_2 , CCl_2F_2 , C_2H_4), (5) exploring our 3D environment and learning Lewis diagram principles, (5) a post-test which is of a similar difficulty of the pre-test (this latter tests if the participant's understood all the instructions presented on the game and could apply them to solve other examples) and (7) finally, evaluation of our environment and self-reported difficulty of each presented mission of our game (easy, medium and hard ranging from 1 to 3) using a questionnaire based on a Likert's scale (see figure 5. 5).

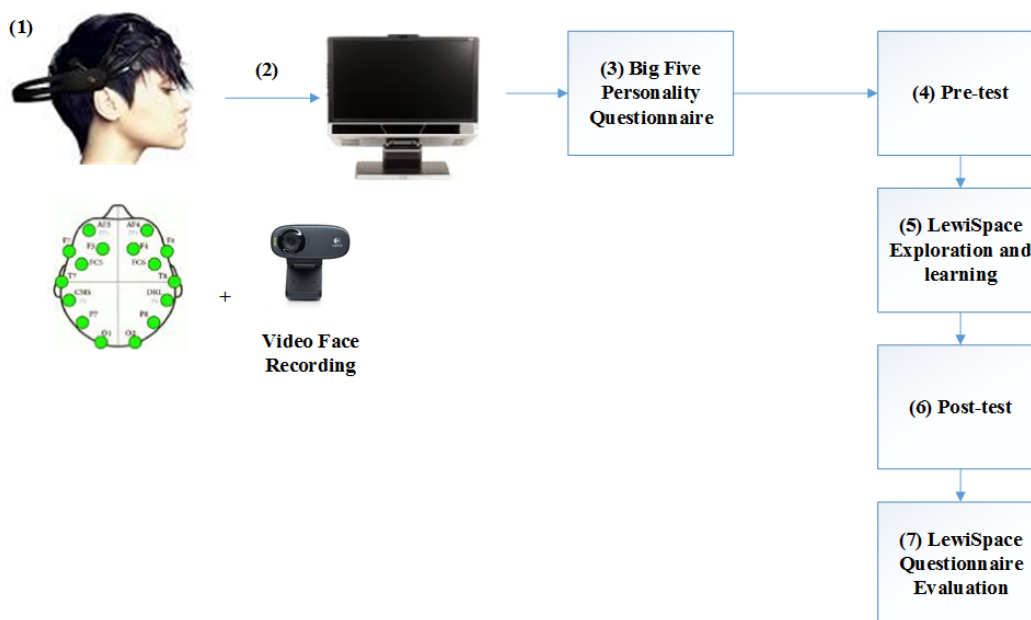


Figure 5. 5 The experimental process

5.4.2 Data pre-processing

Given the data's sequential nature, the data stream was divided in individual sequences according to the learners' trials recorded by the game. For example, the learner could try and fail three times for the first task of the game, with a fourth successful trial. Four sequences would then be available for analysis. Each sequence was then reduced to a feature vector consisting of the 4 metrics, median, standard deviation, maximum, and minimum values for each feature gathered during the game session: short-term excitement, long-term excitement, meditation,

frustration, boredom from the Affectiv suite from Emotiv (on a scale from 0 to 1), pupil diameter from the eye tracking sensors to measure learner's workload, arousal, valence and the seven emotions mentioned above from FaceReader (15 features at total). A total of 633 sequences (across 33 participants) 60-dimensional vectors (15 features multiplied by 4 metrics) were produced. 7 participants were ignored for analysis as technical errors (unrelated to the participants) corrupted data segments essential for a correct synchronization of all data streams. So, only 33 participants out of 40 were taken into consideration for data analysis in the rest of this paper.

Other missing data values (e.g. the eye tracker failing to fit a model to the participant for a small amount of time) were replaced by the mean values for each feature.

Given the difficulty and the nature of the game, which encourages trials and errors, most sequences are labelled as failures. This presents a severely unbalanced dataset, and we used class weighting in order to address this issue. Class weighting inversely penalizes misclassifications according to the frequency of each class, and is implemented in Scikit-learn (Pedregosa et al., 2011), the Python library we used to manipulate data and train machine learning models.

5.5 Results

5.5.1 Descriptive Results

As we mentioned before (experiment section), all the participants completed at the end a subjective evaluation questionnaire which contains questions about the difficulty of each mission of our game as well as the degree of game's appreciation (ranked from 1: not appreciated at all to 5: very appreciated). We noticed then that we obtained a mean of 3.083 (medium appreciated) and a standard deviation (SD) of 1.204. So, we can see clearly that a very big number of participants like the design and the content of our game. However, a small number didn't. This factor could be explained by the value of SD which is a little bit high. We also noticed in general that a small number of participants (20%: 8 from 40) succeed to complete the whole missions of the game because it's a difficult lesson for non-scientific people and needs

more explications and examples (which is our main goal for an improved version of this game according to different types of learners and personalities).

Next, we calculated for each mission some statistics of the average number of failure (see figure 6 below). From this figure, we reported the mean and the standard deviation M(SD) for each mission. For example, mission 1 is ranked the second highest in term of average of failure although it's an easy mission (H₂O):10.04(9.49). Mission 2 has the lowest number of failure (6.15(5.75)) despite it is very similar to mission 1. Mission 3 has the highest number of failure: 14.55(8.55). However, mission 4 and mission 5 are about the same in term of mean of failure. Whereas, it should be noticed that is a large number of participants decided to quit in mission 3 (27 cases) and a small number of them completed mission 4 and more particularly mission 5 (8 cases) which is at the end of our game. Despite these results, we noticed that there are some learners have 0 failures in some missions and some ones have 29 errors or failures (the maximum number of trials which are wrong achieved for our sample of people).

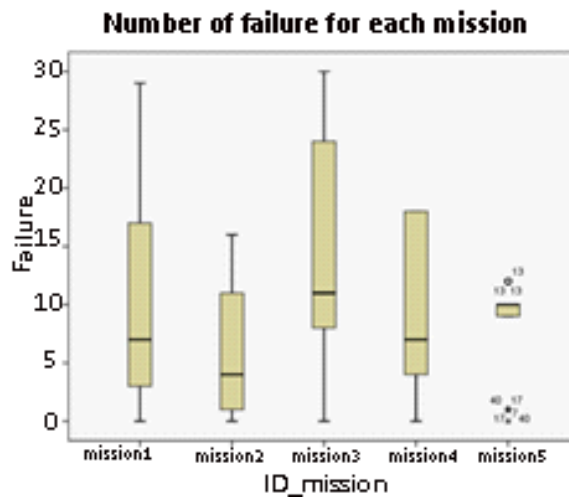


Figure 5. 6 Statistics of failure per mission

Therefore, we calculated in the table below the means and standard errors of the average duration of success in seconds (s) for each mission to study again the difficulty of missions for our sample of students. We noticed from this table that mission 2 is the easiest one (433.80 s) and mission 3 is the hardest one (1433.26 s). However, missions 1 and 4 are about the same in term of difficulty according to the time spent in order to accomplish them. This result is very surprising for us as we assume that the mission 1 is very easy to accomplish.

Table 5. 2 The average success time per mission

Missions	Total Duration
	Mean (Standard Error)
Mission 1	926.68 (69.83)
Mission 2	433.80 (24.37)
Mission 3	1433.26 (67.43)
Mission 4	902.03 (73.96)
Mission 5	626.72 (86.76)

Moreover, we realized an ANOVA to study if the missions are statistically different according to the number of failure and the duration. Results shows that missions depend only on the number of failure and not the duration ($F(4,666)=25.17;p=0.000^{**}<0.01$) but the results of three popular post hoc tests (Scheffe, LSD and Tukey) are not significant. This means that the missions are not totally different according to the failure factor. Finally, a paired samples t-test was conducted to compare the score improvement in the pre-test and post-test. There was a significant difference in the scores for the pre-test ($M=12.28, SD=23.78$) and the post-test ($M=54.83, SD=29.42$) conditions; $t(37)=-9.054, p = 0.000^{**}$. This result proves that our game contributes to improve learners' scores for drawing Lewis structures after learning the lesson.

5.5.2 Selection of the best machine learning model.

Support vector machine (SVM) with a Radial Basis Function (RBF) kernel and logistic regression models were tested with a grid search on values of gamma (for SVMs) and C to produce the highest balanced accuracy with a **leave-one-participant-out** scheme. This scheme was used in order to promote the selection of a model that can generalize well for a new participant from previous participants. Both algorithms performed similarly. For instance with a RBF SVM, the accuracy is about 54.9% with the hyper-parameters ($C=1.0$ and $\text{gamma}=0.05$) using all the features issued from 3 sensors. This value is very near to that of logistic regression (56.4%). In the following, we interest only on communicating results for the best logistic regression models, as the focus of this paper is not the comparison of machine learning

algorithms. Logistic regression model in our case provides the highest accuracy in all cases as we will see in the next section (by taken into consideration all or some features).

5.5.3 Comparison of the importance of each sensor and the Big Five.

After selecting the best ML model that allows us to detect if the user needs help or not, we focus on this section to study the features importance that contributes mainly to prediction. In what follows, we present the difference in accuracies in term of subtracting each time one feature (sensor feature or Big Five questionnaire). Balanced accuracy is determined by the mean of correct classifications for each class while according both classes the same weight. Overall accuracy is the mean number of correct classifications with weighting for the number of samples (therefore giving more weight for the “failure” class), and the mean participant accuracy is the mean number of correct classifications per participant, ignoring whether or not a participant produced more or less samples in the dataset, similarly to the balanced accuracy.

Table 5. 3 Feature selection through classification accuracies

	All features (3 sensors)	Ignores pupil diameter	Ignores Emotiv	Ignores facial expression recognition	Ignores Big Five Questionnaire
Balanced accuracy	0.564	0.564	0.501	0.564	0.584
Overall accuracy	0.603	0.603	0.256	0.603	0.564
Mean participant accuracy	0.593	0.593	0.312	0.593	0.549

Table 5. 3 shows that ignoring the Emotiv has the **highest** impact on performance, whereas the other features do not seem to change the accuracies when ignored. Adding five features from the self-reported Big Five Questionnaire (the values along the five dimensions measured by the questionnaire before starting the game), we note that the balanced accuracy is

highest, but at the cost of the overall accuracy, which means that the classifier predicts more often that a task will be successful but mispredicts more sequences in total. Ideally, the model should be balanced between those two measures of accuracy. A model that measures only the features from the Emotiv headset was therefore tested and produces the best results so far but still very similar to one which uses all features (using 3 sensors and ignoring the big five questionnaire), with a balanced accuracy of 0.570 , an overall accuracy of 0.635 and a mean participant accuracy of 0.609. However, this indicates that other features are not necessary and might even add noise to the dataset. Table 4 shows its confusion matrix for the logistic regression using only Emotiv Affectiv Suite (The most important feature), predicted values are shown vertically, and true values horizontally.

Table 5. 4 Confusion matrix for a logistic regression model with Emotiv headset features

	Failure	Success	Total number
Failure	0.665	0.335	532
Success	0.525	0.475	101
Total number	407	226	633

From table 5. 4, we can see clearly that failure is easiest to predict (with an accuracy of 66.5% which is higher than the random baseline of 50%), whereas, success is more difficult to predict (value of 47.5% which is less than 50%).

As to the relevance of the other features, we can speculate that brightness changes throughout the game (e.g. some scenes or actions producing various lighting effects) have a larger impact on the pupil diameter, a factor unaccounted during the experiment. We could improve this by controlling the brightness in real time. Facial expressions might also be more useful in a game which is more emotionally engaging which is not the goal of our game. LewiSpace focuses more on educational aspect and how to provide the adequate help to learners according the different situations encountered when playing it.

Finally, we present the Receiver Operating Characteristic curves known as ROC curves (figure 7) for each participant compared with the random baseline using our best model, a

logistic regression (C value of 0.1). It is a plot of the true positive rate against the false positive rate for the different possible cutpoints of a diagnostic test. Accuracy for these models was measured by the area under the ROC curve. An area of 1 represents a perfect test, whereas an area of 0.5 represents a worthless test. The figure below illustrates the width of the accuracies encountered due to individual differences between the participants. We noticed a lot of people with a good accuracy (under the curve) however a lot of cases are misclassified. This result suggests that we should more closely investigate individually trained models rather than generalized models.

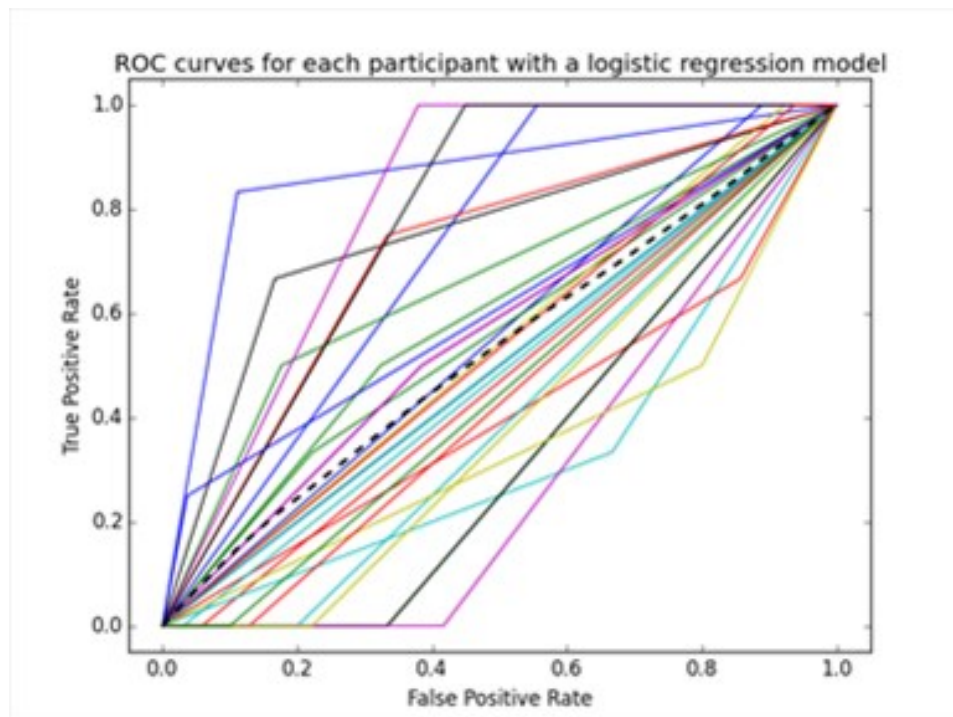


Figure 5. 7 ROC curves for the 33 participants

5.6 Conclusions

In this paper, we presented our educational game aimed at teaching players how to draw Lewis diagrams known as LewiSpace. LewiSpace is a 3D environment integrating EEG and eyetracking SDKs and combining educational and playful aspects. Our study designed also to investigate the use of physiological data (electroencephalography, eye tracking, and facial expression recognition) and personality traits (the Big Five Questionnaire) in order to detect the performance of the learner and his need of help during progressing in our game. We described

also some statistical results in term of failure and duration for each mission. From these results, we can clearly see that the game should be improved by providing more help and examples for students struggling to complete the tasks. This aim will be realized by collecting features from different types of physiological sensors and train a machine learning algorithm. Our finding shows that a logistic regression model using only Emotiv EPOC Affectiv Suite as features is the most suitable for detecting when learner have more difficulty (failure) and needs more help and examples to understand the lesson. We noticed also that personality traits as well as pupil diameter and emotions do not improve the accuracy of our model however they add more noise to our dataset due to the nature of our game.

Future work will involve developing a version of the game that reacts in real-time to the players' physiological data in order to help or challenge them accordingly. Before its development, we will however focus on models that will perform better on our current dataset and on the real-time data gathered in our next experiment.

Acknowledgements

We acknowledge the CRSH (Conseil de Recherche en Sciences Humaines, more precisely LEADS project) and NSERC for funding this work.

Chapitre 6

Une approche physiologique pour la prédiction du besoin d'aide dans l'environnement éducatif LewiSpace

Dans le chapitre précédent, nous avons décrit notre deuxième environnement expérimental *LewiSpace* pour l'apprentissage des diagrammes de Lewis. Nous avons de même proposé d'adapter l'aide aux apprenants en utilisant des algorithmes d'apprentissage machine selon une approche d'adaptation généralisée qui consiste à prédire l'aide pour le participant courant en fonction des résultats des participants précédents.

Ce chapitre est une continuation des travaux entrepris dans le chapitre précédent. Nous étendons la question de prédiction du besoin d'aide en testant et intégrant dans le JS *LewiSpace* deux approches : une approche généralisée telle que décrite dans le chapitre précédent et une approche personnalisée qui prend en considération seulement le participant en cours et non les participants précédents.

Pour ce faire, nous avons proposé une approche physiologique où nous collectons dans la même étude expérimentale décrite dans le chapitre précédent des données sur les états mentaux issus de EEG, le diamètre de la pupille issue de traceur de regard et les données des expressions faciales capturées par une caméra et analysée par le logiciel *FaceReader*. Durant cette approche, nous commençons par collecter les données physiologiques et les données issues d'un questionnaire auto-évaluatif du besoin d'aide (de 1 : pas d'aide nécessaire à 3 : forte aide nécessaire). Ensuite, nous procédons au prétraitement des données en éliminant les données bruitées et en complétant certaines données manquantes. Puis, nous avons développé en Python

des algorithmes d'apprentissage machine (machines à vecteurs de support, les plus proches voisins, régression linéaire, régression logistique, arbres de forêts, etc.) pour tester et trouver le meilleur algorithme qui prédit le besoin d'aide. Nous avons ensuite sélectionné dans chaque modèle les caractéristiques les plus importantes (qui contribuent plus à prédire le besoin d'aide). Finalement, nous avons validé nos modèles afin de sélectionner parmi l'ensemble des algorithmes d'apprentissage machine celui ayant des meilleures performances en termes de précision.

Nous avons alors divisé les tentatives des apprenants en séquences. Par exemple, si l'apprenant a réussi dans le quatrième essai pour une mission donnée dans notre jeu, nous créons 4 séquences ayant pour étiquettes les données de questionnaire du besoin d'aide. Tel que décrit dans le chapitre précédent, nous avons collecté un total pour tous les participants de 633 séquences qui sont composées de vecteurs caractéristiques à 60 dimensions (4 statistiques*15 entrées issues de signaux physiologiques).

Pour sélectionner le meilleur modèle, nous avons considéré deux approches :

- Approche généralisée où nous entraînons les algorithmes d'apprentissage machine sur les anciens participants et nous utilisons ses prédictions pour un nouvel participant,
- Approche personnalisée où nous entraînons et testons les algorithmes d'apprentissage machine sur un seul participant en utilisant deux types de tâches au début de jeu (une tâche facile (qui consiste à construire H₂ molécule) et une tâche difficile (molécule complexe CH₄N₂O₂)). L'accomplissement de ces deux tâches peut fournir des étiquettes sur deux tâches (besoin d'aide ou non) et peut ensuite être utilisée pour calibrer le modèle d'apprentissage machine afin de prédire dans le reste du jeu si l'apprenant a besoin d'aide ou non.

Les résultats de ces deux approches sont comme suit : les machines à vecteurs de support ont tendance à performer mieux pour l'approche généralisée en comparaison avec les précisions des autres algorithmes testés (une précision de **54.1%** qui est supérieure à notre *random baseline* de **33%**) tandis que les plus proches voisins (KNN) a la meilleure performance pour l'approche personnalisée (précision de 67.5%). Les machines à vecteurs de support sont les plus adéquates dans notre cas (en considérant notre échantillon de données) car elles permettent de généraliser

les transformations linéaires des données et sont capables d'accepter un ensemble de données de grande dimension. Leur performance s'est montrée par une précision supérieure aux autres algorithmes d'apprentissage machine testés pour la régression du besoin d'aide. Les caractéristiques les plus importantes restent les états mentaux issus de EEG.

Le reste de ce chapitre est constitué de l'article intitulé « *Towards Real Time Detection of Learners' Need of Help in Serious Games* » publié dans la conférence internationale Flairs 2016.

Towards Real Time Detection of Learners' Need of Help in Serious Games

Ramla Ghali, Sébastien Ouellet, Claude Frasson

Published in: The 29th International [FLAIRS](#) Conference (2016)

<http://www.aaai.org/ocs/index.php/FLAIRS/FLAIRS16/paper/view/12857>

Abstract:

Providing an adequate help to a learner remains a challenge. In this paper we aim to find how to provide learners with real time help in an educational game. Detecting that a player is engaged or motivated is a good sign that he is progressing. For these reasons we need to assess learner's states while learning. In this study we gather a variety of data using three types of sensors (electroencephalography, eye tracking and automatic facial expression recognition) to build a reliable user adaptation system. The data result from an interaction of 40 players with LewiSpace game, that we built for experimental purpose to learn construction of Lewis diagrams. We used machine learning algorithms in order to identify the most important features gathered from each sensor. Two models were trained with these data: a generalized model, trained on all data available, and a personalized model, trained only on the current user during an early phase of the game experience. The predictive results showed that personalized model could outperform the generalized model.

6.1 Introduction

Serious games are environment that teach, train and inform players. Mainly, they combine two fundamental aspects: (1) fun aspect, and (2) educational content. In the last decade, researchers showed that these tools are very beneficial for learning purposes (Jackson et al., 2012; Prensky, 2007). Therefore, we noticed that this learning environment has spread and concerned many population in different kinds of courses: Physics (Shute et al., 2013), data structure (Derbali et Frasson, 2012), and medicine (Lester et al., 2014).

However, a main problem of these tools is that of focusing more on the playful aspect rather than the educational content. For that, detecting when users need more help or challenge, is very important. This task is very delicate and requires that the game detects learners' emotions, engagement and motivation. For instance, emotions could be used to detect if the learner is bored or frustrated and therefore the game should react by changing the learner's state and offer him more encouragement and help if needed. In the other case, if the game detects that learner is very excited it may be an indication that he needs more challenge (no help is required but learner should be in an ascending difficulty level of the game). Moreover, detecting that the player is engaged would be a good sign that he is progressing while playing the game. To summarize, in order to develop more reliable serious games (SG), it would be necessary to detect all the previous mentioned points and react accordingly.

In this paper we are interested in detecting exactly if learners need help in a LewiSpace game. The main objective of this paper is to provide learner with an adequate help when needed in order to avoid gaming behavior. To do that, we conducted an experiment where we collected data from three sensors (Electroencephalogram, Eye tracking, and FaceReader) and integrated them into machine learning algorithms. At the end of the game, the players were submitted to a self-report questionnaire on the need of help, for each mission of the game.

We will adopt two approaches: (1) a **generalized approach** where we train machine learning algorithms **offline on all the data acquired**, then we predict users' need of help **online** on the current participant, and (2) a **personalized approach** where we train several models in **real time** for each user at the beginning of the game session using two types of tasks (an easy

and a difficult task). The prediction will be done also **online** using different descriptive vectors for this participant.

6.2 Related work

In the community of Intelligent Tutoring Systems (ITS) or Serious Games (SG), researchers are always interested in improving these environments in order to offer learners with more adequate content. To do this, they are interested to different kinds of problems, such as enhancing comprehension, improving motivation and engagement, increasing positive emotional states which are more effective for learning, etc. (Baker et Rossi, 2013; Ghali et al., 2015). The main purpose of these works is to create adaptive environments for learning.

For instance, D'Mello and his team (D'Mello et al., 2012) have used *eye tracking* data to dynamically detect emotions of boredom and disengagement. Dynamic tracking of eye movements was integrated into a tutor that identifies when a learner is bored. In the case of student disengagement, the tutor tries to speak and attract the learner's attention. Besides, (Conati, 2002) proposed a probabilistic model to monitor user's emotions and engagement in their educational game, PrimeClimb. Recently, (Ghali et al., 2015) were interested to **unsolicited feedback**: whether or not students **need help** while interacting with an educational game.

Despite the effort of researchers, developing an adaptive tool (ITS or SG) remains a great problem because it depends on different factors, the actual works are mainly based on two approaches: (1) adaptation according learners' behaviors, and (2) adaptation according a physiological sensing approach. For instance, in the first category, we can cite as examples of adaptation: mouse movements, learner's behaviors (off-task and gaming behaviors), time response, learner's emotions, etc. Whereas, in the second category, we can mention as examples the mental states of distraction, workload, and engagement extracted from EEG, learner's gaze data, arousal and valence (of emotions), skin conductivity, etc.

Besides, researchers prove that efficient **help seeking behavior** (students' request of help) can improve learning outcomes and reduce learning duration (Wood et Wood, 1999). However, the abuse of help seeking can reduce learner's performance and reflect more gaming

behavior (Baker et al., 2012). For all these reasons, we focus on the following to detect if a learner really needs help in SG. To do that we studied the possibility of integrating two machine models (Support Vectors Machine (SVM) and K Nearest Neighbors (KNN)) used with a simple feature extraction to detect when it is necessary to provide learner with more help and pedagogical content in our educational game, described in the next section.

6.3 A brief description of Lewispace game

LewiSpace is a game intended to teach Lewis diagrams for college students. For a detailed description, the reader is referred to (Ghali et al., 2015; Ghali, Ouellet, et al., 2016a).

Our game is a puzzle-game designed using Unity 4.5 (a 3D environment) integrating EEG and Eyetracking sensors data using the Emotiv SDK v2.0 LITE and the Tobii SDK 3.0. In this game, the learner appears as an astronaut exploring a planet's surface. The astronaut falls into a cavern and for surviving he has to accomplish five missions elaborated in an ascending order of difficulty (see table 6. 1).

Table 6. 1 Missions' distribution in LewiSpace game

Missions	Molecules to construct
Mission 1	Produce water (H ₂ O)
Mission 2	Produce methane gas (CH ₄)
Mission 3	Produce a sulfuric acid (H ₂ SO ₄)
Mission 4	Craft a refrigerant (C ₂ F ₃ Cl)
Mission 5	Refuel the fuel tank with ethanol (C ₂ H ₆ O)

In each mission, he has to gather atoms randomly distributed in the environment and construct then a molecule's Lewis diagram in order to overcome obstacles, progress in the game and find his lander to return on the earth. The tool for constructing molecules is presented in figure 6.1.

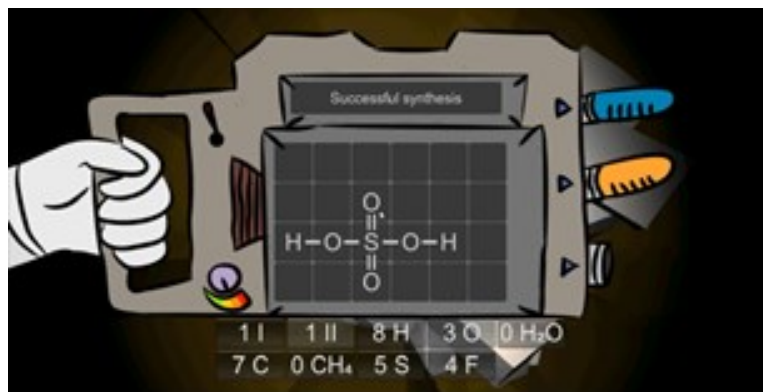


Figure 6. 1 Lewis diagram of the sulfuric acid

6.4 Our approach: physiological sensing

In this section, we describe in detail our proposed approach for realizing real time user's adaptation in our game LewiSpace. The adaptation will be done according to learner's need of help (known as **unsolicited feedback**). In fact, as mentioned before, we use a multimodal approach that combines mainly three types of non-intrusive sensors: **electroencephalography** (EEG) for different bands and indices extraction (more specifically the emotions' classification using the Affectiv Suite provided by Emotiv EPOC (Ghergulescu et Muntean, 2014), **eye tracking** for tracking eyes' motions (more precisely pupil diameter (Bartels et Marshall, 2012) to measure the mental state of workload), and **facial expression recognition** using FaceReader (seven basic emotions defined by Ekman (Ekman, 1970)). Moreover, for users' need of help we used at the end of our game a **self-report questionnaire** that identify if the users need or not more of help in each mission of our game ranging from 1 (no help required) to 3 (more help required).

To build and validate our approach, we conducted an **experiment** and collected data using the above mentioned types of sensors, where **40** students (25 males and 15 females) participated voluntarily in the study and had no prior knowledge about Lewis diagrams. The participants were asked to play LewiSpace while we collected mainly three types of physiological data:

- EEG data with a sampling rate of 128Hz at a second (using Emotiv EPOC),

- pupil diameter to measure workload (Bartels et al. 2012) using eyetracking (Tobii Tx300), and
- 7 basic emotions: happy, sad, angry, surprised, scared, disgusted, and neutral with their valence and arousal (using FaceReader).

The experiment is respectively preceded and finished by a pre and post test to see if the student learnt some concepts from the game. After collecting these data, we merged them and removed the noisy ones. This step (data preprocessing) will be described in detail further (section 5). Then, we developed and trained different statistical machine learning algorithms in order to find the best algorithm that provides the highest accuracy. All the algorithms were developed using the Python library, the **Scikit-learn** (Pedregosa et al., 2011). After training these algorithms, we select from each algorithm the most important features for modelling purposes in order to reduce our feature vectors and take only into consideration those that have an impact on varying the model's accuracy. We validate our models using **leave-one-out method** and **cross-validation**. Finally, after building and validating the statistical machine learning models, a decision stage consists at selecting the best model with the highest accuracy and investigate the most potentially useful features from it (figure 6. 2).

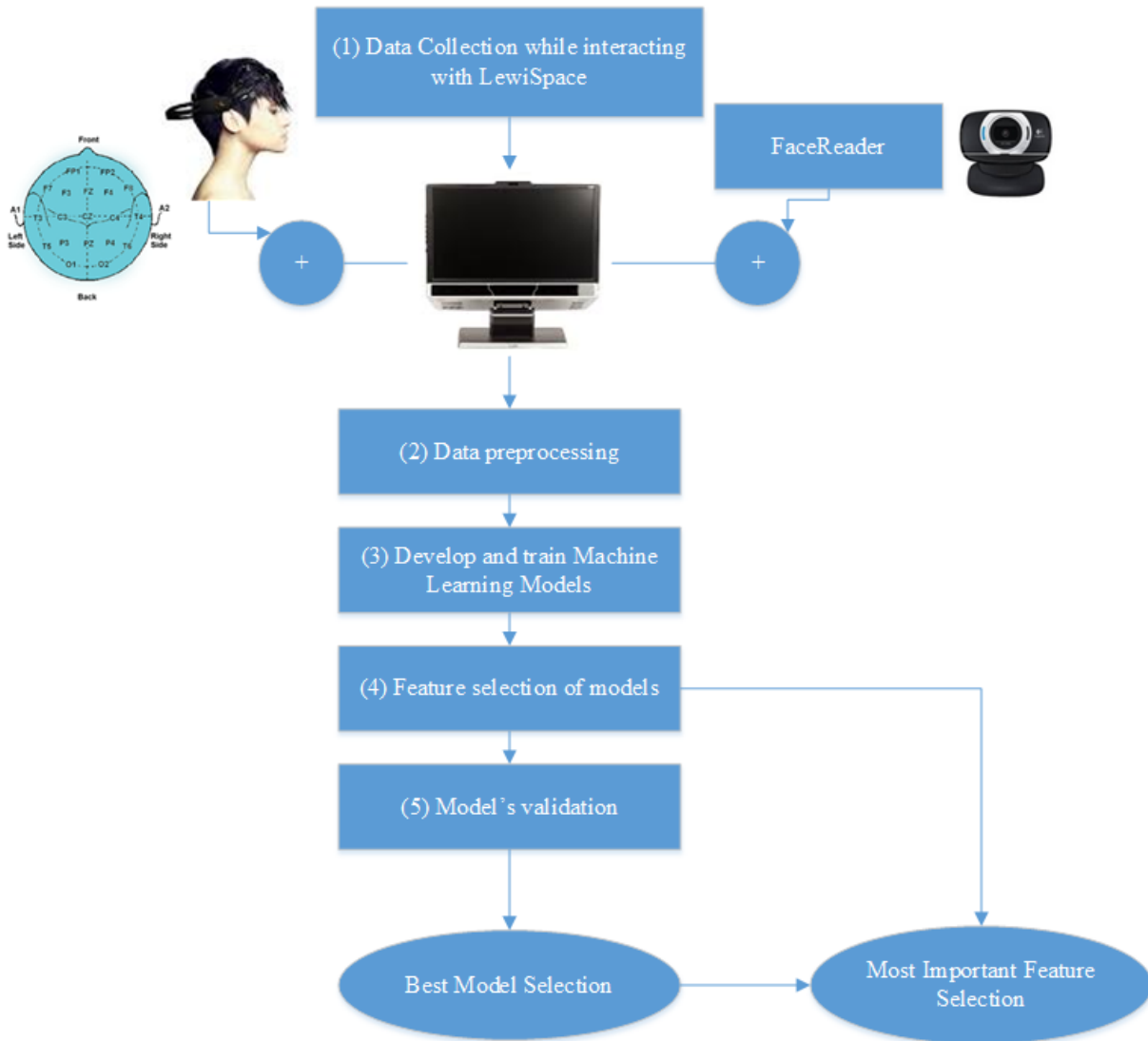


Figure 6. 2 The physiological sensing approach

6.5 Data preprocessing, results and discussion

We eliminated seven participants' data (out of forty) due to technical errors that happened during the recording of the sessions, making data unavailable for analysis (some data segments were lost during the recording, in a manner unrelated to the participants). Fifteen signals are used throughout our analysis: pupil diameter, five Affectiv indices (short-term excitement, long-term excitement, meditation, boredom, and frustration), and nine facial expression indices (neutral, sad, happy, angry, surprised, scared, disgusted, valence (of the current facial expression), and arousal). At the end of the session, participants filled out an evaluation task indicating if more

or less help was needed in each mission. The scale ranges from 1 to 3, where 1 indicates that the mission was too easy (no help required), 2 that it was adequate, and 3 that the mission was too hard (help is required). Signals (EEG, eyetracking and facial expressions recognition) occurring before a mission was completed (and after the completion of the mission) were labeled with the category self-reported by the participant (help or no help required for each mission as described above). Each mission can be further segmented into trials, where one trial is the time spent between two attempted answers for a given mission (in cases where the player immediately grasps the solution, one mission would be such a sequence (or a trial), but it is expected that many players will produce a few incorrect answers before completing the mission). In a total, **633** such **sequences** (or samples) were gathered. We extracted *four statistical parameters* for each signal, producing **feature vectors** with **60 dimensions** ($4 * 15$ outputs of signals) for each sequence. The four parameters are the mean, the standard deviation, the minimum, and the maximum of each signal for a given duration of time.

For feature and model selection of the best machine learning algorithm, we considered two contexts of analysis (and their application): 1) a **generalized model** that consists of training **offline** a single model on **all previous participants** before a game session and using its predictions for a new user, and 2) a **personalized model** that consists of training multiple models, one for each current user, **in real-time** at the beginning of a game session using two types of tasks (easy task: construct H₂ molecule, and a hard one: construct CH₄N₂O₂ molecule). These tasks will be used to calibrate the personalized model.

For the first context, we used a **leave-one-participant-out scheme**, where a model was trained on all participants' data except for the current participant (i.e. the one on which the model was tested). The 633 sequences mentioned above were reduced to 60-dimensional vectors and used as samples. For the second context, we gathered a training sample consisting of a single sequence of each class, testing all other sequences, for a single participant's data. The samples were overlapping time slices of the signals, and each slice was reduced to a 60-dimensional vector.

For the first context, **Support Vector Machine models** (the best algorithm obtained with highest accuracy) were used as classifiers, with the best hyper-parameters selected through a **grid search** through C and gamma (C of 10^{-6} to 10^4 and Gamma of 10^{-9} to 10^3).

For the first context, we tested models by ignoring some features during the training and classification tasks, allowing us to compare the accuracy of each model depending on which available features. Table 1 presents those results, showing that the **Affectiv indices** are the **best features (54.1%)**.

Table 6. 2 Balanced accuracies depending on feature selection

	All features	No pupil diameter	No Affectiv indices	No facial expression	Only Affectiv indices
Correct (%)	53.4	53.1	46.1	53.8	54.1

As for the second context, table 3 presents the confusion matrix for the best model found using a **cross validation**. The best algorithm found was **KNN** with a **Euclidean distance** and **7 neighbors**. Horizontal values are true labels, and vertical values are predicted labels.

Table 6. 3 Confusion matrix for a personalized model

	Too easy	Adequate	Too hard	Total
Too easy	0.776	0.034	0.190	58
Adequate	0.092	0.515	0.392	291
Too hard	0.101	0.165	0.733	907
Total	164	302	790	0.675

6.6 Conclusion

In the current paper, we described a study aimed at exploring whether we could predict in real time if users needed help to solve problems in LewiSpace. We suggested two approaches where machine-learning models can be trained well enough as to offer usefully accurate predictions, using widely available algorithms and a simple feature extraction process from physiological sensors.

Future work will consist of developing a second study where our game is augmented with a combination of the two proposed approaches, testing the impact of unsolicited feedback in learners' performance.

Acknowledgements

We acknowledge the CRSH, more precisely LEADS project, and NSERC for funding this work.

Chapitre 7

Utilisation de stratégies d'aide dans l'environnement virtuel *Inertia*

Nous décrivons dans ce chapitre notre dernière contribution qui s'intéresse aux stratégies d'aide dans un environnement virtuel. Cette troisième et dernière contribution consiste à proposer des stratégies d'aide visant à renforcer le raisonnement intuitif dans un nouvel environnement virtuel intitulé *Inertia* (JSI). Ce dernier est un jeu virtuel qui vise à apprendre les lois de mouvements en physique en recourant au raisonnement intuitif. Ces jeux se reposent sur la physique naïve qui sert à apprendre intuitivement (et pas scientifiquement) les lois de la physique²⁰ (Hayes, 1978). Ils intègrent naïvement les lois de la dynamique (mouvements des balles et des quilles). Il est composé de quatre jeux : Bowling 1, Bowling 2, Pivot et Billard. Dans le jeu de Bowling, le joueur doit faire tomber 10 quilles en franchissant un obstacle (mur). Il doit ainsi déterminer l'angle d'inclinaison et sa vitesse. Dans le jeu de pivot, le but est de lancer un ballon dans un seau. Le ballon ne doit pas tomber par terre en passant par les deux planches placées dans le mur. Finalement, dans le jeu de Billard, le joueur voit deux boules et doit choisir la vitesse appropriée pour tomber une boule sans la faire sortir d'un cercle. Chaque

²⁰ La physique naïve a été formulée par Patrick Hayes dans son œuvre '*Naïve Physics Manifesto*'. Elle est basée sur les connaissances intuitives et les expériences que les gens ont sur le monde des objets physiques qui les entourent. Elle ne repose pas sur les observations directes ou le caractère rationnel et scientifique du monde réel de la physique. Par exemple, le fait de lancer un ballon nécessite des connaissances en mouvement et gravité alors que cette action est évidente pour un humain. En effet, l'humain se base sur ses connaissances intuitives pour réaliser cette tâche.

jeu est constitué de 5 essais où on attribue à chaque essai un score entre 0 et 100 qui dépendent de son échec ou sa réussite ainsi que le nombre de quilles tombées dans le Bowling. Les scores totaux (variables quantitatives continues) pour chaque jeu sont normalisés à la fin du jeu pour qu'ils soient entre 0 et 100.

Ce jeu a été conçu en deux versions : une version V0 qui ne contient pas les stratégies d'aide et une version V1 incluant les stratégies d'aide. Cette version est adaptative dans le sens où elle fournit aux apprenants des stratégies d'aide en fonction de l'évolution des deux états mentaux (la frustration et l'engagement) extraits de EEG et des états d'échec/ réussite dans chaque étape du jeu (C'est pour cette raison que *Inertia* est un JSI). Par exemple, si l'apprenant a réussi et présente un état de frustration élevé, nous présentons un message de support qui sert à calmer ce dernier et l'encourager à continuer. Par contre, si l'apprenant a échoué et présente un engagement faible, nous lui présentons une stratégie d'aide qui est un complément de solution qui sert essentiellement à débloquer cet état mental. En résumé, la version V1 incluant les messages de support et stratégies d'aide sert à guider l'apprenant et à lui offrir des solutions afin qu'il progresse plus rapidement dans son raisonnement et améliore par conséquent ses performances.

Les deux versions de jeu font l'objet d'une étude expérimentale composée de 30 étudiants de l'Université de Montréal. Chaque session d'un participant dure en moyenne 30 minutes pour un total de **1 mois et une semaine**. La version V0 est expérimentée par 15 étudiants est utilisée comme groupe de contrôle, elle permet de déduire l'amélioration des performances et l'efficacité des stratégies d'aide. La version V1 inclut les stratégies d'aide qui sont composées de deux types de messages et trois formes de solutions :

- ME : message qui sert à encourager l'étudiant,
- MC : message qui sert à calmer l'étudiant,
- SO1 : stratégie complément de solution,
- SO2 : stratégie partie de solution,
- SO3 : stratégie solution complète.

Ces stratégies d'aide varient en fonction d'un diagramme qui trace les activités de l'apprenant en termes d'échecs ou de réussites et des deux états mentaux extraits par EEG en utilisant l'*affectiv suite* du casque Emotiv EPOC (la frustration et l'engagement).

Les premiers résultats obtenus par ce jeu sont comme suit : nous avons testé l'efficacité des stratégies d'aide introduites en comparant et étudiant les deux variantes de versions du jeu et nous avons collecté dans un fichier journal les paramètres suivants : le nombre d'échec, le nombre de réussite et le score global obtenu dans chaque type du jeu (4 jeux au total : variables qualitatives nominales). Nous avons constaté qu'en introduisant les stratégies d'aide le nombre d'échecs diminue dans tous les jeux et le nombre de réussite augmentent. En particulier, le nombre de réussite est doublé pour le jeu Billard que nous considérons le plus facile. Nous avons de même remarqué une amélioration de performance globale de tous les participants dans la version V1 du jeu. *Inertia* s'est montré statiquement efficace dans l'amélioration des performances des apprenants en comparant les deux versions de jeu.

Le reste de ce chapitre est constitué de l'article intitulé « *Improving Intuitive Learning Through Assistance Strategies in a Virtual Reality Game* » accepté à la conférence internationale Flairs 2017 et sous publication. Hamdi Ben Abdessalam a contribué dans cet article en développant le jeu *Inertia*. Il a aussi fourni de l'aide durant les expérimentations réalisées dans cet environnement et qui nécessitent l'installation et le réglage de deux casques : *Emotiv EPOC* pour la capture d'EEG et *Gear VR* pour l'affichage du jeu sur un téléphone intelligent en utilisant la technologie de réalité virtuelle.

Improving Intuitive Reasoning Through Assistance Strategies in a Virtual Reality Game

Ramla Ghali, Hamdi Ben Abdessalam, Claude Frasson

Accepted in : The 30th International [FLAIRS](#) Conference (2017)

Abstract

Intuitive reasoning relies on intuition, experience and affective judgment in order to resolve problems and acquire knowledge. This form of intelligence, based on unconscious learning, provides fast access to knowledge and is most of the time filled from learning by experience. Virtual Reality (VR) is a way to offer environments in which a user can deploy intuitive reasoning and acquire knowledge faster than usual academic training. In this paper, we use Inertia, a VR game that we have designed to teach some basic physics rules. We assess emotional behavior of the participants considering their engagement or frustration in the game, through electroencephalograms. Assistance strategies are integrated into the game to provide help according to the emotional measures. Inertia is composed of four types of games (Bowling 1, Bowling 2, Pivot, and Billiard) available on two versions (V0: without assistance strategies and V1: with assistance strategies). According to the data collected from the two versions, we

noticed that assistance strategies could improve participants' performance of more than 50% for some games (more precisely Billiard). The impact appeared not only in terms of reducing the number of failures, increasing the number of success, but also increasing the scores in all categories of the game.

7.1 Introduction

Nowadays, we can consider different kinds of reasoning such as inductive reasoning, deductive reasoning, comparative reasoning, example reasoning, clinical reasoning, intuitive reasoning, to mention just some of them. In this paper, we are concerned by intuitive reasoning which consists of instinctive and unconscious knowing without deduction or logical reasoning. Intuition itself is defined as the power of obtaining knowledge that cannot be acquired either by inference, observation, logical reasoning or calculation. However, this knowledge could be acquired through experience, associative learning and information stored in long-term memory (Betsch, 2008). People who use intuitive reasoning rely on instant responses including affective valences, which are reached with little apparent effort. We can therefore deduce that intuitive reasoning is based on spontaneous judgment or decision which relies on practice and experience stored in long-term memory rather than logical and analytic reasoning (Kolb, 1984).

Intuitive reasoning has many benefits. For example, intuition can greatly enhance the effectiveness of the decision making. It's also fast and theory free. It relies only on practice. Whereas, we should note that intuition does not provide accurate predictions of the problem like analytic reasoning. To summarize, using either intuitive reasoning or analytic reasoning have some advantages and disadvantages. So, choosing which reasoning is most suitable relies on the nature of the problem and the resources provided to solve it.

As we said above **intuitive reasoning** can emerge from subconscious state. One of the advantages of VR, compared to other interactive environments, is that the user is isolated from external distractions and this condition allows better access to the subconscious. Immersion in VR can trick the subconscious mind of the user, so that he believes he is in a real world. As intuition is an automatic non-conscious process (Frankish, 2010), it resides in the subconscious mind. So, using a VR environment is a way to favor and develop intuitive reasoning. VR allows to trigger intuitive interpretation of the situation with actions or decisions emerging from the

common sense and this is a very interesting channel for strengthening reasoning and more generally learning. In addition, if we combine VR environment with a game, which can increase engagement and motivation of the user, we improve his/her capability of knowledge acquisition. With this goal in mind we designed **Inertia**, a VR game intended to teach basic physics rules to : 1) provide a learning environment able to trigger intuitive reasoning, and 2) to test assistance strategies able to help the learner according to his cognitive state.

- **Mental assessment.** To evaluate the reasoning state of the learner we focus on two mental states: *engagement* (the learner is fully involved and interested in the reasoning), and *frustration* condition (the learner is lost and does not find the solution). These mental states are analyzed using electroencephalograms.
- **Assistance strategies** can be defined as strategies to help the learner to reason correctly and progress in the knowledge acquisition. To help him, we focus on three types of *solutions* and two types of *messages* according to the learner's mental state and his success or failure in each exercise. The strategies are grouped into: complement of solution using a suggestion provided to the learner in order to help him find the solution (Green et al., 2012), part of solution using some elements of the solution, and solution with the whole solution so that he can reach the next level of the game. Messages are decomposed into: messages of encouragement to encourage the learner to continue playing, and messages designed to calm the learner and decrease his frustration.

The paper is organized as follows. In section 2 we talk about how to improve intuitive reasoning. Section 3 describes the VR environment and the main characteristics of the game. Section 4 describes the principles of the assistance strategies. In section 5 we present the experiment which led to the results presented and discussed in Section 6.

7.2 Improving Intuitive Reasoning

7.2.1 Intuitive reasoning

As we mentioned before, intuitive reasoning relies on intuition, experience and affective decision used to resolve problems. Intuition is also known as a form of unconscious intelligence

(Gigerenzer, 2007; Kruglanski et Gigerenzer, 2011), very different from deliberate and analytic reasoning which are based on calculation. It's rather based on affective and unconsciousness behaviors. We can see this type of intuition while practicing sports, shopping, choosing friends, etc. We also consider that this behavior occurs when *playing games*. Games and particularly serious games (SG) are tools that can be used to enhance learning (Dunwell et al., 2012). We believe that one method to enhance intuitive reasoning is first to choose games able to trigger this type of reasoning, and particularly VR games.

7.2.2 Virtual Reality Games

Games and applications could be presented to the player according to a variety of forms. It could be presented on a normal computer monitor, smartphone screen, 3D computer monitors which work with 3D glasses, etc. However, the most immersive way to present a video game to the player is VR. With this technology, the player don't see anything that could disturb him except the virtual world (he is not solicited by external factors). VR immersion tricks the mind of the player, so he thinks that he is really inside the game and he acts like in normal life. This immersion is driven especially by the subconscious mind of the player, so his reactions will be based more on his intuition than on his logic. Thus, the first advantage of VR is the immersion which could be useful for both the conscious and subconscious mind. The user can focus more, learn faster, and think better, since he is isolated in the VR world. The immersion is even more efficient to the subconscious mind because it lets the user learn by experience and use his intuition to succeed and find the solution.

VR can be used and can be useful in several domains such as Education (Freina et Ott, 2015; Merchant et al., 2014; Pan et al., 2006), Social life (Didehbani et al., 2016; Freeman et al., 2014; Ip et al., 2016), Medicine and health (Didehbani et al., 2016; Miller et al., 2014; Price et al., 2014), and neuroscience (Bohil et al., 2011; Frankish, 2010; Yates et al., 2016). In education domain, (Freina et Ott, 2015) show that the immersive property of VR offers many advantages and allows the students to feel objects and events that are not physically reachable. In neuroscience domain, Bohil and his team show that VR is useful for neurosurgery and cognitive assessment. In fact he mentions that "*VR allows naturalistic interactive behaviors to take place while brain activity is monitored via imaging or direct recording* (Bohil et al., 2011)."

Thus, intuitive simulation is a powerful approach to problem solving situations. This is particularly useful in 3D environments where the result of an action depends on position, force to apply, resistance of elements and physical factors. For all these reasons we have developed the Inertia game, which is described in the next section.

7.2.3 Assistance strategies

With such an environment able to trigger intuitive reasoning, assistance strategies can then be used to facilitate teaching and to progress more effectively in the game (Ghali, Frasson, et al., 2016; Ghali, Ouellet, et al., 2016b; McNamara et Jackson, 2010). For instance, McNamara et al. described five main characteristics that describe SG as game based features. Game based features aim to motivate learners and increase their playful aspect. However, Ghali et al. (2015, 2016) mention that one problem of SG is that they concentrate more on playful aspect and don't focus on educational aspect. In order to solve this problem, the authors proposed real time models for providing learners with help if necessary. Their models used machine learning techniques and predicted the need of help with the accuracy of 54.1%. The prediction was based on a collection of physiological data (electroencephalography, eye tracking and emotional recognition). The authors showed that it is possible to build reliable machine model classification of three states: (1) no help required, (2) medium help required, and (3) strong help required. Instead of using only messages of help, we propose in this paper to introduce also assistance strategies that are composed of three types of help: (1) complementary of solution (Green et al., 2012), (2) part of solution, and (3) whole solution. We suppose that introducing these strategies in our SG will be a second method to improve intuitive reasoning and then enhance learning.

As proposed by Ghergulescu (Ghergulescu et Muntean, 2014), the use of EEG indices of mental states is effective and can produce reliable results. In order to guide our instructions of assistance strategies, we propose to extract two mental states from electroencephalograms: (1) **engagement** that gives a clue of learner's attention and concentration while learning, and (2) **frustration** that gives a clue if the learner is bored and/or not interested in the content. We choose these two mental states because they give us an indication of the whole mental state of a learner and if he/she is progressing in reasoning or stacked. According to the evolution of these

two mental states, we generate our assistance strategies in order to provide learners with an adequate help.

7.3 Inertia: a Virtual Reality Game for Learning Dynamic Rules

In order to monitor the mental state of the learner and test our assistance strategies, we created a game called Inertia. Inertia is based on the principle of force transformation and quantity of movement, known in physics of dynamics. For example, what force and direction should we apply to a billiard ball in order to reach a specific target? We created a platform of several games (3 games: Bowling, Pivot, and Billiard, and 4 levels: Bowling 1, Bowling 2, Pivot, and Billiard) using Unity 5.3.1 which contains a built-in physics engine able to handle a simulation of dynamics.

Inertia was optimized in order to run and react quickly in a VR mobile headset (Samsung Gear VR headset). The player can control the game with a bluetooth game-pad, and with the rotation of his head he can see the VR environment in 360 degrees. In this way the learner is immersed into an intuitive environment in which rules to apply come from the subconscious (sometimes called rules of thumb).

In order to cover a variety of physics rules, we built three categories of games in Inertia: Bowling, Pivot and Billiard. For the Bowling, we made two difficulty levels and just one for the other games. For each game, the player has five attempts to play and the score appears after each attempt. A help message is displayed after every trial, and its content depends both on the failure or success and on his mental state as explained in the previous section.

The score of each game is calculated as follows: for Bowling, the score depends of the number of pins that the ball hits, it varies between 0 and 10 pins. The success is considered if the player shoots 5 pins (50%). Whereas, the score of Pivot game depends only on the success or failure. If he succeeds, he gets 100%, else he gets 0%. The success in this game is considered if he can get the red ball into a bucket. Similarly, the score of Billiard bowling depends only on the success or failure. If he succeeds, he gets 100%, else he gets 0%. For this game, the success is considered if the player can drop the coin which is situated on the orange ball inside the black

circle. Some screenshots of each game developed in Inertia are presented in Figure 7. 1, together with a description of the difficulty to overcome.





Screenshot	Description
	<p>Bowling1: in this difficulty level, we put an obstacle in the middle of the track. The player should choose the right angle and thrust, so that the ball can pass the obstacle and reach the pins.</p>
	<p>Bowling2: in this difficulty level, we changed the type of ball to a soccer ball, so that the player has to consider the bounce of the ball and find the right angle and thrust as in Bowling1.</p>
	<p>Pivot: in this game, we placed two tilting wooden beams initially in equilibrium. A black metal ball falls onto one beam and inclines more or less the beam which hits the other beam. The player has to choose the mass of the black metal ball in order to tilt the wooden beams and let the red ball fall inside the bucket.</p>
	<p>Billiard: in this game, a balanced coin is placed on the orange ball. The player has to apply the right thrust to hit the white ball in order to drop the coin inside the black circle.</p>

Figure 7. 1 Details of Inertia

7.4 Assistance Strategies in Inertia

The principle of assistance strategies is based on two types of messages (ME, MC) and three types of solutions (SO1, SO2 and SO3) which are sent to the learner according to their mental state and result:

- ME: presents a message in order to encourage the player to progress in the game,
- MC: presents a message in order to calm the student and decrease his state of frustration,
- SO1: complement of solution that tries to present the adequate message according the learner's choice of parameters of the game,
- SO2: part of solution that gives exactly one parameter of the game,
- SO3: whole solution given to the learner in order to let him progress in the game.

These messages and solutions vary according to two mental states: frustration and engagement, extracted from Electroencephalograms, and two evaluation metrics: success or failure in each game. An overview of the distribution of these strategies is presented in Figure 7. 2.

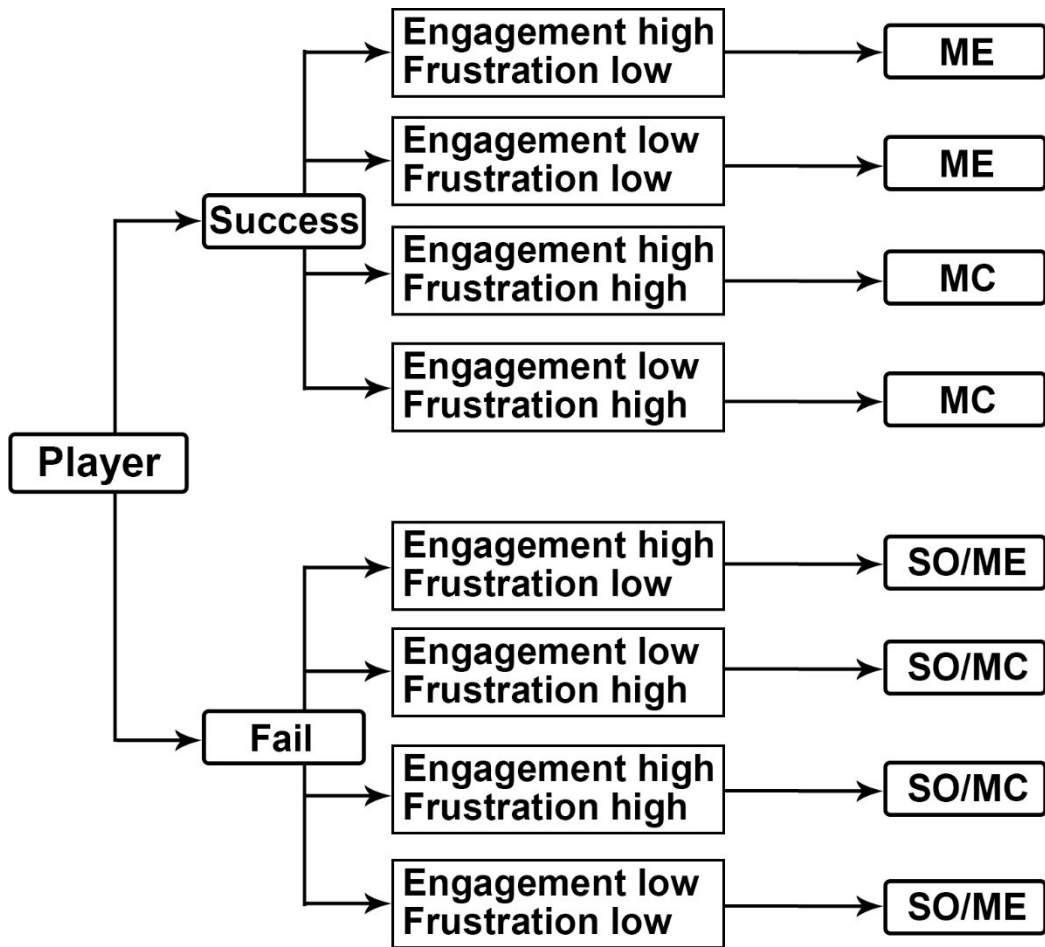


Figure 7. 2 An overview of the distribution of messages and solutions in Inertia

This distribution also varies according to each trial of a level. For example, for the best game bowling, the player can try maximum five times. If he successes, we normally generate a message (ME or MC). However, if he fails we present in this case the strategies with the corresponding messages. Table 7. 1 gives an example of assistance strategies integrated into a game (more specifically in the bowling game).

Table 7. 1 Examples of strategies in bowling 2 game

Messages and Solutions	Examples
ME: E high, F low, Success	Well done! You are a good player! Continue like this.
MC: E Low, F high, Failure	Don't worry, try to play again. Be careful, your behavior seems to be disengaged, so try to be more concentrated and more engaged
SO1: complement of solution	If $Speed > Speed_{Optimal}$: your speed is too high. You must apply less power
SO2: part of solution	Okay, we will help you. Let's choose an angle between 5 and 10 degrees
SO3: whole solution	The right angle is 7 and the right power is 90

7.4 Experiment

In order to study the effectiveness of assistance strategies, we built two versions of Inertia: V0, a version without assistance strategies, and V, a version including assistance strategies. The two versions were experimented on 30 participants from University of Montreal (11 female and 19 male, mean age 28.76). In the first step of the experiment (see Figure 7. 3), the participant signs an ethic form that explains the study and mentions its advantages and disadvantages. In the second step, the participant is equipped with the Emotiv EPOC headset in order to extract the two mental states: frustration and engagement. These states are used further to generate the assistance strategies.

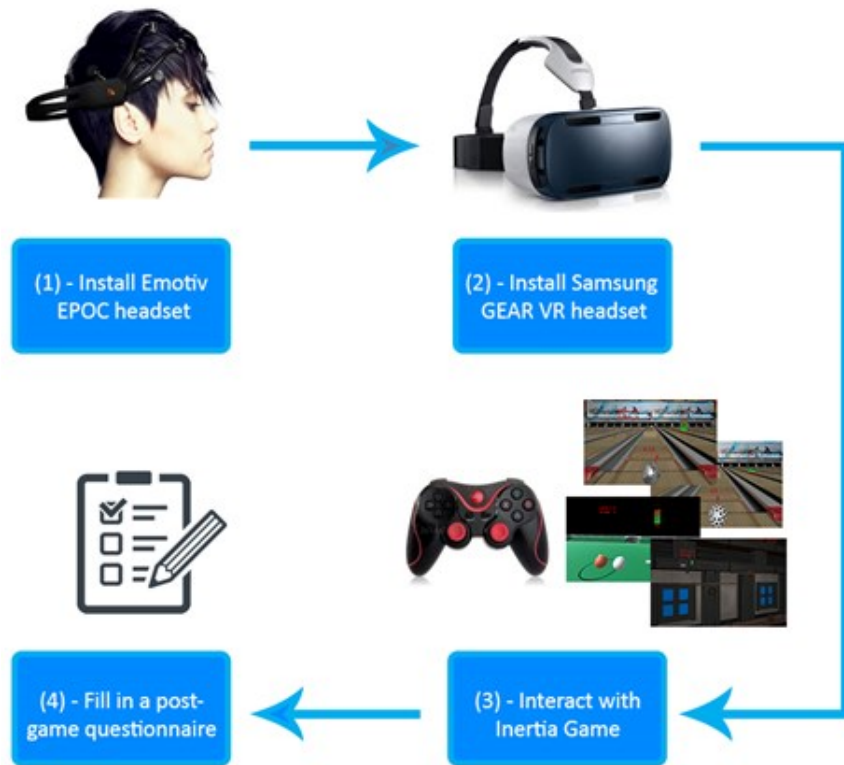


Figure 7. 3 The experimental process

In the third step, we set up the game into a galaxy 6 smart phone which is placed in the headset GEAR VR and install this headset on the participant.

After these steps, the participant could manipulate the 3D environment and practice his skills with Inertia game described above. Finally, we ask the participant to fill in a post-game questionnaire. This questionnaire is used to collect demographic information from the participants as well as their need of help in each section of the game.

7.5 Results and discussions

To compare the two versions of Inertia, we calculated in each version the number of failure and success as well as the scores' variation, grouped by the type of game. The two following figures show the variation of the number of failure and success in each type of game. Figure 7. 4 gives statistics on Inertia V0 in terms of the number of failure and success, whereas Figure 7. 5 illustrates the same statistics on inertia V1 with assistance strategies.

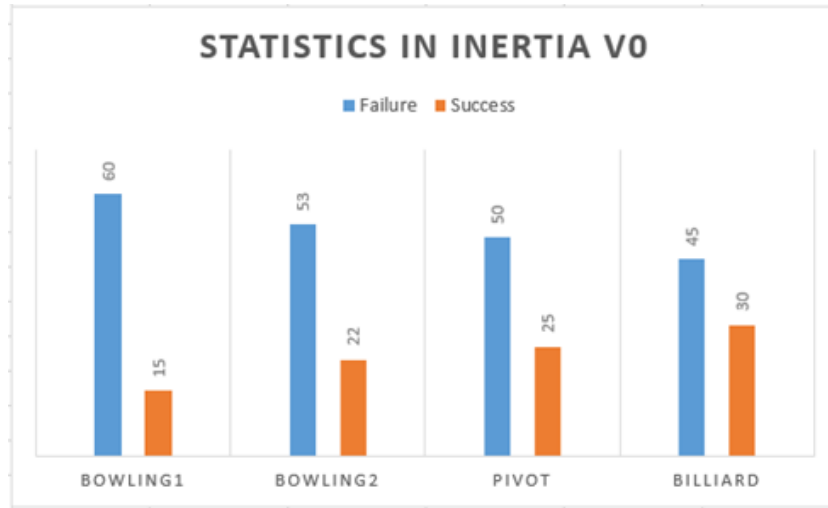


Figure 7. 4 Descriptive Results in Inertia V0

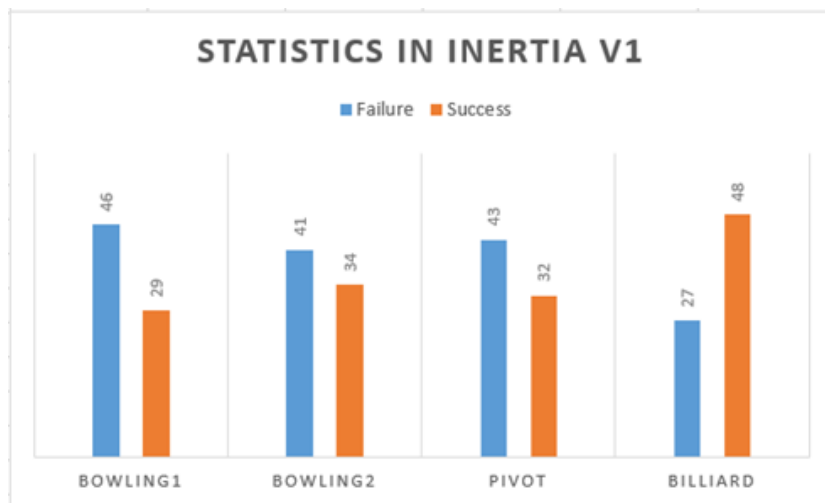


Figure 7. 5 Descriptive Results in Inertia V1

From Figure 7. 4 and Figure 7. 5, we can notice clearly that the Bowling is the hardest game with a maximum of failure in both versions (60% vs 46%). Moreover, we can see clearly that the assistance strategies added to Inertia V1 have a strong impact on reducing the number of failures and in-creasing the number of success in all the games.

From Figure 7. 5, we also notice that the success in Billiard game increased significantly and the failure was reduced by quite half (45% vs 27%). This means that assistance strategies

were specifically more successful in this game. To prove this result, we also conducted descriptive results on scores and run a t-test that will be described later.

For the descriptive results, we calculate for Inertia V0 and V1 the average of scores for all the participants, as well as the standard deviation (SD), minimum (Min) and maximum (Max). These results are presented respectively in tables 7. 2 and 7. 3. It should be noted that all the scores for all the games are normalized between 0 and 100.

Table 7. 2 Descriptive statistics of scores in Inertia V0

Game	Mean	SD	Min	Max
Bowling1	16.8	17.42	0	46
Bowling2	28.53	17.06	0	62
Pivot	33.33	16.19	0	60
Billiard	40	16.19	20	80

From table 7. 2, we confirm our finding that the Billiard is the easiest game for most of the participants because it has the highest score's average (40%). However the Bowling1 remains the most difficult one (16.8%). This can be explained by the fact that a success in a bowling needs that a player shoots more than five pins out of 10 in only one trial.

Table 7. 3 Descriptive Statistics of scores in Inertia V1

Game	Mean	SD	Min	Max
Bowling1	33.5	20.12	15	52
Bowling2	41.73	19.73	28	68
Pivot	42.66	23.83	0	80
Billiard	64	22.59	20	80

From table 7. 3, we see clearly that for **all the games** the mean score is higher in Inertia V1 compared to Inertia V0. This result could be explained by the impact of assistance strategies and their role to guide the players in almost all the games. The assistance strategies can improve learners' reasoning and let them progressing faster in the game.

Finally, to confirm that hypothesis and to study the scores' improvement in Inertia V1, an independent-samples t-test was conducted to compare scores for games in both versions. There was a significant difference in the scores for V0 (M=29.66, SD=18.27) and V1 (M=47.4, SD=23.62) conditions; $t(59)=5.09$, $p = 0.00^{**}<1\%$. These results are very significant and show clearly the difference between the two versions of the game. Moreover, we calculated the effect size (a Cohen'd of 0.84) which is large. This proves that there is an improvement in overall learners' scores. We also conducted an analysis of variance ANOVA to see if the games are statistically different. Results considering four types of games (independent variables) are also very significant ($F(3.86)=10.98$, $p=0.00^{**}$) showing that the design of the games is different. Despite this difference in the games' design, the assistance strategies proposed play an important role in enhancing learning.

7.6 Conclusions

In this paper, we studied the importance of adding assistance strategies in an intuitive reasoning VR game. We conducted an experiment and tests. Results showed that for all the games, learners' performance increased when adding assistance strategies, both taking in consideration the global score and the number of failure and success. This can be a useful finding for improving the training in intuitive situations such as physical interventions in many domains (sport, surgery, and marketing) where the decision is based on intuitive behavior and contributes to learning by experience. Future work will aim to experiment more strategies on the Inertia platform.

Acknowledgements

We acknowledge the CRSH, more precisely LEADS project, and NSERC for funding this work. We want also to thank professor Chokri Awada for his help in Physics.

Chapitre 8

Conclusion

Dans ce dernier chapitre, nous faisons une synthèse de nos trois contributions et examinons nos travaux futurs.

8.1 Contributions

La première contribution de cette thèse a été de faire une classification et une prédiction des performances de l'apprenant dans des tâches cognitives (l'environnement logique). À cet effet, nous avons entraîné et testé plusieurs algorithmes d'apprentissage machine (modèles linéaires et non linéaires) afin de regrouper les résultats des apprenants selon quatre groupes (allant des résultats très faibles à des résultats très forts). Nous avons ainsi proposé une approche physiologique où nous collectons les données physiologiques issues de trois états mentaux extraits de EEG (engagement, charge cognitive et distraction (Berka et al., 2004; Stevens et al., 2007a; Stevens et al., 2007b)) et des données issues de l'activité électrodermale de la peau.

Pour valider cette approche, nous avons mené une étude expérimentale où nous avons collecté les données de **20 participants** (états mentaux, informations sur la tâche, sexe, âge et émotions auto-évaluatives). Cette étude expérimentale a duré **1 mois** où nous avons collecté une base de données des mesures physiologique des participants sur notre environnement expérimental logique (les données issues de EEG et GSR). Deux modèles (linéaires et non linéaires) ont été proposés pour la prédiction des scores : la régression multiple et les arbres des forêts. Avec la régression multiple, nous avons constaté que l'information sur la tâche (difficulté de la tâche, nature de la tâche, durée et catégorie) constitue un attribut essentiel dans la

prédiction des scores et que deux caractéristiques principales déterminent le résultat de l'apprenant : la difficulté de la tâche et la catégorie de la tâche. Les autres caractéristiques sont très négligeables (le sexe est le dernier à considérer). Pour la classification, nous avons montré aussi que les arbres de forêts ont tendance dans notre échantillon de données à mieux performer pour la prédiction des performances de l'apprenant avec les caractéristiques les plus importantes sont en ordre ascendant : le niveau de difficulté, la durée de la tâche, l'engagement, la distraction et la charge cognitive. Les autres caractéristiques (en particulier le sexe) sont négligeables. Cette contribution a fait l'objet de publications présentées dans le chapitre 3 et le chapitre 4.

Notre deuxième contribution a consisté à prédire le besoin d'aide dans l'environnement éducatif (JS) *LewiSpace*. Pour ce faire, nous avons opté principalement pour deux approches : une approche généralisée où les données des algorithmes d'apprentissage machine sont entraînées sur les anciens participants sauf le participant en cours. Les données des anciens participants permettent de prédire le besoin d'aide chez le participant en cours. De même, nous avons opté pour une approche individualisée où les algorithmes d'apprentissage machine sont entraînés et testés sur le même participant. La calibration de ce modèle se fait en utilisant au début de jeu deux tâches (facile et complexe) pour pouvoir prédire dans le reste du jeu s'il a besoin d'aide.

Pour ce faire, une étude expérimentale a été faite où nous avons collecté les données de **40 participants** de l'Université de Montréal. Cette étude a duré **2 mois et une semaine**. Durant cette étude, nous avons collecté différentes données physiologiques de l'apprenant (états mentaux extraits de *Emotiv Epoc*, diamètre de la pupille, émotions, valence, activation et données de personnalité). Ces données ont servi à construire notre approche et entraîner et tester différents algorithmes d'apprentissage machine. Les résultats préliminaires ont montré que les machines à vecteurs de supports ont tendance à performer mieux pour la prédiction des résultats de l'apprenant (précision de 54,1% qui est **supérieur** à un random baseline de 33.3%) et que les algorithmes des plus proches voisins sont meilleurs pour l'approche individualisée avec 7 voisins.

Notre troisième contribution a consisté à recourir à des stratégies d'aide qui permettent de renforcer le raisonnement intuitif dans un jeu virtuel (JSI) *Inertia*. Ces stratégies sont sous forme de deux types de messages d'aide (encouragement et calme) et trois types de solutions

(complément de solution, partie de solution et solution complète). Pour étudier l'impact de ces stratégies, nous avons collecté les données d'échec, de réussite et la performance globale durant les quatre types de jeux formant *Inertia* (Bowling 1, Bowling 2, Pivot et Billard).

Pour ce faire, une étude expérimentale a été réalisée où nous avons collecté les données de deux états mentaux extraits d'EEG (frustration et engagement) pour le JSI *Inertia* sur **30 participants** de l'Université de Montréal. Cette étude a duré **1 mois et une semaine**. La variation de ces deux états ainsi que le nombre d'échecs et de réussites servent à définir une version V1 adaptative du jeu (un JSI). La version V1 fournit de l'aide aux apprenants en fonction de l'évolution des états mentaux et l'échec ou la réussite dans chaque tentative du jeu. Cette version a été expérimentée par 15 étudiants de l'Université de Montréal qui ont participé volontairement à l'étude expérimentale. Les performances des apprenants ont été comparées à la version V0 (groupe de contrôle composé de 15 autres participants) où les apprenants expérimentaient l'environnement *Inertia* sans recourir à des stratégies d'aide. Les résultats préliminaires avec l'expérimentation de cet environnement ont montré que pour tous les participants et pour les types de jeux, il existe une amélioration significative des performances. De plus, le nombre d'échec a diminué dans tous les jeux et le nombre de réussite a augmenté. Plus précisément, nous avons reporté que cette valeur a été doublée pour le jeu Billiard que nous constatons comme le plus facile.

8. 2 Travaux futurs

Les résultats de nos travaux ont des implications sur les jeux sérieux, et plus généralement les interactions homme machine, qui visent à établir un suivi régulier de l'utilisateur et lui offrir des méthodes plus adéquates pour améliorer ses performances et accroître son expérience d'interaction. Nos travaux présentent cependant quelques limitations qui laissent entrevoir de nombreuses perspectives de recherche. En voici quelques exemples :

- **Étendre la base de données de mesures physiologiques :**

Notre base de données des mesures physiologique contient un nombre restreint des participants dans chacune de nos trois expériences (20, 40 et 30 respectivement). Pour améliorer les résultats des algorithmes d'apprentissage

machine et généraliser les résultats trouvés, nous comptons ajouter d'autres participants dans les expérimentations et collecter d'avantage d'autres mesures physiologiques. Ces mesures servent à tester et valider les résultats trouvés par les algorithmes d'apprentissage machine.

- **Tester la fiabilité des mesures et des algorithmes sur d'autres environnements d'apprentissage :**

Pour que les résultats d'apprentissage machine soient générales et valides pour les environnements cognitifs, jeux sérieux ou jeux sérieux intelligents, il serait utile d'expérimenter d'autres environnements dans le même type et collecter les mesures physiologiques. Par la suite, nous menons des expérimentations sur les environnements développés et validons nos hypothèses et résultats.

- **Se focaliser sur les données d'états mentaux dans l'environnement LewiSpace :**

Les résultats obtenus montrent que les états mentaux sont les caractéristiques les plus importantes à considérer dans les algorithmes d'apprentissage machine. De ce fait, nous proposons d'utiliser uniquement ces données pour entraîner et tester les algorithmes d'apprentissage machine et étudier par la suite les résultats en termes de précision et de score F1 (le score F1 est une mesure plus élaborée de précision qui dans sa formule prend en considération les mesures de la précision et de rappel). Nous envisageons d'aller plus loin en examinant la contribution des ondes cérébrales (alpha, beta, gamma, delta) dans la prédiction d'aide. Nous proposons alors d'étudier des fonctions d'apprentissage machine plus complexes qui prennent en compte la pondération des poids pour les différents ondes cérébrales. Cette proposition sera étudiée selon les deux approches proposées (généralisée et individualisée).

- **Prédire les performances de l'apprenant dans l'environnement LewiSpace :**

Dans l'environnement JS *LewiSpace*, nous n'avons pas tenu compte de l'évolution et de la variation des performances de l'apprenant. De ce fait, nous proposons d'entraîner et de tester les algorithmes d'apprentissage machine pour

prédire les résultats des apprenants en fonction de la difficulté de chaque mission (facile ou difficile). Nous allons de même étudier les caractéristiques les plus pertinentes qui contribuent à la variation de la performance de l'apprenant.

- **Expérimenter de nouvelles stratégies dans l'environnement Inertia :**

Dans l'environnement Inertia, nous avons expérimenté seulement cinq stratégies d'aide. Nous proposons alors d'intégrer dans cette plateforme d'autres stratégies d'aide facilitant plus les interactions avec l'environnement et améliorant l'état cognitif global de l'apprenant.

- **Apprentissage machine pour la génération des stratégies d'aide dans Inertia :**

Dans l'environnement actuel JSI *Inertia*, nous avons défini des seuils pour décider quand les mesures d'engagement et de frustration sont élevées ou basses. Cette méthode est acceptable mais n'est pas assez intelligente. Nous proposons plutôt à partir des performances de l'apprenant et des mesures de frustration et d'engagement d'entraîner des algorithmes d'apprentissage machine qui permettent de prédire quand il faut générer une stratégie d'aide et quelle stratégie sera plus adéquate à présenter. Cette méthode est beaucoup plus flexible et intelligente surtout que nous avons assez de données (30) pour l'entraînement et le test des algorithmes d'apprentissage machine.

- **Prédire le besoin d'aide dans l'environnement Inertia :**

Lors de l'interaction avec l'environnement *Inertia*, nous avons collecté un questionnaire auto-évaluatif du besoin d'aide avant et après le jeu. Les données collectées peuvent servir comme étiquettes pour entraîner les algorithmes d'apprentissage machine pour prédire le besoin d'aide dans cet environnement et le fournir en temps réel tel que proposé dans l'environnement *Lewispace*.

- **Mesure du stress et adaptation selon cet état dans les jeux sérieux :**

Nous pouvons conduire une étude expérimentale soit dans l'environnement *LewiSpace* ou *Inertia* pour mesurer l'état de stress grâce à la variation de

l'activité électrodermale. Cet état physiologique mesuré peut être combiné avec un questionnaire auto-évaluatif du stress afin d'entraîner et de valider des modèles d'apprentissage machine. Ces jeux peuvent ainsi être adaptatifs en fonction de l'évaluation des états mentaux d'EEG, de besoin d'aide et de stress.

Bibliographie

- Aleven, V., McLaren, B., Roll, I. et Koedinger, K. (2006). Toward Meta-cognitive Tutoring: A Model of Help Seeking with a Cognitive Tutor. *International Journal of Artificial Intelligence Education*, 16(2), p. 101-128.
- Aliseda, A. (2006). What is abduction? Overview and Proposal for Investigation, *Abductive Reasoning*: Springer, p.26-50.
- Alvarez, J. (2007). *Du Jeu Vidéo au Serious Games*. (Université de Toulouse, France).
- Alvarez, J. et Michaud, L. (2008). Serious Games : Advergaming, edugaming, training and more Dans IDate (dir.), *Executive Summary*, France, p. 83.
- Andresen, K. et Gronau, N. (2005). *Seeking Optimal IT Strategies by The Determination of Adaptability in Domain-Specific Software Applications*. Proceeding of the 2005 Information Resources Management Association International Conference, San Diego, CA, USA.
- Atkinson, R. C. et Shiffrin, R. M. (1968). Human Memory: A Proposed System and its Control Processes1. Dans W. S. Kenneth & S. Janet Taylor (dir.), *Psychology of Learning and Motivation*: Academic Press, 2, p.89-195.
- Azevedo, R., Johnson, A., Chauncey, A. et Burkett, C. (2010). Self-regulated Learning with MetaTutor: Advancing the Science of Learning with MetaCognitive Tools. Dans M. S. Khine & I. M. Saleh (dir.), *New Science of Learning: Cognition, Computers and Collaboration in Education*, New York: Springer, p.225-247.
- Baddeley, A. (1986). *Working Memory*. Oxford:: Clarendon Press.
- Baikadi, A., Rowe, J., Mott, B. et Lester, J. (2014). Generalizability of Goal Recognition Models in Narrative-Centered Learning Environments. Dans V. Dimitrova, T. Kuflik, D. Chin, F. Ricci, P. Dolog & G.-J. Houben (dir.), *User Modeling, Adaptation, and Personalization: 22nd International Conference, Aalborg, Denmark, July 7-11*, Cham: Springer International Publishing, p. 278-289.
- Baker, R. J., Godwa, S. M. et Corbett, A. T. (2012). *Automatically detecting a student's preparation for future learning: Help use is key*. 4th International conference of Data mining.

- Baker, R. J. et Rossi, L. (2013). Assessing the Disengaged Behaviors of Learners, *Design Recommendations for Intelligent Tutoring Systems*: <http://www.columbia.edu/~rsb2162/BakerRossi2013.pdf>.
- Baker, R. S., Corbett, A. T. et K. R. Koedinger, K. R. (2004). *Detecting student misuse of intelligent tutoring systems*. The international Intelligent tutoring systems conference, August 30 - September 3, Maceió, Alagoas, Brazil.
- Baker, R. S. J. d. (2007). *Modeling and understanding students' off-task behavior in intelligent tutoring systems*. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, San Jose, California, USA. doi: 10.1145/1240624.1240785.
- Baker, R. S. J. d., Corbett, A. T. et Aleven, V. (2008). More Accurate Student Modeling through Contextual Estimation of Slip and Guess Probabilities in Bayesian Knowledge Tracing. Dans B. P. Woolf, E. Aïmeur, R. Nkambou & S. Lajoie (dir.), *Intelligent Tutoring Systems: 9th International Conference, ITS 2008, Montreal, Canada, June 23-27*, Berlin, Heidelberg: Springer Berlin Heidelberg, p. 406-415.
- Barrick, M. R. et Mount, M. K. (1991). The Big Five Personality Dimensions And Job Performance: A Meta-Analysis. *Personnel Psychology*, 44(1), p. 1-26. doi: 10.1111/j.1744-6570.1991.tb00688.x.
- Bartels, M. et Marshall, S. P. (2012). *Measuring cognitive workload across different eye tracking hardware platforms*. Proceedings of the Symposium on Eye Tracking Research and Applications, Santa Barbara, California. doi: 10.1145/2168556.2168582.
- Beck, J. E., Chang, K.-m., Mostow, J. et Corbett, A. (2008). Does Help Help? Introducing the Bayesian Evaluation and Assessment Methodology. Dans B. P. Woolf, E. Aïmeur, R. Nkambou & S. Lajoie (dir.), *Intelligent Tutoring Systems: 9th International Conference, ITS 2008, Montreal, Canada, June 23-27*, Berlin, Heidelberg: Springer Berlin Heidelberg, p. 383-394.
- Bergqvist, E. (2007). Types of reasoning required in university exams in mathematics. *The Journal of Mathematical Behavior*, 26(4), p. 348-370. doi: <http://dx.doi.org/10.1016/j.jmathb.2007.11.001>.
- Berka, C., Levondowski, D. J. et Cvetinovic, M. (2004). Real-time Analysis of EEG Indices of Alertness, Cognition, and Memory Acquired with a Wireless EEG Headset. *International Journal of Human-Computer Interaction*, 17(2), p. 151-170.
- Bernardini, S., Porayska-Pomsta, K. et Smith, T. J. (2014). ECHOES: An intelligent serious game for fostering social communication in children with autism. *Information Sciences*, 264, p. 41-60. doi: <http://dx.doi.org/10.1016/j.ins.2013.10.027>.

- Betsch, T. (2008). The nature of intuition and its neglect in research on judgment and decision making. Dans H. Plessner, Betsch, C. and Betsch, T. Eds (dir.), *Intuition in judgment and decision making*. New York: London: US: Lawrence Erlbaum Associates Publishers, 334 p.
- Bisson, F., Larochelle, H. et Kabanza, F. (2015). *Using a recursive neural network to learn an agent's decision model for plan recognition*. Proceedings of the 24th International Conference on Artificial Intelligence, Buenos Aires, Argentina.
- Bohil, C. J., Alicea, B. et Biocca, F. A. (2011). Virtual reality in neuroscience research and therapy. *Nat Rev Neurosci*, 12(12), p. 752-762.
- Boutilier, C. (1992). *Conditional Logics for Default Reasoning and Belief Revision*: University of British Columbia.
- Breiman, L., Freidman, J., Stone, C. J. et Olshen, R. A. (1984). *Classification and regression trees*. Taylor & Francis. 368 p..
- Bunge, S. A., Dudukovic, N. M., Thomason, M. E., Vaidya, C. J. et Gabrieli, J. D. E. (2002). Immature Frontal Lobe Contributions to Cognitive Control in Children. *Neuron*, 33(2), 301-311. doi: 10.1016/S0896-6273(01)00583-9.
- Calvo, R. A. et D'mello, S. K. (2010). Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications. *IEEE Transactions on Affective Computing*, 1(1).
- Cardoso, R., de Souza, E., Camano, L. et Roberto Leite, J. (2004). Meditation in health: an operational definition. *Brain Research Protocols*, 14(1), p. 58-60. doi: <http://dx.doi.org/10.1016/j.brainresprot.2004.09.002>.
- Cetintas, S., Si, L., Xin, Y. P. P. et Hord, C. (2010). Automatic Detection of Off-Task Behaviors in Intelligent Tutoring Systems with Machine Learning Techniques. *IEEE Transactions on Learning Technologies*, 3(3), p. 228-236.
- Chaffar, S. et Frasson, C. (2006). *Predicting Learners'Emotionl Response in Intelligent Distance learning Systems*.The 19th International FLAIRS Conference, 15-17 May, Melbourne , Florida , USA.
- Chaouachi, M., Chalfoun, P., Jraidi, I. et Frasson, C. (2010). *Affect and Mental Engagement: Towards Adaptability for Intelligent Systems*. Proceedings of the 23rd International FLAIRS Conference, May 19-21, Daytona Beach, Florida, USA.

- Chaouachi, M. et Frasson, C. (2012). *Mental workload, engagement and emotions: an exploratory study for intelligent tutoring systems*. International Conference on Intelligent Tutoring Systems, June 14-18, Chania, Crete, Greece.
- Chaouachi, M., Jraidi, I. et Frasson, C. (2015). MENTOR: A Physiologically Controlled Tutoring System. Dans F. Ricci, K. Bontcheva, O. Conlan & S. Lawless (dir.), *User Modeling, Adaptation and Personalization: 23rd International Conference, UMAP 2015, Dublin, Ireland, June 29 -- July 3, 2015*, Cham: Springer International Publishing, p. 56-67.
- Clements, M. A. (1982). Careless Errors Made by Sixth-Grade Children on Written Mathematical Tasks. *Journal for Research in Mathematics Education*, 13(2), p. 136-144. doi: 10.2307/748360.
- Conati, C. (2002). Probabilistic assessment of user's emotions in educational games. *Applied Artificial Intelligence*, 16(7-8), p. 555-575. doi: 10.1080/08839510290030390.
- Conati, C., Jacques, N. et Muir, M. (2013). Understanding Attention to Adaptive Hints in Educational Games: An Eye-Tracking Study. *International Journal of Artificial Intelligence in Education*, 23(1), p. 136-161.
- Conati, C. et Maclaren, H. (2009). Empirically building and evaluating a probabilistic model of user affect. *User Modeling and User-Adapted Interaction*, 19(3), p. 267-303. doi: 10.1007/s11257-009-9062-8.
- Conn, H. O. (1977). Trailmaking and number-connection tests in the assessment of mental state in portal systemic encephalopathy. *The American Journal of Digestive Disease*, 22(6), p. 541-550.
- D'Mello, S., Olney, A., Williams, C. et Hays, P. (2012). Gaze tutor: A gaze-reactive intelligent tutoring system. *International Journal of Human-Computer Studies*, 70(5), p. 377-398. doi: <http://dx.doi.org/10.1016/j.ijhcs.2012.01.004>.
- D'Mello, S. K. et Kory, J. (2015). A Review and Meta-Analysis of Multimodal Affect Detection Systems. *ACM Computers Survey*, 47(3), p. 1-36. doi: 10.1145/2682899.
- D'Mello, S. D., Taylor, R. S., Tapp, K., King, B. et Graesser, A. C. (2007). Posture as a Predictor of Learners' Affective Engagement: Boredom and Flow. *Annual AERA Meeting*. Chicago, Illinois, USA.
- Damasio, A. (1995). *L'erreur de Descartes : La raison des émotions*. Odile Jacob (Eds.), 396 p.

- Demos, J. N. (2005). *Getting Started with Neurofeedback*. WW Norton, 304 p.
- Dempsey, J. V., Rasmussen, K. et Lucassen, B. (1994). Instructional Gaming: Implications for Instructional Technology. *Annual Meeting of the Association for Educational Communications and Technology*. Nashville, TN.
- Dempster, A., Laird, P. et Rubin, N. M. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society*, 39(1), p. 1-38.
- Derbali, L. et Frasson, C. (2010). *Prediction of Players Motivational States Using Electrophysiological Measures during Serious Game Play*. 10th IEEE International Conference on Advanced Learning Technologies, 5-7 July, Sousse, Tunisia. doi: 10.1109/ICALT.2010.143.
- Derbali, L. et Frasson, C. (2011). *Physiological evaluation of attention getting strategies during serious game play*. The 15th international conference on Artificial intelligence in education, Auckland, New Zealand. Springer.
- Derbali, L. et Frasson, C. (2012). *Exploring the effects of prior video-game experience on learner's motivation during interactions with heapmotiv*. Proceedings of the 11th international conference on Intelligent Tutoring Systems, Chania, Crete, Greece. doi: 10.1007/978-3-642-30950-2_17.
- Derbali, L., Ghali, R. et Frasson, C. (2013). *Assessing Motivational Strategies in Serious Games Using Hidden Markov Models*. The Twenty-Sixth International FLAIRS Conference, May 22-24, St. Pete Beach, Florida, USA.
- Didehbani, N., Allen, T., Kandalaf, M., Krawczyk, D. et Chapman, S. (2016). Virtual Reality Social Cognition Training for children with high functioning autism. *Computers in Human Behavior*, 62, p. 703-711. doi: <http://dx.doi.org/10.1016/j.chb.2016.04.033>
- Dollard, J., Miller, N. E., Doob, L. W., Mowrer, O. H. et Sears, R. R. (1939). *Frustration and aggression*. New Haven, CT, US: Yale University Press.
- Dunwell, I., Petridis, P., Hendrix, M., Arnab, S., Smadi, M. A. et Guetl, C. (2012). *Guiding Intuitive Learning in Serious Games: An Achievement-Based Approach to Externalized Feedback and Assessment*. The Sixth International Conference on Complex, Intelligent, and Software Intensive Systems, 4-6 July. doi: 10.1109/CISIS.2012.205.
- Eggemeier, F. T., Glenn, F., Arthur, F. et Damos, D. (1991). Workload assessment in multi-task environments, Multiple task performance, London: Taylor & Francis, p. 207-216.

- Ekman, P. (1970). *Universal facial expressions of emotion*. Published in California Mental Health Research Digest, 8(4).
- Ekman, P. (1999). Basic Emotions. Dans L. John Wiley & Sons (dir.), *Handbook of Cognition and Emotion*, p. 138-198.
- Elliot, A. J. et Pekrun, R. (2007). Emotion in the Hierarchical Model of Approach-Avoidance Achievement Motivation. Dans P. Schutz, Pekrun, R. (dir.), *Emotion in Education*, London: Elsevier.
- Frankish, K. (2010). Dual-Process and Dual-System Theories of Reasoning. *Philosophy Compass*, 5(10), p. 914-926. doi: 10.1111/j.1747-9991.2010.00330.x.
- Freeman, D., Evans, N., Lister, R., Antley, A., Dunn, G. et Slater, M. (2014). Height, social comparison, and paranoia: An immersive virtual reality experimental study. *Psychiatry Research*, 218(3), p. 348-352. doi: <http://dx.doi.org/10.1016/j.psychres.2013.12.014>.
- Freina, L. et Ott, M. (2015). *A Literature Review on Immersive Virtual Reality in Education: State Of The Art and Perspectives*. The ELSE Conference, Bucharest, April 2015.
- Fuster, J. M. (2002). Frontal lobe and cognitive development. *J Neurocytol*, 31(3-5), p. 373-385.
- Galan, F. et Beal, C. (2012). *EEG estimates of engagement and cognitive workload predict math problem solving outcomes*. Proceedings of the 20th international conference on User Modeling, Adaptation, and Personalization, Montreal, Canada. doi: 10.1007/978-3-642-31454-4_5.
- Garris, R., Ahlers, R. et Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & gaming*, 33(4), p. 441-467.
- Ghali, R., Chaouachi, M., Derbali, L. et Frasson, C. (2014). *Motivational Strategies to Support Engagement of Learners in Serious Games*. The 6th International ICAART Conference. 6-8 March, ESEO, Angers, France.
- Ghali, R., Frasson, C. et Ouellet, S. (2016). Towards Real Time Detection of Learners' Need of Help in Serious Games. The international *Flairs 2016*. Key Largo, Florida, USA.
- Ghali, R., Ouellet, S. et Frasson, C. (2015). *LewiSpace: An Educational Puzzle Game Combined with a Multimodal Machine Learning Environment*. Joint German/Austrian Conference on Artificial Intelligence (Künstliche Intelligenz: KI), 21-25 September, Dresden, Germany.

- Ghali, R., Ouellet, S. et Frasson, C. (2016a). LewiSpace: An Exploratory Study with a Machine Learning Model in an Educational Game. *Journal of Education and Training Studies*, 4(1), p. 192-201.
- Ghali, R., Ouellet, S. et Frasson, C. (2016b). Using Electrophysiological Features in Cognitive Tasks: An Empirical Study. *International Journal of Information and Education Technology*, 6(8), p. 584-590.
- Ghergulescu, I. et Muntean, C. H. (2014). A Novel Sensor-Based Methodology for Learner's Motivation Analysis in Game-Based Learning. *Interacting with Computers*, 26(4), p. 305-320. doi: 10.1093/iwc/iwu013.
- Gigrenzer, G. (2007). *Gut Feeling: The intelligence of unconsciousness*. Penguin Books; Reprint edition (June 24, 2008).
- Gobert, J. D., Baker, R. S. et Wixon, M. B. (2015). Operationalizing and Detecting Disengagement Within Online Science Microworlds. *Educational Psychologist*, 50(1), p. 43-57. doi: 10.1080/00461520.2014.999919.
- Godefroy, O. (2008). *Fonctions exécutives et pathologies neurologiques et psychiatriques: évaluation en pratique clinique*. De Boeck Université. 312 p.
- Green, J. L., Camilli, G. et Elmore, P. (2012). *Handbook of complementary methods in education research*. United States of America: Routledge; 3 edition.
- Ha, E. Y., Rowe, J. P., Mott, B. W. et Lester, J. C. (2011). *Goal recognition with Markov logic networks for player-adaptive games*. Proceedings of the Seventh AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment, Stanford, California, USA.
- Haapalainen, E., Kim, S., Forlizzi, J. F. et Dey, A. K. (2010). *Psycho-physiological measures for assessing cognitive load*. Proceedings of the 12th ACM international conference on Ubiquitous computing, Copenhagen, Denmark. doi: 10.1145/1864349.1864395.
- Harrison, B., Ware, S. G., Fendt, M. W. et Roberts, D. L. (2015). A Survey and Analysis of Techniques for Player Behavior Prediction in Massively Multiplayer Online Role-Playing Games. *IEEE Transactions on Emerging Topics in Computing*, 3(2), p. 260-274. doi: 10.1109/TETC.2014.2360463.
- Hastie, T., Tibshirani, R. et Friedman, J. (2012). *The Elements of Statistical Learning: second edition*. Springer.

- Hayes, P. J. (1978). *The naive physics manifesto*. Université de Genève, Institut pour les études séantiques et cognitives.
- Hays, R. (2005). The Effectiveness of Instructional Games: A Literature Review and Discussion. Technical report.
<http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA441935>.
- Hershkovitz, A., de Baker, R. S. J., Gobert, J., Wixon, M. et Sao Pedro, M. (2013). Discovery With Models: A Case Study on Carelessness in Computer-Based Science Inquiry. *American Behavioral Scientist*. doi: 10.1177/0002764213479365
- Hocine, N., Gouaich, A., Cerri, S. A., Mottet, D., Froger, J. et Laffont, I. (2015). Adaptation in serious games for upper-limb rehabilitation: An approach to improve training outcomes. *User Modeling and User-Adapted Interaction*, 25(1), p. 65-98. doi: 10.1007/s11257-015-9154-6.
- Hocine, N., Gouaich, A., Loreto, I. et Abrouk, L. (2011). techniques d'adaptation dans les jeux ludiques et sérieux. *Revue des Sciences et Technologies de l'Information*, 25(2), p. 253-280.
- Holyoak, K. J. et Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science*, 13(3), p. 295-355. doi: [http://dx.doi.org/10.1016/0364-0213\(89\)90016-5](http://dx.doi.org/10.1016/0364-0213(89)90016-5).
- Ip, H. H. S., Wong, S. W. L., Chan, D. F. Y., Byrne, J., Li, C., Yuan, V. S. N., Lau, K., Wong, J. Y. W. (2016). Virtual Reality Enabled Training for Social Adaptation in Inclusive Education Settings for School-Aged Children with Autism Spectrum Disorder (ASD). Dans S. K. S. Cheung, L.-f. Kwok, J. Shang, A. Wang & R. Kwan (dir.), *Blended Learning: Aligning Theory with Practices : 9th International Conference, ICBL 2016, Beijing, China, July 19-21, 2016*. Cham: Springer International Publishing, p. 94-102.
- Isen, A. M. (2005). Positive Affect. *Handbook of Cognition and Emotion* : John Wiley & Sons, Ltd. p. 521-539.
- Jackson, G. T., Dempsey, K. B. et McNamara, D. S. (2012). Game-Based Practice in a Reading Strategy Tutoring System: Showdown in iSTART-ME. Dans H. Reinders (dir.), *Digital Games in Language Learning and Teaching*. London: Palgrave Macmillan UK, p. 115-138.
- Jaques, N., Conati, C., Harley, J. M. et Azevedo, R. (2014). *Predicting affect from gaze data during interaction with an intelligent tutoring system*. International Conference on Intelligent Tutoring Systems. 5-9 June, Honolulu, Hawaii, USA.

- Jaume, L. (1983). La Théorie de la personne fictive; Dans le Léviathan De Hobbes. *Revue française de science politique*, 33(6), p. 1009-1035.
- Johnson, W. L. et Wu, S. (2008). Assessing Aptitude for Learning with a Serious Game for Foreign Language and Culture. Dans B. P. Woolf, E. Aïmeur, R. Nkambou & S. Lajoie (dir.), *Intelligent Tutoring Systems: 9th International Conference, ITS 2008, Montreal, Canada, June 23-2*. Berlin, Heidelberg: Springer Berlin Heidelberg, p.520-529.
- Jraidi, I., Chaouachi, M. et Frasson, C. (2013). *A dynamic multimodal approach for assessing learners' interaction experience*. Proceedings of the 15th ACM on International conference on multimodal interaction, Sydney, Australia. doi: 10.1145/2522848.2522896.
- Jraidi, I. et Frasson, C. (2010). *Subliminally enhancing self-esteem: impact on learner performance and affective state*. International Conference on Intelligent Tutoring Systems, June 14-18, Pittsburgh, PA, USA.
- Kahneman, D. (2002). Maps of bounded rationality: A perspective on intuitive judgment and choice. *Nobel prize lecture*, 8, p. 351-401.
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.98.9867&rep=rep1&type=pdf>
- Kai, s., Paquette, L., Baker, R. S., Bosch, N., D'mello, S. K., Ocumpaugh, J., Shute, V., Ventura, M. (2015). *Comparison of face-based and Interaction-based affect detectors in Physics Playground*. International Conference on Educational Data Mining, Madrid, Spain.
- Kapoor, A. et Picard, R. W. (2005). *Multimodal affect recognition in learning environments*. Proceedings of the 13th annual ACM international conference on Multimedia, Hilton, Singapore. doi: 10.1145/1101149.1101300.
- Keller, J. M. (1987). Development and use of the ARCS model of instructional design. *Journal of instructional development*, 10(3), p. 2. doi: 10.1007/bf02905780.
- Kim, J. M., Randall W. Hill, J., Durlach, P. J., Lane, H. C., Forbell, E., Core, M., Marsella, S., Pynadath, D., Hart, J. (2009). BiLAT: A Game-Based Environment for Practicing Negotiation in a Cultural Context. *International Journal Artificial Intelligence Education*, 19(3), p. 289-308.
- Klem, G. H., Luders, H. O., Jasper, H. H. et Elger, C. (1999). The ten-twenty electrode system of the International Federation. The International Federation of Clinical Neurophysiology. *Electroencephalography and Clinical Neurophysiology Supplement*, 52, p. 3-6.

- Koedinger , K. R., Aleven , V., Heffernan , N., McLaren , B. et Hockenberry, M. (2004). *Opening the Door to Non-Programmers: Authoring Intelligent Tutor Behavior by Demonstration*. Intelligent Tutoring System. August 30-September 3, Maceió, Alagoas, Brazil.
- Kolb, D. A. (1984). *The nature of intuition and its neglect in research on judgments and decision making, Intuition in judgment and decision making*, N. J.: Prentice Hall.
- Krathwohl, D. R. (2002). A Revision of Bloom's Taxonomy: An Overview. *Theory Into Practice*, 41(4), p. 212-218.
- Krishnamurti, J. et Bohm, D. (1987). *Le temps aboli*. Rocher.
- Kruglanski, A. W. et Gigerenzer, G. (2011). Intuitive and deliberate judgments are based on common principles. *Psychological review*, 118(1), p. 97.
- Lang, P. J. (1995). The emotion probe: Studies of motivation and attention. *American Psychologist*, 50(5), p. 372-385. doi: 10.1037/0003-066X.50.5.372
- Lee, S. Y., Mott, B. W. et Lester, J. C. (2012). Real-Time Narrative-Centered Tutorial Planning for Story-Based Learning. Dans S. A. Cerri, W. J. Clancey, G. Papadourakis & K. Panourgia (dir.), *Intelligent Tutoring Systems: 11th International Conference, ITS 2012, Chania, Crete, Greece, June 14-18*. Berlin, Heidelberg: Springer Berlin Heidelberg, p.467-481.
- Lester, J., Ha, E., Lee, S., Mott, B., Rowe, J. et Sabourin, J. (2013). Serious Games Get Smart: Intelligent Game-Based Learning Environments. *AI MAGAZINE* 34(4).
- Lester, J., Mott, B., Rowe, J. et Sabourin, J. (2013). Modeling Learner Affect with Theoretically Grounded Dynamic Bayesian Networks. *Design Recommendations for Adaptive Intelligent Tutoring Systems Learner Modeling*, p. 1-10.
- Lester, J. C., Spires, H. A., Nietfeld, J. L., Minogue, J., Mott, B. W. et Lobene, E. V. (2014). Designing game-based learning environments for elementary science education: A narrative-centered learning perspective. *Information Sciences*, 264, p. 4-18. doi: <http://dx.doi.org/10.1016/j.ins.2013.09.005>
- Lewinski, P., den Uyl, T. M. et Butler, C. (2014). Automated facial coding: Validation of basic emotions and FACS AUs in FaceReader, 7, p. 227-236. US: Educational Publishing Foundation.

- Maydeu-Olivares, A. et D'Zurilla, T. J. (1995). A Factor Analysis of the Social Problem-Solving Inventory using Polychoric Correlations. *European Journal of Psychological Assessment*, 11(2), p. 98-107. doi: 10.1027/1015-5759.11.2.98.
- McNamara, D. et Jackson, G. (2010). Intelligent Tutoring and Games (ITAG) (*Gaming for Classroom-Based Learning: Digital Role Playing as a Motivator of Study*: IGI Global, p.44-65.
- McNamara, D. S., O'Reilly, T., Rowe, M., Boonthum, C. et Levinstein, I. (2007) iSTART: A Web-based tutor that teaches self-explanation and metacognitive reading strategies. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- McQuiggan, S. W. et Lester, J. C. (2006). Diagnosing Self-efficacy in Intelligent Tutoring Systems: An Empirical Study. Dans M. Ikeda, K. D. Ashley & T.-W. Chan (dir.), *Intelligent Tutoring Systems: 8th International Conference, ITS 2006, Jhongli, Taiwan, June 26-30*. Berlin, Heidelberg: Springer, p.565-574.
- Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W. et Davis, T. J. (2014). Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis. *Computers & Education*, 70, p. 29-40. doi: <http://dx.doi.org/10.1016/j.compedu.2013.07.033>
- Miller, E. et Wallis, J. (2009). Executive function and higher-order cognition: definition and neural substrates. *Encyclopedia of neuroscience*, 4, p. 99-104.
- Miller, K. J., Adair, B. S., Pearce, A. J., Said, C. M., Ozanne, E. et Morris, M. M. (2014). Effectiveness and feasibility of virtual reality and gaming system use at home by older adults for enabling physical activity to improve health-related domains: a systematic review. *Age Ageing*, 43(2), p. 188-195. doi: 10.1093/ageing/aft194.
- Min, W., Ha, E., Rowe, J. P., Mott, B. W. et Lester, J. C. (2014). *Deep Learning-Based Goal Recognition in Open-Ended Digital Games*. The 12th Annual AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment.
- Min, W., Mott, B., Rowe, J., Liu, B. et Lester, J. (2016). *Player Goal Recognition in Open-World Digital Games with Long Short-Term Memory Networks*. The 25th International Joint Conference on Artificial Intelligence, 9-15 July, New York, USA.
- Mitchell, A., Savill-Smith, C. et Britain, G. (2011). The use of computer and video games for learning : a review of the literature. <http://dera.ioe.ac.uk/5270/>.
- Mohammad, Y. et Nishida, T. (2010). Using physiological signals to detect natural interactive behavior. *Applied Intelligence*, 33(1), p. 79-92. doi: 10.1007/s10489-010-0241-4.

- Mostow, J., Chang, K.-M. et Nelson, J. (2011). *Toward exploiting EEG input in a reading tutor*. Proceedings of the 15th international conference on Artificial intelligence in education, Auckland, New Zealand.
- Mott, B. W. et Lester, J. C. (2006). *U-director: a decision-theoretic narrative planning architecture for storytelling environments*. Proceedings of the fifth international joint conference on Autonomous agents and multiagent systems, Hakodate, Japan. doi: 10.1145/1160633.1160808.
- Nakasone, A., Prendinger, H. et Ishizuka, M. (2005). *Emotion recognition from electromyography and skin conductance*. The Fifth International Workshop on Biosignal Interpretation. September 6-8, Tokyo, Japan.
- Naser, S. S. A. (2012). Predicting learners performance using artificial neural networks in linear programming intelligent tutoring system. *International Journal of Artificial Intelligence and Applications*, 3(2), p. 65.
- Niedermeyer, E., Schomer, D. L. et Lopes da Silva, F. H. (2011). *Niedermeyer's electroencephalography : basic principles, clinical applications, and related fields*. Philadelphia: Wolters Kluwer/Lippincott Williams & Wilkins Health.
- Nourbakhsh, N., Wang, Y., Chen, F. et Calvo, R. A. (2012). *Using galvanic skin response for cognitive load measurement in arithmetic and reading tasks*. Proceedings of the 24th Australian Computer-Human Interaction Conference, Melbourne, Australia. doi: 10.1145/2414536.2414602
- Ortony, A. (1988). *The cognitive structure of emotions*. Cambridge University Press, New York.
- Ouellet, S. (2016). *Environnement d'adaptation pour un jeu sérieux*. (Université de Montréal, Canada). <https://papyrus.bib.umontreal.ca/xmlui/handle/1866/16179>
- Pan, Z., Cheok, A. D., Yang, H., Zhu, J. et Shi, J. (2006). Virtual reality and mixed reality for virtual learning environments. *Computers & Graphics*, 30(1), p. 20-28. doi: <http://dx.doi.org/10.1016/j.cag.2005.10.004>
- Pedregosa, F., Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel et al. (2011). Scikit-learn: Machine Learning in Python. *JMLR*, 12, p. 2825–2830.
- Piaget, J. (1936). *La naissance de l'intelligence chez l'enfant*. http://www.fondationjeanpiaget.ch/fjp/site/textes/ve/jp36_ndi_avpropos_intro.pdf.

- Pittalis, M. et Christou, C. (2010). Types of reasoning in 3D geometry thinking and their relation with spatial ability. *Educational Studies in Mathematics*, 75(2), p. 191-212. doi: 10.1007/s10649-010-9251-8.
- Pope, A. T., Bogart, E. H. et Bartolome, D. S. (1995). Biocybernetic system evaluates indices of operator engagement in automated task. *Biol Psychol*, 40(1-2), p. 187-195.
- Prendinger, H. et Ishizuka, M. (2005). The empathic companion: a character-based interface that addresses users' affective states. *Applied Artificial Intelligence*, 19(3-4), p. 267-285. doi: 10.1080/08839510590910174.
- Prensky, M. (2007). *Digital Game-Based Learning* (2007^e éd., p. 464 p.): Paragon House.
- Price, M., Yuen, E. K., Goetter, E. M., Herbert, J. D., Forman, E. M., Acierno, R. et Ruggiero, K. J. (2014). mHealth: a mechanism to deliver more accessible, more effective mental health care. *Clinical Psychology & Psychotherapy*, 21(5), p. 427-436. doi: 10.1002/cpp.1855.
- Prinzel, L. J., Freeman, F. G., Scerbo, M. W., Mikulka, P. J. et Pope, A. T. (2000). A Closed-Loop System for Examining Psychophysiological Measures for Adaptive Task Allocation. *The International Journal of Aviation Psychology*, 10(4), p. 393-410. doi: 10.1207/S15327108IJAP1004_6.
- Ramesh, A., Goldwasser, D., Huang, B., Hal Daume, I. et Getoor, L. (2014). *Uncovering hidden engagement patterns for predicting learner performance in MOOCs*. Proceedings of the first ACM conference on Learning @ scale conference, Atlanta, Georgia, USA. doi: 10.1145/2556325.2567857.
- Rowe, J., Mott, B. et Lester, J. (2014). *Optimizing Player Experience in Interactive Narrative Planning: A Modular Reinforcement Learning Approach*. Proceedings of the Tenth International Conference on Artificial Intelligence and Interactive Digital Entertainment,, Raleigh, North Carolina.
- Rowe, J., Mott, B., McQuiggan, S., Robison, J., Lee, S. et Lester, J. (2009). *Crystal island: A narrative-centered learning environment for eighth grade microbiology*. Communication présentée Workshop on Intelligent Educational Games at the 14th International Conference on Artificial Intelligence in Education, Brighton, UK.
- Rowe, J. P. et Lester, J. C. (2010). *Modeling user knowledge with dynamic Bayesian networks in interactive narrative environments*. Proceedings of the Sixth AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment, Stanford, California, USA.

- Rowe, J. P., Shores, L. R., Mott, B. W. et Lester, J. C. (2011). Integrating learning, problem solving, and engagement in narrative-centered learning environments. *International Journal of Artificial Intelligence Ed.*, 21(1-2), p. 115-133. doi: 10.3233/jai-2011-019.
- Sabourin, J., Mott, B. et Lester, J. C. (2011). *Modeling learner affect with theoretically grounded dynamic bayesian networks*. Proceedings of the 4th international conference on Affective computing and intelligent interaction, Memphis, TN, USA.
- Sabourin, J., Rowe, J. P., Mott, B. W. et Lester, J. C. (2011). When Off-Task is On-Task: The Affective Role of Off-Task Behavior in Narrative-Centered Learning Environments. Dans G. Biswas, S. Bull, J. Kay & A. Mitrovic (dir.), *Artificial Intelligence in Education: 15th International Conference, AIED 2011, Auckland, New Zealand*. Berlin, Heidelberg: Springer, p. 534-536.
- Sabourin, J. L. et Lester, J. C. (2014). Affect and Engagement in Game-Based Learning Environments. *IEEE Transactions on Affective Computing*, 5(1), p. 45-56. doi: 10.1109/T-AFFC.2013.27.
- San Pedro, M. O., Baker, R. J. et Rodrigo, M. M. (2011). *Detecting carelessness through contextual estimation of slip probabilities among students using an intelligent tutor for mathematics*. Proceedings of the 15th international conference on Artificial intelligence in education, Auckland, New Zealand.
- Sawyer, B. et Rejeski, D. (2002). Serious games: Improving public policy through game-based learning and simulation.
- Setz, C., Arnrich, B., Schumm, J., Marca, R. L., Troster, G., Ehlert, U. (2010). Discriminating stress from cognitive load using a wearable EDA device. *IEEE Transactions on Information Technology in Biomedicine.*, 14(2), p. 410-417. doi: 10.1109/titb.2009.2036164.
- Shute, V., Ventura, M. et Kim, Y. J. (2013). Assessment and Learning of Qualitative Physics in Newton's Playground. *The journal of Educational research*, 29, p. 579-582.
- Sottolare, R. A. (2014). Using Learner Data to Influence Performance during Adaptive Tutoring Experiences. Dans D. D. Schmorow & C. M. Fidopiastis (dir.), *Foundations of Augmented Cognition. Advancing Human Performance and Decision-Making through Adaptive Systems: 8th International Conference, AC 2014, Held as Part of HCI International 2014, Heraklion, Crete, Greece, June 22-27*. Cham: Springer International Publishing, p.265-275.
- Stevens, R. H., Galloway, T. et Berka, C. (2007a). Exploring Neural Trajectories of Scientific Problem Solving Skill Acquisition. Dans D. D. Schmorow & L. M. Reeves (dir.),

Foundations of Augmented Cognition: Third International Conference, FAC 2007, Held as Part of HCI International 2007, Beijing, China, July 22-27. Berlin, Heidelberg: Springer, p.400-408.

Stevens, R. H., Galloway, T. et Berka, C. (2007b). Integrating Innovative Neuro-Educational Technologies (I-Net) into K-12 Science Classrooms. *Augmented Cognition: Past, Present, & Future*, 16, p. 47-56.

Teplan, M. (2002). Fundamentals of EEG measurement. *Measurement Science Review*, 2(2). <http://www.measurement.sk/2002/S2/Teplan.pdf>.

Vezneva, M. (2011). *Développement du raisonnement analogique: rôle de la composante exécutive d'inhibition.* (Université de Bourgogne, France).

Viola, P. et Jones, M. (2001). *Rapid object detection using a boosted cascade of simple features.* Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, 8-14 December 2001, Kauai, HI, USA. doi: 10.1109/CVPR.2001.990517.

Vogel, J. J., Vogel, D. S., Cannon, C. A., Bowers, K. M. et Wright, M. (2006). Computer Gaming and Interactive Simulations for Learning: A Meta-Analysis. *Journal of Educational Computing Research*, 34(3), p. 229–243.

Vuilleumier, P. (2005). How brains beware: neural mechanisms of emotional attention. *Trends Cogn Sci*, 9(12), p. 585-594. doi: 10.1016/j.tics.2005.10.011.

Vygotsky (1934). *Pensée et langage.* https://www.scienceshumaines.com/lev-vygotski-1896-1934-pensee-et-langage_fr_9754.html.

Wilson, G. F. (2004). An analysis of mental workload in pilots during flight using multiple electrophysiological measures. *International Journal of Aviation in Psychology, Chapter 12*, p. 3-18.

Wislow, B. (2006). *Affective learning companions : strategies for empathetic agents with real-time multimodal affective sensing to foster meta-cognitive and meta-affective approaches to learning, motivation, and perseverance.* (Massachusetts Institute of Technology). <http://hdl.handle.net/1721.1/37404>

Wixon, M., Baker, R. S. J. d., Gobert, J. D., Ocumpaugh, J. et Bachmann, M. (2012). WTF? Detecting Students Who Are Conducting Inquiry Without Thinking Fastidiously. Dans J. Masthoff, B. Mobasher, M. C. Desmarais & R. Nkambou (dir.), *User Modeling, Adaptation, and Personalization: 20th International Conference, UMAP 2012, Montreal, Canada, July 16-20.* Berlin, Heidelberg: Springer, p. 286-296.

Wood, H. et Wood, D. (1999). Help seeking, learning and contingent tutoring. *Computers & Education*, 33(2-3), p. 153-169. doi: [http://dx.doi.org/10.1016/S0360-1315\(99\)00030-5](http://dx.doi.org/10.1016/S0360-1315(99)00030-5)

Yang-Ting, J. (2011). Une application sérieuse des jeux vidéo (*Gamentrepreneurs*, 134p.).

Yates, M., Kelemen, A. et Sik Lanyi, C. (2016). Virtual reality gaming in the rehabilitation of the upper extremities post-stroke. *Brain Inj*, 30(7), p. 855-863. doi: 10.3109/02699052.2016.1144146

Zyda, M. (2005). From Visual Simulation to Virtual Reality to Games. *Computer*, 38(9), p. 25-32. doi: 10.1109/mc.2005.297.

Publications

Articles de Journaux

- **Ghali**, R., Ouellet, S., Frasson, C., 2016. "Using Electrophysiological Features in Cognitive Tasks: an Empirical Study", *International Journal of Information and Education Technology*, vol. 6, no. 8, August 2016.
- **Ghali**, R., Ouellet, S., Frasson, C., 2016. "LewiSpace: An Exploratory Study with a Machine Learning Model in an Educational Game". *Journal of Education and Training Studies*, vol. 4, no. 1, 2016.
- **Ghali**, R., Ouellet, S., Frasson, C., 2015. "Classification and regression of learner's scores in Logic environment". *Journal of Education and Training Studies*, vol. 3, no. 4, July 2015.
- **Ghali**, R., Frasson, C., 2015. "Stratégies Émotionnelles pour l'apprentissage d'une langue étrangère". *Revue le Langage et l'homme*, vol. I, no.1, Juin 2015.

Articles de Conférences

- **Ghali**, R., Ben Abdessalem, H., Frasson, C., 2017. Improving Intuitive Reasoning through Assistance Strategies in a Virtual Reality Game, Full paper, Flairs 2017, the 30th International FLAIRS Conference, Marco Island, Florida, USA.
- Benlamine, M. S., Villata, S., **Ghali**, R., Frasson, C., Gandon, F., Cabrio, E., 2017. "Persuasive Argumentation and Emotions: an Empirical Evaluation with Users", *Proceeding in International HCI 2017, the 19th International Conference on Human-Computer Interaction*, Vancouver, 14-19 July 2017.
- **Ghali**, R., Frasson, C., Ouellet, S., 2016. "Using Electroencephalogram to Track Learner's Reasoning in Serious Games", *ITS 2016, The 13th International Conference on Tutoring Systems*, 6-10 June, Zagreb, Croatia.
- **Ghali**, R., Ouellet, S., Frasson, C., 2016. "Towards Real Time Detection of Learners' Need of Help in Serious Games", *the 29th international Flairs conference 2016*, 16-18 May, Key Largo, Florida, USA.

- **Ghali, R., Ouellet, S., Frasson, C.** 2015. "LewiSpace: an educational Puzzle Game with a Multimodal Machine Learning Environment". KI 2015: the 38th German Conference on Artificial intelligence, short paper, Dresden, Germany, September 21-25, 2015.
- **Ghali, R., Ouellet, S., Frasson, C.** 2015. "Using Electrophysiological Features in Cognitive Tasks: An Empirical Study", ICEIT (International Conference on Educational and Information Technology, Firenze, Italy, March 2015. **Best paper Award.**
- **Ghali, R., Chaouachi, M., Derbali, L., Frasson, C.** 2014. "Motivational Strategies to Support Engagement of Learners in Serious Games", the 6th International ICAART Conference, March 2014.
- **Derbali, L., Ghali, R., Frasson, C.** 2013. "Assessing motivational strategies in Serious Games using Hidden Markov Models". FLAIRS 2013, the 26th International FLAIRS Conference, AAAI Press, May 22-24, St. Pete Beach, Florida, USA, 2013.
- **Ghali, R., Frasson, C.** 2010. "Agent émotionnel pour l'apprentissage du vocabulaire d'une langue étrangère", TICE 2010 : 7ème colloque Technologies de l'Information et de la Communication pour l'Enseignement, Nancy, France, 6-8 Décembre, 2010.
- **Ghali, R., Frasson, C.** 2010. "Emotional Strategies for Vocabulary Learning", 10th IEEE International Conference on Advanced Learning Technologies (ICALT), Sousse, Tunisie, 5-7 July, 2010, P.249-253.

Articles de workshops

- **Ghali, R., Frasson, C.** 2014. "A study of learner's mental states in different cognitive tasks", EEG Workshop in the 12th International Conference in Intelligent Tutoring System", June 5-9, 2014, Honolulu, Hawaii, USA.
- **Bilodeau, G.-A., Ghali, R., Desgent, S., Langlois, P., Farah, R., St-Onge, P.-L., Duss, S., Carmant, L.** 2011. "Where is the rat? Tracking in low contrast thermographic images", IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops), Colorado Springs, USA, 20-25 Juin, 2011, P. 55-60.

Posters

- **Ghali, R., Frasson, C.** 2014. "An exploratory study of learner's brain states", the 12 th International ITS Conference, Poster, June 2014, Honolulu, Hawaii, US.