Université de Montréal

Environnements Virtuels Émotionnellement Intelligents

par

Benlamine Mohamed Sahbi

Département d'Informatique et de Recherche Opérationnelle

Faculté des Arts et des Sciences

Thèse présentée

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

en Informatique

Avril, 2019

© Benlamine, 2019

Résumé

Les émotions ont été étudiées sous différents angles dans le domaine de l'interaction homme-machine y compris les systèmes tutoriel intelligents, les réseaux sociaux, les plateformes d’apprentissage en ligne et le e-commerce. Beaucoup d’efforts en informatique affective sont investis pour intégrer la dimension émotionnelle dans les environnements virtuels (tel que les jeux vidéo, les jeux sérieux et les environnements de réalité virtuelle ou de réalité augmenté). Toutefois, les stratégies utilisées dans les jeux sont encore empiriques et se basent sur des modèles psychologiques et sociologiques du joueur : Courbe d’apprentissage, gestion de la difficulté, degré d’efficience dans l’évaluation des performances et de la motivation du joueur. Or cette analyse peut malmener le système dans la mesure où les critères sont parfois trop vagues ou ne représentent pas les réelles compétences du joueur, ni ses vraies difficultés. Étant donné que la stratégie d’intervention est très influencée par la précision de l’analyse et l’évaluation du joueur, de nouveaux moyens sont nécessaires afin d’améliorer les processus décisionnels dans les jeux et d’organiser les stratégies d’adaptation de façon optimale.

Ce travail de recherche vise à construire une nouvelle approche pour l’évaluation et le suivi du joueur. L’approche permet une modélisation du joueur plus efficace et moins intrusive par l’intégration des états mentaux et affectifs obtenus à partir de senseurs physiologiques (signaux cérébraux, Activité électrodermale, …) ou/et instruments optiques (Webcam, traceur de regard, …). Les états affectifs et mentaux tels que les émotions de base (basées sur les expressions faciales), l’état d’engagement, de motivation et d’attention sont les plus visés dans cette recherche. Afin de soutenir l’adaptation dans les jeux, des modèles des émotions et de la motivation du joueur basé sur ces indicateurs mentaux et affectifs, ont été développés. Nous avons implémenté cette approche en développant un système sous forme d’une architecture modulaire qui permet l’adaptation dans les environnements virtuels selon les paramètres affectifs du joueur détectés en temps-réel par des techniques d’intelligence artificielle.

Mots-clés : Jeux vidéo, Reconnaissance des émotions, Motivation, Apprentissage machine, Adaptation dans les jeux, Capteurs physiologiques.

Abstract

Emotions were studied from different angles in the field of human-machine interaction including intelligent tutorial systems, social networks, online learning platforms and e-commerce. Much effort in affective computing are invested to integrate the emotional dimension in virtual environments (such as video games, serious games and virtual reality environments or augmented reality). However, the strategies used in games are still empirical and are based on psychological and sociological models of the player: Learning Curve, trouble management, degree of efficiency in the evaluation of performance and motivation of the player. But this analysis can mislead the system to the extent that the criteria are sometimes too vague and do not represent the actual skills of the player, nor his real difficulties. Since the intervention strategy is influenced by the accuracy of the analysis and evaluation of the player, new ways are needed to improve decision-making in games and organizing adaptation strategies in optimal way.

This research aims to build a new approach to the evaluation and monitoring of the player. The approach enables more effective and less intrusive player modeling through the integration of mental and emotional states obtained from physiological sensors (brain signals, electro-dermal activity, ...) or/and optical instruments (Webcam, eye-tracker, ...). The emotional and mental states such as basic emotions (based on facial expressions), the states of engagement, motivation and attention are the most targeted in this research. In order to support adaptation in games, models of emotions and motivation of the player based on these mental and emotional indicators, have been developed. We have implemented this approach by developing a system in the form of a modular architecture that allows adaptation in virtual environments according to the player's emotional parameters detected in real time by artificial intelligence methods.

Keywords: Video Games, Emotion Recognition, Motivation, Machine Learning, Adaptation in games, Physiological sensors.

Table des matières

Liste des tableaux

Liste des figures

Liste des abréviations

EEG : Électroencéphalogramme

IA : Intelligence Artificielle

IE : Intelligence Émotionnelle

EDA : Activité électrodermale de la peau

GSR: Galvanic Skin Response

HR: Heart Rate

STI : Système Tutoriel Intelligent

IHM : Interaction humain machine

JVEI : Jeu vidéo émotionnellement intelligent

OCC : Modèle d'évaluation cognitive de Orthony, Clore et Collins (1988)

RV : Réalité virtuelle

VR: Virtual Reality

AR: Augmented Reality

SA : Stratégie d'adaptation

EM : État Mental

Att : Attention

Eng : Engagement

Distr : Distraction

DL : Difficulty Level

Du : Duration

ML: Machine Learning

SVM : Support Vector machines

RF : Random Forest

RBF: Radial Basis Function

ROC: Receiver Operator Curves

KNN: K Nearest Neighbors

LR : Linear Regression

ANOVA: ANalysis of Variance

T-test: Student Test

"À ma chère mère Saïda et à l'âme de mon cher père Habib."

Remerciements

Merci à tous ceux et celles qui ont contribué, de près ou de loin, à ce que cette thèse voit le jour. Je tiens à remercier mes deux directeurs de thèse Claude Frasson et Aude Dufresne pour les conseils et leur rigueur scientifique, les encouragements et le soutien apportés tout au long de cette thèse. Je remercie aussi Fabien Gandon le directeur du laboratoire Wimmics Inria pour son aide, ses conseils, sa rigueur scientifique et aussi son hospitalité lors de mon séjour en France. Bien sûr, Je remercie Yan Cyr le directeur de la société BMU pour son soutien financier dans le cadre de ce projet CRSNG et les autres projets du labo et aussi tous ses conseils.

J'ai eu l'occasion de travailler dans des environnements variés, que cela soit parmi les étudiants et professeurs du Laboratoire «Higher Educationnal Research ON Tutoring Systems» (HERON) de l'Université de Montréal, l'équipe «Web-Instrumented Man-Machine Interactions, Communities and Semantics » (Wimmics) de l'Inria à l'université de Nice en France et la société BMU Games à Montréal.

Je tiens aussi à remercier les membres du laboratoire HERON. J'adresse une pensée particulière à Maher Chaouachi, Hamdi Abdessalem, René Doumbouya, Jason Harley, Ramla Ghali et Samira Bouslimi qui ont participé à certains travaux de cette thèse et tous mes collègues au Diro. Je tiens aussi à remercier les membres du laboratoire Wimmics. J'adresse une pensée particulière à Serena Villata et Elena Cabrio qui ont participé à certains travaux de cette thèse. Je tiens à remercier tous les membres de l'équipe de la société BMU qui ont été très coopératifs et enthousiastes avec moi.

Je salue également tous mes amis en Tunisie et à Montréal qui m'ont toujours entouré et qui ont rendu ma vie agréable sans oublier mes colocataires et tous mes collègues.

Je pense enfin à mes parents, frères et sœurs et leurs maris et aussi mes cousins qui m'ont toujours encouragé à persévérer et à continuer jusqu'au bout.

Pour vous tous mille mercis!

Chapitre 1 : Introduction

1.1 Contexte

Avec les avances technologiques dans les domaines de l'interaction humain machine et de l'intelligence artificielle, les environnements virtuels (tel que les jeux vidéo, les discussions en ligne, les environnements de réalité virtuelle, …) deviennent de plus en plus centrés sur l'utilisateur et plus particulièrement sur son état émotionnel et cognitif. Selon plusieurs recherches en psychologie, interaction humain machine (IHM) et Affective Computing (AC), le comportement de l'utilisateur dans un environnement virtuel dépend de son émotion ressentie durant l'interaction, de son état mental (tels que la motivation, le niveau d'engagement et d'attention) et de son expérience subjective (sa performance, sa persuasion, son immersion, sa satisfaction, …), d'où l'émergence d'un nouveau concept qui est l'intelligence émotionnelle dans les environnements virtuels.

L'intelligence émotionnelle (IE) d'un système implique la capacité de comprendre et de gérer les émotions des usagers. L'intelligence émotionnelle d'un individu est non seulement la capacité d'exprimer et de contrôler ses émotions mais aussi la capacité de comprendre, d'interpréter et de réagir aux émotions des autres (Goleman, 2006; Salovey & Mayer, 1990). Les personnes émotionnellement intelligentes sont profondément sensibles à leurs propres sentiments, capables d'exprimer leurs émotions de manière appropriée, de même qu'elles sont empathiques et comprennent ce que ressentent les autres (Mayer & Geher, 1996). L'intégration d'intelligence émotionnelle dans les environnements virtuels devrait permettre de faire progresser non seulement le domaine des jeux pour le divertissement mais aussi les jeux sérieux.

Vu l'importance des émotions dans les interactions entre les humains et aussi entre l'humain et la machine, cette thèse vise à intégrer différents aspects de l'intelligence émotionnelle dans les environnements virtuels afin de déterminer leurs impacts sur la réaction émotionnelle et mentale de l'utilisateur et son expérience subjective (rapporté par des questionnaires sur la qualité de l'interaction). Nous voulons ainsi expérimenter différents aspects de l'intelligence émotionnelle dans plusieurs environnements où il y a interaction humain-humain et humain-machine tels que les discussions en ligne et les jeux commerciaux et sérieux.

Pour cela, notre recherche s'articule autour de trois axes destinés à approfondir les questions de recherche selon plusieurs dimensions :

1- Le premier axe concerne l'**évaluation** et la **reconnaissance** des émotions en situation d'interaction.

2- Le deuxième axe concerne la capacité de **déterminer les éléments qui influencent les émotions, la motivation et l'expérience de l'utilisateur** dans un environnement virtuel.

3- Le troisième axe vise à l'**adaptation émotionnelle** des éléments de jeu dans les environnements virtuels pour améliorer l'expérience de l'utilisateur.

Cette thèse vise à la mise au point de systèmes émotionnels intelligents dans les environnements virtuels. Nous nous intéressons à différents types d'environnements virtuels tels que les discussions en ligne, les jeux vidéo, les environnements de réalité virtuelle, … etc. Dans un premier temps, nous évaluons les émotions dans des sessions de débats en ligne pour étudier si le sujet du débat, la relation entre les arguments et la stratégie d'argumentation dans un débat influencent les émotions des participants et leur changement d'opinion (persuasion). Dans un deuxième temps, nous nous intéressons aux jeux vidéo (commerciaux et sérieux) sur écran et en réalité virtuelle pour étudier si les caractéristiques des joueurs, les évènements dans le jeu et les objectifs de design de la scène de jeu ont un effet sur l'expérience du joueur, ses émotions, sa motivation et aussi la persuasion au sein du jeu (dans le cas de l'interaction avec des personnages non-joueur dans un jeu sérieux).

Les environnements virtuels (et surtout les jeux vidéo et les environnements de réalité virtuelle) intègrent plusieurs caractéristiques essentielles qui favorisent l'intelligence émotionnelle. Parmi ces caractéristiques, nous distinguons les caractéristiques des jeux classées en trois composantes dans le modèle design de jeu DPE (Winn, 2009)- Design, Play, Experience -, qui peuvent susciter chez les utilisateurs différents états émotionnels et mentaux (tels que la surprise, la motivation et l'engagement). Selon Goleman et ses collègues (Goleman, 2006), il existe cinq composantes principales de l'intelligence émotionnelle que les environnements virtuels peuvent intégrer pour devenir émotionnellement intelligents:

- **Conscience de soi** (*Self-awareness*): c'est non seulement la capacité de reconnaître et de comprendre ses propres émotions mais aussi de savoir l'effet de ses propres actions sur les émotions des autres. L'environnement virtuel doit reconnaître les sources des réactions émotionnelles des joueurs et les effets des éléments du jeu sur les émotions et les états mentaux des joueurs. Cette intégration entre le Design du jeu et le modèle du joueur rend le système plus conscient du joueur et efficace pour adapter le jeu afin d'optimiser l'expérience de l'utilisateur en utilisant des techniques d'IA. En utilisant ce système, le concepteur sera capable de reconnaître les forces et les limites de l'environnement virtuel qu'il conçoit, de l'optimiser et d'apprendre à partir des interactions avec les utilisateurs et leurs expériences dans l'environnement.

- **Autorégulation** (*Self-regulation*): c'est la capacité de réguler et de gérer ses émotions et aussi d'influencer les autres de manière bien réfléchie. Ceci peut être intégré dans les environnements virtuels par la modélisation de la réaction de l'utilisateur et les stratégies d'interventions pour atteindre l'objectif visé par l'environnement. Cela signifie par exemple qu'il faut attendre le bon moment, le bon endroit et le bon moyen pour donner un feedback ou un commentaire à l'utilisateur ou bien effectuer un changement du contenu de l'environnement pour s'adapter à l'état émotionnel de l'utilisateur et à sa performance.

- **Motivation**: c'est la capacité d'être motivé par la réponse à ses propre besoins et objectifs internes et aussi par des simples récompenses externes, telles que l'argent, la reconnaissance et les primes. Ceci peut être intégré dans les environnements virtuels par la détermination des orientations motivationnelles soit de l'objectif du joueur (Maîtrise/Performance) et son orientation (Approche/Évitement) dans l'environnement ainsi que l'adaptation des messages et les récompenses en favorisant la compétence et le challenge.

- **Empathie** (*Empathy*): c'est la capacité de comprendre ce que ressentent les autres, de reconnaître leurs états émotionnels selon leurs situations et de savoir en quoi nos actions influencent leurs sentiments et leurs comportements. Si la situation est difficile, on peut intégrer des messages empathiques ou une musique qui reflète l'émotion reconnue.

- **Compétences sociales** (*Social skills*): c'est la capacité de bien interagir avec les autres, ce qui implique, outre la compréhension de ses propres émotions et celles des autres, l'utilisation de ces informations dans les interactions. Les plus importantes compétences sociales sont l'écoute active, la communication verbale, la communication non verbale, le leadership et la persuasion. Celles-ci peuvent être intégrées dans les environnements virtuels par la considération de la dimension sociale avec des personnages non-joueurs ou d'autres joueurs.

L'environnement résultant qui combine ces principes de l'intelligence émotionnelle et les caractéristiques des jeux vidéo peut être appelé *Jeux émotionnellement Intelligents* (JEI). L'idée consiste ainsi à intégrer les principes de l'intelligence émotionnelle en modifiant les caractéristiques des jeux vidéo (Winn, 2009) pour créer des jeux qui s'adaptent aux émotions des joueurs. Dans notre cas, nous allons créer des JEI en s'inspirant des composantes mentionnées de l'intelligence émotionnelle.

1.2 Problématique

Pour faire évoluer l'intelligence artificielle dans les environnements virtuels et les jeux vidéo, il est nécessaire d'y intégrer la dimension émotionnelle en reconnaissant les émotions de l'utilisateur et en adaptant leurs contenus à son état émotionnel et mental détecté. Nous souhaitons dans cette thèse améliorer l'expérience de l'utilisateur en adaptant le contenu de l'environnement virtuel selon son émotion et sa motivation en utilisant des techniques d'apprentissage machine. Le but ultime est d'arriver à construire un *environnement virtuel émotionnellement intelligent* qui contrôle mieux l'interaction avec l'utilisateur pour offrir une expérience optimale que ce soit pour un jeu de divertissement ou pour un jeu sérieux. Face à ce grand défi, nous avons choisi d'aller par étapes. Une première étape est d'évaluer les états émotionnels et mentaux de vraies personnes qui interagissent entre elles avec des arguments dans un environnement de débat en ligne, et d'analyser les stratégies d'argumentation et leurs effets sur le changement d'attitude (persuasion). Une deuxième étape est d'évaluer des environnements de jeux vidéo et de construire un jeu émotionnellement intelligent qui prend en considération les émotions détectées du joueur et qui est capable de s'adapter selon l'état émotionnel du joueur. Nous avons été guidés par les résultats trouvés dans les débats en ligne à l'analyse des réactions émotionnelles et cognitives au sein du jeu émotionnellement intelligent. Les problèmes de recherche ciblés dans cette thèse relèvent des limites que l'on rencontre pour :

a) Analyser des émotions et l'expérience subjective des utilisateurs dans les environnements virtuels
b) Identifier les éléments du jeu liés à l'émotion de l'utilisateur et sa motivation lors de son interaction avec l'environnement virtuel
c) Définir et évaluer des stratégies d'adaptation émotionnelle dans un environnement virtuel de jeu vidéo.

1.2.1 Analyse des émotions et de l'expérience subjective des utilisateurs dans les environnements virtuels

Pour analyser les réactions émotionnellement intelligentes, nous avons d'abord analysé les interactions entre humains dans un environnement de débat en ligne. Dans ce contexte, nous avons cherché à analyser l'effet des arguments et leurs relations (attaque/support) sur les émotions (expressions faciales) et l'état mental (engagement et charge mentale) des participants et aussi leur changement d'attitude (persuasion) à la fin du débat.

Pour les environnements de jeu vidéo et de réalité virtuelle, nous avons cherché à analyser l'effet des évènements dans le jeu (la dynamique du jeu lors de l'interaction du joueur) sur les émotions (expressions faciales) et l'état mental (engagement, motivation, …) du joueur et aussi sur son expérience du jeu rapporté à la fin de la session du jeu (immersion et satisfaction). Conçus pour divertir ou atteindre des objectifs plus «sérieux», les environnements virtuels tels que les jeux, la réalité virtuelle et la réalité augmentée ont le potentiel d'influencer les croyances, les connaissances, les attitudes, les émotions, les capacités cognitives et les comportements des joueurs. Or dans le cas des environnements de réalité virtuelle et de réalité augmentée, nous n'avons pas accès aux expressions faciales de l'utilisateur via la caméra puisqu'il porte un casque de réalité virtuelle ou augmentée qui cache une partie de son visage. On doit donc utiliser d'autres sources de données, telles que l'activité électrodermale (EDA) ou les signaux cérébraux EEG (Électro-encéphalographie), ce qui donne aussi accès aux mesures cognitives (engagement, charge mentale et attention).

1.2.2 Identification des éléments du jeu liés à l’émotion du joueur et sa motivation lors de son interaction avec l’environnement virtuel

Les joueurs en difficulté et qui ont tendance à abandonner le jeu, ont souvent des émotions négatives et des problèmes d’engagement et de motivation. Il sera alors intéressant de trouver les éléments du jeu qui ont influé sur l’émotion et la motivation du joueur. Si on veut mesurer l’effet sur l’émotion dominante du joueur et sa motivation dans une scène de jeu, une question se pose avec acuité « Dans quelle situation se trouve le joueur dans la scène de jeu? ». Pour répondre à cette question, on doit connaître comment décrire la situation planifiée par le concepteur pour chacune des scènes du jeu. Par exemple, le modèle OCC (Andrew Ortony, Clore, & Collins, 1990) permet de formaliser la description de diverse situations de jeu et des réactions émotionnelles qui peuvent en découler, par exemple en cas d’échec provoqué par un autre agent (colère) ou par lui-même (déception, honte). Le recours à des modèles de détection de l’émotion ressentie par le joueur et de sa motivation selon la scène de jeu nous paraît alors très important. Ces modèles sont utiles pour garantir une approche systématique favorisant une séquence logique du développement d'un diagnostic d’émotion dominante du joueur dans la scène de jeu et de sa motivation, ce qui est intéressant pour optimiser la conception du jeu émotionnellement intelligent. Les critères suivants, classés par ordre d'importance, vont nous guider dans le choix de ces modèles à appliquer à la présente recherche :

a. Le modèle doit chercher à détecter l’émotion du joueur et sa motivation selon les objectifs de design dans une scène de jeu (description de la situation dans la scène).
b. Le modèle doit pouvoir identifier les éléments du jeu avec lesquels le joueur interagit dans une scène de jeu (évènements de jeu, narration, …).
c. Le modèle doit s'adresser aux caractéristiques individuelles du joueur (trait de personnalité, style du joueur, …) pour favoriser une intervention engageante adaptée par la suite pour optimiser son expérience dans le jeu.

Plusieurs modèles théoriques (Isbister & Schaffer, 2008; Lazzaro, 2004; Nakamura & Csikszentmihalyi, 2002; Prensky, 2002; Quick, Atkinson, & Lin, 2012) ont été proposés pour modéliser l’émotion du joueur et sa motivation dans un jeu vidéo. L’application de ces modèles dans les études des émotions dans les jeux se limite généralement à modéliser les caractéristiques du joueur et le niveau de difficulté du jeu. Une recherche détaillée des effets affectifs et motivationnels des éléments de jeu est donc importante dans le cadre des jeux sérieux et commerciaux.

1.2.3 Définition et intégration des stratégies d'adaptation émotionnelle dans les jeux vidéo

Le défi dans notre recherche est de concevoir des jeux vidéo émotionnellement intelligents qui puissent s'adapter dynamiquement au comportement émotionnel du joueur durant le jeu. Nous devons intervenir dans la boucle de rétroaction, appelée jouabilité (gameplay[1]**)**, qui existe entre le joueur et les éléments du jeu. Ainsi, si l'état émotionnel de l'utilisateur montre par exemple qu'il est joyeux et sa valence émotionnelle est positive, alors il pourrait rapporter une expérience positive dans l'environnement virtuel à la fin de l'expérience. Si dans le cas contraire, son état émotionnel est négatif (peur ou fâché), il est alors plus opportun d'intervenir avec la bonne stratégie comme le supporter en lui donnant des commentaires d'encouragement ou d'aide ou bien changer le contenu de l'environnement pour modifier son état émotionnel.

Le jeu final est le résultat d'un processus itératif de conception appelé **Game design**. À chaque cycle on obtient un prototype qui passe par une phase de test puis un raffinage pour avoir un prototype avec un *gameplay* plus adapté. Ainsi une analyse erronée ou superficielle de la situation du joueur peut avoir des conséquences directes sur la durée du processus de conception, la qualité des prototypes du jeu et le coût total du projet. Or les critères utilisés dans les tests sont globaux, imprécis et souvent subjectifs, ce qui ralentit tout le processus de conception en exigeant de réaliser plus de prototypes et plus de tests afin d'arriver à un résultat satisfaisant ce qui n'est pas toujours possible si on considère des contraintes économiques qui peuvent arrêter le processus. Ainsi, rendre le *gameplay* plus motivant et adaptatif est une des clefs du bon fonctionnement d'un jeu vidéo. Les jeux vidéo doivent fournir des mesures d'adaptation en temps réel selon l'état émotionnel du joueur pour prévenir la perte de motivation (Przybylski,

---

[1] Dans ce document, nous utilisons le terme anglais « gameplay » au lieu de « jouabilité » puisque « *gameplay »* désigne l'ensemble des règles du jeu alors que « jouabilité » peut être traduite en « playability » qui correspond à la capacité d'une situation précise à intégrer des éléments de jeu.

Rigby, & Ryan, 2010), et soutenir sa performance (Schoenau-Fog, 2011). Il faut ainsi tenir compte de l'évolution de l'émotion du joueur pour optimiser le *gameplay*. Idéalement, le jeu améliorera l'expérience du joueur en faisant appel à des stratégies d'adaptation dans le jeu selon l'état émotionnel du joueur.

1.3 Objectifs de recherche

La recherche dans cette thèse vise à optimiser l'expérience de l'utilisateur par la modélisation de ses réactions affectives et cognitives couplée avec l'ajustement en temps-réel des éléments du jeu selon des stratégies d'adaptation émotionnelle. L'optimisation de l'expérience de l'utilisateur dans un environnement virtuel signifie l'amélioration de certains états émotionnels (tels que la joie et la surprise) et mentaux (tels que l'engagement et la motivation) en concrétisant les trois objectifs suivants:

1- **Analyser les réactions émotionnelles et mentales des utilisateurs dans différents environnements virtuels et la reconnaissance des émotions**.

Notre premier objectif consiste à analyser les réactions émotionnelles et mentales des utilisateurs dans le cas d'interactions entre humains dans un environnement de débat en ligne et d'interactions avec des stimuli visuels (images) et interactifs (jeu vidéo). En effet, l'interaction de l'utilisateur dans un environnement virtuel permet d'analyser son état émotionnel et mental selon les évènements dans l'environnement. Pour cela, nous envisageons de construire des modèles de prédiction des expressions faciales de l'usager et leurs intensités à partir de ses données EEG (électro-encéphalographie) à l'aide d'algorithmes d'apprentissage machine. Ces modèles sont utiles pour l'analyse des expressions faciales dans les jeux de réalité virtuelle ou augmentée à partir de données EEG de l'usager qui porte un casque qui lui cache la face.

2- **Identifier l'émotion du joueur et sa motivation lors de son interaction avec un jeu vidéo et de les lier avec des éléments de jeu**.

Notre deuxième objectif consiste à détecter l'émotion dominante du joueur et sa motivation dans une scène de jeu. La description de la scène est représentée selon les variables situationnelles du modèle OCC (Andrew Ortony et al., 1990). Les modèles de reconnaissance de l'émotion dominante du joueur et sa motivation prennent en considération la structure du jeu et la situation dans la scène et les caractéristiques du joueur.

3- **Définir et expérimenter un outil d'aide à la conception de jeu vidéo émotionnellement intelligent par prototypage et évaluer son efficacité sur l'expérience de l'utilisateur.**
Notre troisième objectif consiste à adapter le jeu vidéo selon l'état émotionnel du joueur en utilisant un outil qui permet de définir et d'appliquer des règles d'adaptation émotionnelle. L'outil proposé permet au concepteur d'adapter le comportement des éléments du jeu selon l'état émotionnel du joueur. C'est une interface visuelle et générique qui peut être intégrée dans n'importe quel jeu vidéo (en Unity 3D). Cet outil applique les règles d'adaptation sur les éléments du jeu au cours de l'interaction avec le joueur selon son état émotionnel.

Pour répondre au **premier objectif**, sur l'analyse affective dans les environnements virtuels (débat en ligne et jeux vidéo), nous avons fait des études empiriques et développé une application temps-réel de reconnaissance des expressions faciales à partir des signaux EEG, nommée « NeuroExpress ». Concernant les études empiriques sur les débats en ligne, la première étude comporte des sessions de débats en ligne entre quatre participants sur des sujets conflictuels. Nous avons analysé la relation entre les états émotionnels, le niveau d'engagement des participants et le nombre d'arguments d'attaque ou de support dans les débats. Dans la deuxième étude, un cinquième participant (un participant persuasif) intervient et donne des arguments selon l'une des stratégies de persuasion (Ethos, Logos, Pathos) pour chacun des débats dans le but de changer l'avis des opposants. Nous avons analysé l'impact de chacune des stratégies de persuasion sur les émotions des participants et leurs engagements.

Nous avons conduit deux études sur l'analyse et la reconnaissance des émotions. Dans la première étude, nous avons développé un jeu « Danger Island », un jeu d'aventure à la troisième personne, qui met le joueur dans une expérience à la fois excitante et amusante. Dans l'interface du jeu, nous avons mis un bouton affichant l'état émotionnel du joueur. À tout moment, si le joueur sent que son état émotionnel a changé, il peut mettre le jeu en pause et cliquer sur ce bouton pour indiquer son nouvel état émotionnel parmi 12 catégories d'émotions d’une liste déroulante qui s’affiche sous le bouton. Le joueur reprend le jeu une fois la nouvelle émotion choisie. Cette méthode d’auto-évaluation au cours du jeu permet au joueur de rapporter l’état émotionnel sans l’interrompre et le désengager. Grâce à cette méthode nous avons procédé à une étude empirique sur la différence entre les « *Gamers »* et les « *non-Gamers »* sur les états émotionnels rapportés et leurs niveaux d’engagement mesurés à partir des données EEG. Dans la deuxième étude on a prédit les expressions faciales des participants à partir de leurs données physiologiques EEG. Pour pouvoir reconnaître les émotions, nous avons utilisé des images de la banque d’images affectives (IAPS[2]) comme stimuli pour les participants et enregistré leurs expressions faciales et leurs signaux cérébraux (EEG). Nous avons entraîné et testé différents algorithmes d'apprentissage machine pour prédire les expressions faciales et des émotions à partir des données EEG. Nous avons effectué la sélection des caractéristiques les plus pertinentes pour maximiser la précision des modèles générés. En appliquant la même méthode décrite dans cette étude sur l’ensemble des données expérimentales sur les images IAPS et le jeu Danger Island nous avons construit des modèles plus robustes qu’on a utilisé dans l’application NeuroExpress.

Pour répondre à notre **deuxième objectif**, qui est d’identifier l’émotion du joueur et sa motivation et de les lier aux éléments du jeu, nous avons développé une solution d'analyse émotionnelle de session de jeu, nommée « Emograph » (Emotional Graph). Nous avons utilisé pour l’expérimentation le jeu commercial sur PC «Outlast » qui est un jeu d’horreur dans lequel le joueur visite des lieux sombres plein de corps et de sang où il ne peut que fuir et se cacher des monstres et des fantômes. Pour vérifier ce but, nous avons adopté une approche multimodale qui combine les mouvements oculaires (à l’aide d’un traceur de regard), les données d'EEG (électroencéphalographie) et des expressions faciales pour annoter les scènes de jeu. Notre système Emograph permet d’accéder plus facilement aux informations à propos de la session de jeu sous forme de graphes de transitions émotionnelles représentant l’émotion dominante d’une scène à l’autre. Le système Emograph permet aussi de prédire l’émotion dominante du joueur et son orientation motivationnelle pour une nouvelle scène de jeu ou bien un nouveau joueur.

[2] IAPS : International Affective Picture System, Librairie standardisée d’images visant à déclencher des émotions et à mesurer l’activation et la valence chez le sujet.

Les scènes de jeu sont décrites par des variables du modèle OCC pour représenter la situation dans la scène du jeu. La prédiction de l'émotion dominante et l'orientation motivationnelle d'une scène du jeu sont faites à partir des caractéristiques du joueur et la description de la scène par les variables OCC. Nous avons conduit deux études avec ce système, la première étude est sur l'analyse émotionnelle dans les environnements de jeux vidéo dans le but de prédire l'émotion dominante du joueur et la deuxième étude est sur l'analyse des orientations motivationnelles dans les scènes de jeu.

Pour répondre à notre **troisième objectif**, nous avons développé une interface intégrable dans n'importe quel projet unity3D permettant la création de règles d'adaptation émotionnelle dans les environnements virtuels, que nous avons nommée « BARGAIN ». Notre interface BARGAIN permet de définir des règles adaptatives en reliant l'état émotionnel mesuré du joueur, le contexte dans le jeu à des modifications des éléments du jeu (feedback, musique, niveau de difficulté, …etc) et de les appliquer dans l'environnement du jeu. Nous avons ensuite intégré BARGAIN dans un environnement de réalité virtuelle pour le développement socio-morale « DilemmasVR » que nous avons développé dans notre labo. Nous avons analysé l'efficacité des règles adaptatives sur la réaction du joueur durant le jeu et son évaluation après le jeu.

1.4 Organisation du document

Ce travail de thèse est organisé comme suit : le chapitre 2 donne un résumé sur l'état de littératures sur l'intelligence émotionnelle, les modèles de design de jeu et la mesure de l'émotion. Nous présentons dans ce chapitre le cadre général de nos recherches, soit rendre les environnements virtuels émotionnellement intelligents par la reconnaissance des émotions et l'adaptation de leurs contenus selon l'état émotionnel de l'usager.

Les chapitres 3 à 9 sont consacrés à nos contributions. Ces chapitres sont présentés sous la forme de six articles de recherche selon les objectifs de la thèse. Les quatre premiers articles visent à répondre au **premier objectif**, sur l'analyse affective dans les environnements virtuels: le premier article a été publié dans la conférence *International Joint Conferences on Artificial Intelligence*, IJCAI 2015, qui est la top-conférence dans le domaine de l'intelligence artificielle.

Le deuxième article a été publié à la conférence *International Florida Artificial Intelligence Research Society conference*, Flairs 2018. Le premier et le deuxième article sont sur les environnements de débats en ligne alors que le troisième t le quatrième article son sur les jeux vidéo. Le troisième article a été publié à la conférence *World Conference on Educational Media and Technology*, Edmedia 2015. Le quatrième article a été publié à la *International conference on Physiological Computing systems*, PhyCs 2016. Les cinquième et sixième articles visent à répondre au **deuxième objectif**, qui est d'identifier l'émotion du joueur et sa motivation et de les lier aux éléments du jeu : Le cinquième article a été publié dans la conférence *International Conference on Intelligent Tutoring Systems*, ITS 2018. Le sixième article a été publié dans la conférence *International Conference on Brain Functions Assessment in Learning*, BFAL 2017. Enfin, le dernier article vise à répondre au **troisième objectif**, sur l'adaptation émotionnelle dans les jeux, il a été soumis le 21 novembre 2018 au journal User Modeling and User Adapted Interaction, UMUAI sous le numéro : UMUI-D-18-00119.

Le chapitre 3 décrit l'étude empirique réalisée dans l'environnement de débats en ligne pour examiner la variation des états émotionnels et mentaux des participants en fonction du nombre de support et d'attaque dans l'argumentation. Le chapitre 4 présente l'étude empirique dans l'environnement de débats en ligne à propos de l'effet des stratégies d'argumentation sur les émotions et l'engagement des participants et leurs changements d'attitudes. Le chapitre 5 introduit le jeu d'aventure « Danger Island » et présente une approche d'auto-évaluation pour rapporter l'état émotionnel du joueur au cours du jeu sans interruption. Le chapitre 6 présente la possibilité de prédire les expressions faciales de l'usager à partir de ses signaux cérébraux (EEG). Nous détaillons notre design expérimental et notre méthode de construction des caractéristiques à partir des données EEG utilisées dans notre application temps-réel de reconnaissance des expressions faciale « NeuroExpress ». Le chapitre 7 présente le système « EmoGraph » pour l'analyse et la prédiction de l'émotion dominante du joueur dans les scènes de jeu grâce à l'utilisation du modèle OCC pour la description de la scène et les caractéristiques du joueur. Le chapitre 8 présente notre approche de classification de la motivation du joueur dans les scènes de jeu par des algorithmes d'apprentissage machine. Le chapitre 9 décrit le système « BARGAIN » notre interface de conception et d'intégration des règles d'adaptation des éléments du jeu selon les états émotionnels du joueur. Il présente également le jeu de réalité virtuelle « DilemmasVR » pour le développement socio-moral. En conclusion, le chapitre 10 résume les contributions et les limites de nos recherches dans cette thèse, et propose des perspectives pour nos futurs travaux.

Chapitre 2 : État de l’art

2.1 Introduction

Ce chapitre présente l'état de la littérature en relation avec nos travaux de recherche. Puisque l'objectif de cette thèse est d'intégrer l'intelligence émotionnelle dans des environnements virtuels par la reconnaissance et d'adaptation à l'état émotionnel de l'usager, nous débutons ce chapitre par des définitions de l'intelligence émotionnelle et ses différents principes. Nous avons réalisé différents outils pour intégrer ces principes de l'intelligence émotionnelle dans les environnements virtuels (principalement les jeux vidéo). Ensuite, nous présentons des modèles de design de jeux vidéo (pour divertissement et apprentissage). Puis, nous citons quelques projets de recherche qui ont traite de l'analyse de l'état émotionnel et mental du joueur. Nous motivons l’importance d'ajouter cette dimension émotionnelle dans le design des jeux vidéo pour arriver à des jeux émotionnellement intelligents. Enfin, nous donnons quelques exemples d'adaptations dans les jeux selon différents critères (Par exemple: le niveau de difficulté, les performances du joueur, les mesures physiologiques, l’émotion du joueur et sa motivation ...).

2.2 Les émotions, la motivation et l’intelligence émotionnelle

2.2.1 Les émotions

La prise de décision et la sélection des actions de l'utilisateur sont régulées et contrôlées non seulement par des stimuli externes, mais également par sa personnalité, ses émotions et ses humeurs. Selon des études réalisées par des chercheurs dans différents domaines, tels que la psychologie, les neurosciences et la philosophie, les principaux rôles de l’émotion sont: le contrôle de l’action, la motivation, l’adaptation, la régulation sociale, la gestion des objectifs, la concentration de l’attention et le modèle de soi. Selon (LeDoux, 1992), l’émotion est un processus cérébral qui évalue une expérience. Dans son livre « *How emotions are made* » (Barrett, 2017), Barrett définit les émotions comme des mots que nous attribuons à certaines configurations d'états corporels, de pensées et de facteurs situationnels.

Ainsi, les émotions influencent l'être humain dans plusieurs aspects, tels que :

- Cognitif: les émotions influencent ou sont influencés par la pensée, l'apprentissage, le processus de décision, la mémoire, le comportement et d'autres fonctions cognitives.
- Physiologique: les émotions sont liées aux hormones, signaux cérébraux, fréquence cardiaque, fréquence respiratoire, activité électrodermale …
- Expressif: les émotions peuvent se manifester par les expressions faciales, les gestes, la posture et le ton vocal.
- Motivation: les émotions sont influencé par les objectifs et les motives (*drives*)
- Feeling: par le fait d'être conscient de son état émotionnel.

Les théories de l'émotion diffèrent en ciblant différentes composantes de l'émotion. Par exemple les théories d'évaluation cognitive sont liées aux antécédents cognitifs d'émotion (événements, situation), les théories des émotions discrètes se concentrent sur les conséquences physiologiques et expressives de l'émotion alors que les théories de l'émotion dimensionnelles représentent les catégories d'émotions dans un espace à repère bidimensionnel ou tridimensionnel. Les techniques de reconnaissance des émotions font souvent appel à la théorie des émotions discrètes ou dimensionnelles et évitent les modèles d'évaluation.

2.2.1.1 Les modèle de l'évaluation cognitive

Le modèle OCC (Andrew Ortony et al., 1990) a été développé pour comprendre les émotions. Il est considéré comme le modèle standard pour la catégorisation des émotions (22 catégories d'émotions). Le modèle OCC explique les émotions humaines et prédit les émotions expérimentées selon la situation. Les émotions sont réparties dans les groupes suivants: réactions (positive / négative) aux événements, actions et objets. Ainsi, les conséquences d'un événement peuvent plaire ou déplaire à l'individu (*pleased / displeased*), qui peut accepter ou refuser les actions (*approve / disapprove*), et les caractéristiques d'un objet peuvent attirer ou non son attention (*like / dislike*).

Un modèle d’émotions basé sur les événements a été proposé par (Roseman, Spindel, & Jose, 1990). Ils ont classé les événements en deux groupes: les événements cohérents avec le but (*motive-coherent*) et les événements incohérents avec le but (*motive-incoherent*). Ce modèle détermine l'occurrence d’émotion en fonction de la certitude qu'un événement se produirait réellement. Ce modèle ne fournit pas une compréhension et une explication complètes des processus émotionnels, car il ne décrit pas de méthode pour catégoriser les événements perçus. Certains événements peuvent être perçus simultanément cohérents et incohérents avec le but pouvant ainsi générer des émotions contradictoires.

Une autre approche a été présentée dans la théorie de l'émotion de Frijda (Frijda, 1986). L'idée centrale de cette théorie est le terme intérêt. Un intérêt représente la tendance d'un système à cibler certains états de l'environnement. Ainsi, l'intensité des émotions est déterminée essentiellement selon les intérêts pertinents.

2.2.1.2 Les théories d’émotion discrètes

Selon la théorie des émotions discrètes, des émotions spécifiques sont des réactions émotionnelles déterminées par la biologie, dont l'expression et la reconnaissance sont fondamentalement les mêmes pour tous les individus, indépendamment des différences ethniques ou culturelles. Après la réalisation de plusieurs études interculturelles, Paul Ekman (Ekman, 2007) conclut que tous les humains sont très similaires dans la façon de produire et reconnaître les expressions faciales d'au moins six émotions de base: la joie, la colère, la tristesse, le dégoût, la surprise et la peur (Figure 1).

La reconnaissance des expressions faciale est basée sur un système de codage des actions faciales (FACS) (Ekman & Rosenberg, 1997) qui est un système de classification standardisé des expressions faciales pour les codeurs humains experts, basé sur les mouvements presque involontaires des muscles du visage (caractéristiques anatomiques).

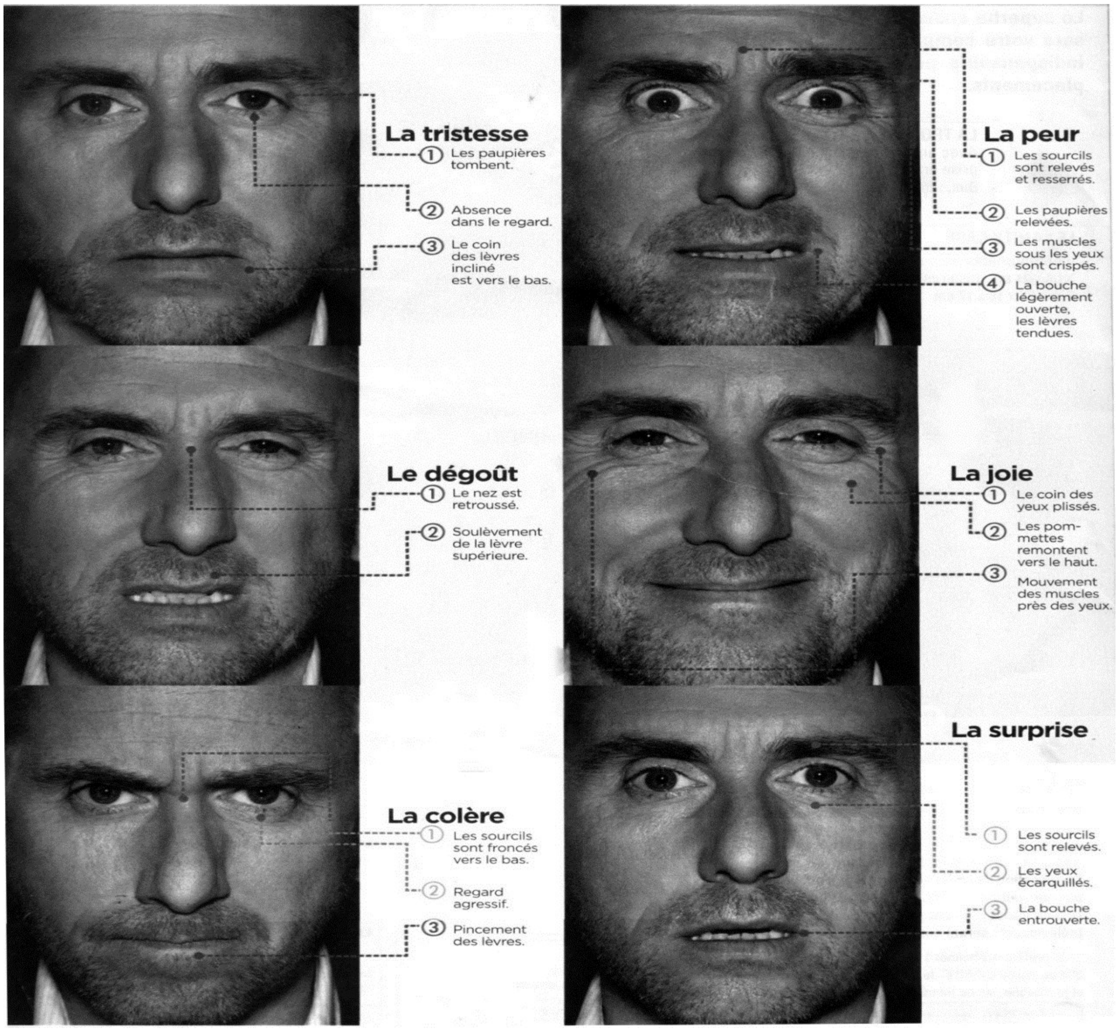


Figure 1. Les émotions de bases (Ekman & Rosenberg, 1997)[3]

Plusieurs recherches supportent l'hypothèse de l'émotion discrète puisqu'ils ont trouvé que les émotions de bases sont similaires indépendamment de l'âge et du genre (Barrett, Gendron, & Huang, 2009). Les émotions sont innées puisque des circuits cérébraux spécifiques sont activés selon les émotions ressenties (LeDoux, 1992). Par exemple, la peur active l'amygdale qui prend le contrôle du comportement de l'individu et de ses émotions (LeDoux, 2017).

[3] http://tpe-corps-et-emotions.eklablog.com/b-les-expressions-faciales-a106253464 (Accédé le 15/11/2018)

2.2.1.3 Les théories d'émotion dimensionnelles

Les modèles dimensionnels de l'émotion représentent les émotions humaines en les plaçant dans un espace de deux ou trois dimensions. Le modèle dimensionnel d'émotions le plus connu est le modèle circumplex de Russell (Figure 2) composé de deux dimensions la valence et l'activation (Russell, 1980). Les modèles dimensionnels de l'émotion supposent que les émotions sont gérées par un système commun neurophysiologique (Posner, Russell, & Peterson, 2005).

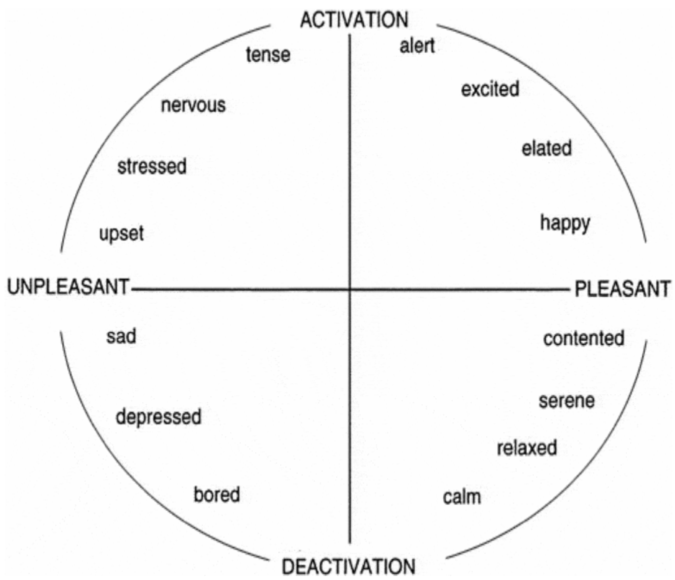


Figure 2. Le système circumplex de Russell (Posner et al., 2005)

Les modèles dimensionnels les plus importants sont le modèle circumplex (Russell, 1980), le modèle de Plutchik (Plutchik, 1982) et le modèle PAD (Mehrabian & Russell, 1974).

Le modèle de Plutchik représente les émotions dans un modèle tridimensionnel sous forme d'un cône, illustré à la Figure 3. La face circulaire est composée de plusieurs quadrants contenant les émotions de base organisées en fonction de leur proximité, la troisième dimension est l'intensité de l'émotion qui permet de former le cône. Les émotions qui s'opposent sont considérées comme contradictoires tandis que les émotions des quadrants adjacents ont des propriétés communes. Deux émotions adjacentes peuvent ensuite être combinées pour former une nouvelle émotion composée.

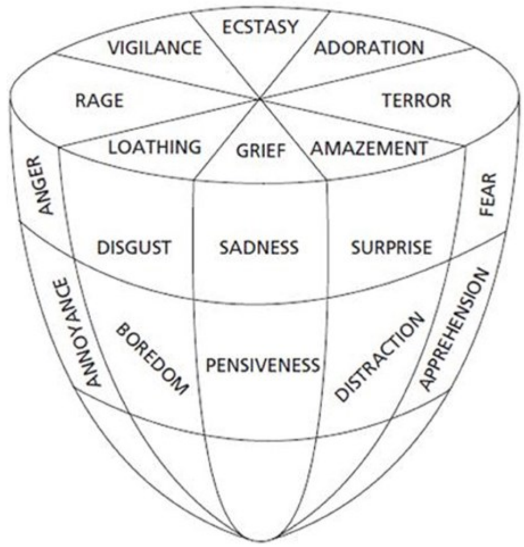


Figure 3. Le modèle d'émotions de Plutchik (Plutchik, 1982)

Le modèle PAD (Mehrabian & Russell, 1974) représente les émotions dans un modèle tridimensionnel (Pleasure, Arousal, Dominance). Ce modèle a été introduit lors d'une étude sur l'influence des environnements physiques sur les émotions des gens. Une théorie physiologique de l'émotion utilisant le modèle PAD a été proposée par Lang et ses collègues (Lang, Bradley, & Cuthbert, 1990). De plus, le modèle PAD a été utilisé comme base dans l'outil d'évaluation visuelle des émotions SAM - Self-Assessment Manikin (Bradley & Lang, 1994) et aussi dans l'outil d'évaluation AffectButton (Broekens & Brinkman, 2013).

2.2.2 La motivation

La motivation est le processus qui régule l'engagement d'un individu pour une activité précise. Elle favorise certaines actions pour un but visé avec l'intensité souhaitée et assure l'accomplissement de la tache ou son interruption (en cas de manque de motivation ou perte d'intérêt). La motivation prend en considération plusieurs paramètres relatifs à l'environnement et à la possibilité d'atteindre une situation souhaitée.

La motivation (état motivationnel) est vue comme une variable hypothétique prise comme cause pour les réactions comportementales. Selon (Vallerand & Thill, 1993)), « Le concept de **motivation** représente le construit hypothétique utilisé afin de décrire les forces internes et/ou externes produisant le **déclenchement**, la **direction**, **l'intensité** et la **persistance** du comportement ». Parmi les théories motivationnelles contemporaines, la théorie de l'autodétermination (Ryan & Deci, 2000) qui se base sur l'existence de deux types de motivation (**intrinsèque** vs **extrinsèque**), chacune menant à des comportements différents.

2.2.2.1 Modèle de l'auto-détermination

Le modèle de l'auto-détermination de Deci et Ryan (Ryan & Deci, 2000) étudie la dynamique motivationnelle d'un individu envers une activité par rapport au « locus de contrôle ». Le modèle de l'auto-détermination considère l'existence de différentes formes de motivation selon un degré d'auto-détermination qui présente le sentiment de libre choix et de cohérence interne (locus interne/externe). Dans ce modèle trois grands types de motivation ont été organisés selon un continuum : la motivation intrinsèque, la motivation extrinsèque et l'amotivation.

Motivation intrinsèque

La motivation intrinsèque caractérise les individus qui pratiquent une activité pour l'intérêt qu'elle présente en elle-même et pour le plaisir et la satisfaction qu'ils en retirent (Ryan & Deci, 2000). La motivation intrinsèque est caractérisée par un locus de contrôle interne pour répondre aux besoins individuels de compétence et d'autodétermination.

Motivation extrinsèque

Les motivations extrinsèques sont reliées aux intérêts d'un individu envers une activité qui sont plutôt de locus de causalité qui tend à être externe, essentiellement dirigées par des facteurs externes (récompenses, obligations, pressions, etc.) (Ryan & Deci, 2000). Le sentiment d'autodétermination décroît alors selon que l'individu perd la maîtrise de la régulation de ses comportements. La motivation extrinsèque est classée sous quatre formes qu'on peut mettre sur un continuum ordonné par degrés décroissants de motivation autodéterminée, à savoir la motivation extrinsèque à :

- régulation intégrée : le sujet perçoit une relative concordance entre l'activité et ses motifs internes.
- régulation identifiée : le sujet s'engage parce qu'il juge l'activité valable et qu'il a identifié l'importance de son engagement.

- régulation introjectée (Vallerand et al., 1989) : le sujet s'engage dans une activité pour éviter des sentiments négatifs, tels que la culpabilité, ou pour chercher l'approbation des autres.
- régulation externe : le sujet est motivé par des éléments extérieurs à l'activité comme des récompenses matérielles ou l'évitement de punitions.

L'amotivation

L'amotivation caractérise les individus qui effectuent une tache en l'absence de motivation. Un état d'aliénation se développe au détriment d'une recherche de satisfaction du besoin d'autodétermination. L'action est perçue comme sans intérêt ou valeur pour l'individu, ou il ne se sent pas compétent pour l'accomplir, ou il ne croit pas que ça va lui apporter un résultat appréciable. La personne ne se sent pas reliée à l'action, ne voit plus de lien entre les gestes et les résultats attendus.

Le continuum de (Ryan & Deci, 2000) a été utilisé par (Denis, 2006) dans le contexte du jeu vidéo (Figure 4). Ce continuum identifie les étapes du parcours d'un joueur de l'amotivation vers l'autodétermination.

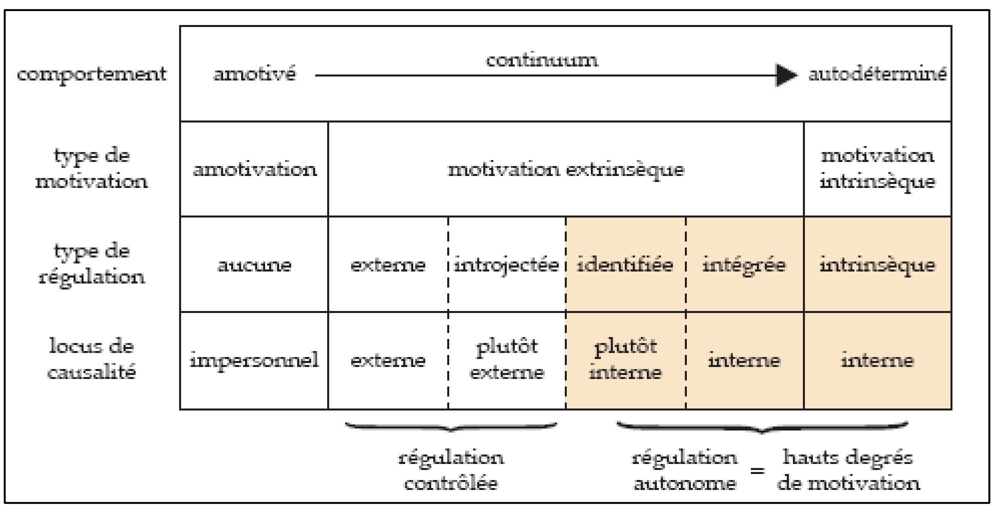


Figure 4. Continuum de motivation de Deci et Ryan (Denis, 2006).

Derbali (Derbali, Ghali, & Frasson, 2013) a utilisé le modèle de Keller et les stratégies de motivation dans un jeu sérieux nommé HeapMotiv pour enseigner la structure de données du tas binaire.

2.2.2.2 La motivation intrinsèque dans les jeux

En regardant comment les utilisateurs sont engagés en jouant avec des jeux intrinsèquement motivants, Malone (Malone, 1981) s'est intéressé à l'étude de la théorie derrière l'apprentissage intrinsèquement motivant, ou l'apprentissage à lequel l'individu s'engage sans motivation externes (récompenses ou punitions). Dans son article, Malone décrit les caractéristiques des environnements qui les rendent intrinsèquement motivant, avec des motivations individuelles telles que le défi, la fantaisie, la curiosité. Plus tard, (Malone & Lepper, 1987), ont ajouté le contrôle comme une motivation individuelle supplémentaire (Tableau 1). Ces caractéristiques peuvent être considérées comme les théories sur la façon de rendre l'apprentissage amusant. Ils ne sont pas seulement en mesure d'être appliqués aux jeux, mais les jeux par leur nature possèdent un nombre de ces fonctions (Malone, 1981).

Tableau 1. Motivation intrinsèque (Malone & Lepper, 1987).

Défi (challenge)	Résultats incertains, Buts à multi-niveaux, Difficulté progressivement variable, Informations cachées, Hasard, Feedback sur la performance.
Fantaisie	- Pour représenter une situation - Fantaisie extrinsèque : la fantaisie est ajoutée au curriculum. - Fantaisie intrinsèque : le *gameplay* et la fantaisie interdépendantes.
Curiosité	- évoquée avec un niveau de complexité correct. - curiosité sensorielle : audio et visuel - curiosité intellectuelle évoquée par des manques de connaissances. - feedback intéressant pour aider à construire de la connaissance.
Contrôle	- variété de choix pour avoir variétés de résultats - contingence : les actions de l'utilisateur influencent le résultat

2.2.3 L'intelligence émotionnelle et les environnements virtuels

Dans leur article intitulé "Intelligence émotionnelle", (Salovey & Mayer, 1990) ont défini l'intelligence émotionnelle (IE) comme "*la capacité de surveiller ses sentiments et ses émotions et ceux des autres, de les distinguer et d'utiliser ces informations pour guider ses* *pensées et ses actions*". Inspirés par les modèles psychologiques de l'émotion, de nombreux chercheurs ont reconnu dans Intelligence artificielle (IA) l'importance et l'utilité d'améliorer des environnements virtuels complexes, dynamiques et interactifs à l'aide de modèles informatiques des émotions. Cependant, la plupart des modèles informatiques d'émotion conçus ne représentent que des situations spécifiques et y répondent de manière prédéterminée. L'idée d'intelligence émotionnelle découle des travaux de Damasio (Damasio, 2002) et de Goleman (Goleman, 2006) et implique une prise de conscience de ses propres émotions et de celles des autres. Un modèle basé sur l'intelligence émotionnelle a été proposé par (Aguirre et al., 2008) pour représenter l'humeur et les émotions des avatars dans un environnement virtuel 3D et pour contrôler leurs comportements. La plupart des modèles des émotions dans les jeux se sont concentrés sur l'aspect physiologique ou cognitif de l'émotion. Cependant, nous pensons qu'il est nécessaire d'ajouter des caractéristiques humaines, telles que la personnalité, afin de prédire les émotions et la motivation dans les jeux.

2.2.4 Les modèles de l'intelligence émotionnelle :

Il existe deux principaux types de modèles d'intelligence émotionnelle: les modèles de capacités mentales et les modèles mixtes. Les modèles de capacités mentales permettent de prédire la structure interne de l'intelligence, ainsi que ses implications dans la vie quotidienne. C'est-à-dire qu'un modèle de capacités mentales se concentre sur les émotions et leurs interactions. Alors que les modèles mixtes se concernent sur les capacités mentales et les caractéristiques telles que la motivation, les états de conscience et l'activité sociale.

Ainsi, les modèles de capacités mentales fonctionnent dans une région définie par l'émotion et la cognition, alors que les modèles mixtes qualifient une multitude de composantes d'Intelligence émotionnelle. L'intelligence émotionnelle a été définit par (Salovey & Mayer, 1990) comme la capacité de percevoir et d'exprimer les émotions, d'assimiler des émotions dans les pensées, de comprendre et de raisonner avec les émotions, et de réguler les émotions en soi et chez les autres. Ils ont présenté un modèle de compétences élémentaires comprenant les domaines de compétences suivants: perception et expression de l'émotion, assimilation de l'émotion dans la pensée, compréhension et analyse de l'émotion, et la régulation réflexive de l'émotion. Le concept de l'intelligence émotionnelle a été défini par Goleman (Goleman, 1995) comme un modèle mixte avec les compétences suivantes: connaissance des émotions, gestion des émotions, l'auto-motivation, reconnaissance des émotions chez les autres et la gestion des relations sociales.

Nous concluons que les modèles de capacité mentale peuvent être décrits comme des modèles standards d'Intelligence émotionnelle. Ainsi, l'intelligence émotionnelle comprend quatre tâches spécifiques: la reconnaissance des émotions, la motivation, la compréhension des émotions et la gestion des émotions.

2.3 Les environnements virtuels

2.3.1 Les environnements de débats en ligne

Dans les médias sociaux, nous pouvons publier, discuter et lire via différents types de média (texte, image, audio et vidéo). Nous échangeons des messages et nous partageons des opinions en se justifiant par des arguments. Nous échangeons également des émotions via les médias sociaux. On peut influencer les gens via les médias sociaux. La persuasion a pour but d'influencer les gens dans leurs croyances, leurs attitudes, leurs intentions, leurs motivations ou leurs comportements (Gass & Seiter, 2015). La Rhétorique est la capacité de trouver les moyens de persuasion disponibles dans tous les cas. Il y a 2330 ans, Aristote avait décrit comment persuader les gens et les faire passer à l'action. Il décrivit les stratégies d'argumentations (Figure 5) basées sur la logique (Logos), la crédibilité (Ethos) et les émotions (Pathos).

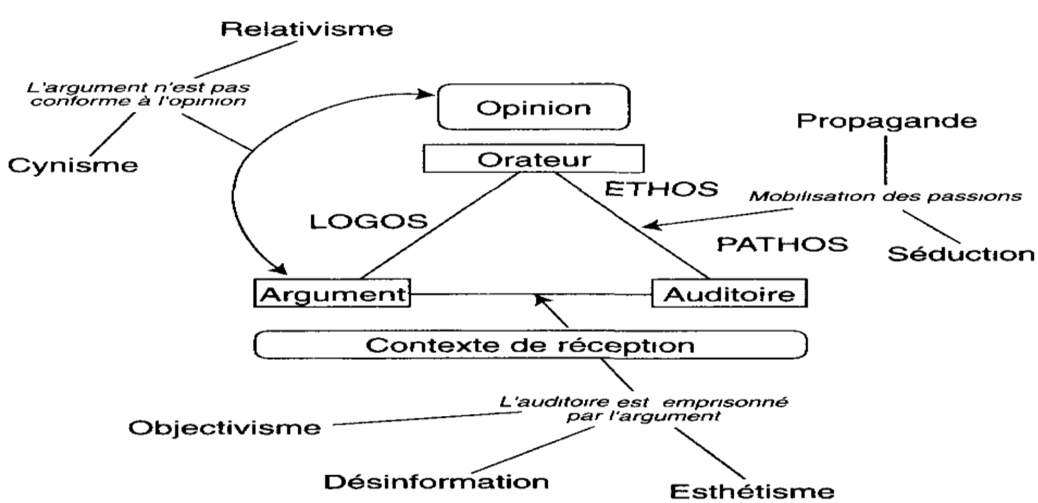


Figure 5. Les stratégies d'argumentation dans le modèle de communication (Breton, 2006).

2.3.2 Les environnements de jeux vidéo

On peut définir le jeu comme une activité de loisirs d'ordre physique ou bien psychique, soumise à des règles conventionnelles, à laquelle on s'adonne pour se divertir, tirer du plaisir et de l'amusement (Wikipédia[4]). En 40 ans, le jeu vidéo est devenu une industrie, une culture, un art, un sport et un média de masse. Les jeux vidéo d'aujourd'hui sont plus proches des films de cinéma tout en permettant l'interaction. Un jeu vidéo est un jeu utilisant un dispositif informatique où le joueur utilise des périphériques pour agir sur le jeu et percevoir l'environnement virtuel. Sous l'appellation jeu vidéo, on peut trouver une variété de produits comme les jeux de réflexion (puzzle, jeux de stratégie, …) sur internet ou les réseaux sociaux, les jeux d'aventure et jeux de rôle ainsi que les jeux de tir (les First Person Shooter (FPS), les Third Person Shooter (TPS), …), les jeux de rôle massivement multi-joueurs (MMORPG), en passant par les jeux de sport et de simulateurs (course de F1, avions, …). Il existe plusieurs types de systèmes sur lesquels il est possible de jouer à un jeu vidéo, et de nombreux jeux sont disponibles sur plusieurs de ces environnements (les ordinateurs, les consoles de jeux, les téléphones portables, les bornes d'arcade).

Pour bien commencer un projet de jeu vidéo, on doit écrire un scénario qui sert de plan au designer et au programmeur. Pour concevoir un jeu vidéo en trois dimensions (3D), nous avons besoin d'une application 3D tel Autodesk Maya, 3ds Max ou bien Blender (Gratuit). Ensuite vient l'intégration grâce à un moteur de jeux vidéo tels que: Unreal Engine et Unity 3D. La conception d'un jeu vidéo est basée sur un processus itératif de design de jeu (***game design***). Chaque cycle commence par un prototype qui passe par une phase de test puis de raffinage jusqu'à en arriver à un prototype ayant un *gameplay* optimal et qui soit la version finale du jeu.

2.4 Les mesures affectives et mentales dans les environnements virtuels

Plusieurs études ont eu recours à diverses méthodes, subjectives ou objectives, afin d'arriver à évaluer l'émotion et la motivation des usagers. Pour déterminer le niveau de réactions

[4] https://fr.wikiversity.org/wiki/Facult%C3%A9:Jeux (accedé le 26/11/2018)

affectives chez les joueurs, nous utilisons des analyses multimodales afin d'explorer plusieurs sources d'informations. Nous avons eu recours à plusieurs sources d'informations telles que : les questionnaires, les fichiers log, le traçage du regard et les données physiologiques. Ceci nécessite des analyses statistiques et des techniques d'IA pour la sélection et l'extraction des caractéristiques et la construction du modèle émotionnel.

2.4.1 Les questionnaires d'auto-évaluation « self-report »

Une des méthodes subjectives pour évaluer les émotions est l'utilisation de questionnaires d'auto-évaluation. Ils sont utilisés pour recueillir un rapport du participant sur ses émotions et sa motivation. Ils sont sous forme d'un ensemble de questions à choix multiple pour éviter toute ambiguïté. L'analyse des questionnaires permet d'obtenir un profil de la personne évaluée. Les questionnaires offrent la possibilité d'évaluer: des caractéristiques personnelles, la perception d'un élément ou d'une activité pour une personne, etc. Par exemple, le questionnaire Big Five (John & Srivastava, 1999), est un questionnaire de évaluation psychologique des traits de personnalité. Il se présente sous la forme de 50 items, sur lesquels le participant prend position sur une échelle en 5 points. Ce questionnaire diagnostique trois types de personnalités (ouverture, conscience, extraversion, agréabilité et névrotisme). Nous en avons eu besoin à des questionnaires pré-test pour recueillir les données sociodémographiques et le profil de l'usager (Expérience, style d'apprentissage et traits de personnalité). En post-test, nous avons eu recours aux questionnaires sur l'utilisabilité et l'expérience de l'utilisateur à propos des environnements virtuels utilisés. Ces questionnaires sont composés d'items de type Likert, c'est-à-dire dont la réponse est exprimée sur une échelle graduée entre « complètement d'accord » et « pas d'accord du tout ». En outre, les questionnaires post-tests peuvent servir comme indicateur de l'état émotionnel et la motivation du joueur/apprenant.

2.4.2 Les fichiers journaux (log-files)

De nouvelles informations peuvent être acquises durant une session de jeu (par exemple les évènements dans le jeu : attaque par un ennemi, réussite d'une étape, taux de clics, …) grâce à l'utilisation des outils d'évaluation et de fouille de données de traces dans les fichiers journaux (logfiles). Les fichiers log contiendront des données quantitatives et qualitatives à partir des réponses et des interactions du joueur avec le jeu. Avec des techniques d'apprentissage machine et de fouille de données (*data-mining*), on peut analyser ces données afin d'obtenir des informations utiles pour améliorer la motivation dans les jeux. Par exemple (Hershkovitz & Nachmias, 2008), ont développé un outil de mesure de la motivation basé sur l'analyse des fichiers de log dans un environnement d'apprentissage à distance. Leur étude porte sur la possibilité de prédire le niveau de motivation des apprenants.

2.4.3 Indicateurs électro-physiologiques

Nous croyons que le suivi et l'analyse des signaux électro-physiologiques représente la méthodologie de reconnaissance émotionnelle et cognitive la mieux adaptée dans le contexte de jeux. Il existe plusieurs études sur la reconnaissance (détection) de l'état émotionnel et motivationnel de l'usager à partir des signaux électro-physiologiques. Nous avons utilisé plusieurs équipements (Figure 6) pour recueillir les données physiologique tels que : le casque Emotiv Epoch pour les données EEG, le bracelet Affectiva Qsensor pour les mesures l'activité électrodermale des joueurs, FaceReader 6.0 pour les expressions faciales et Tobii Tx300 pour le suivi oculaire.

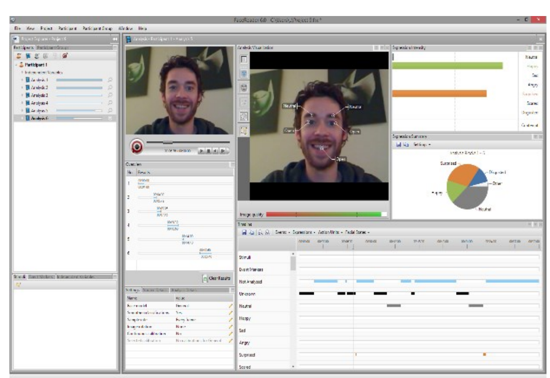
 **Casque Emotiv Epoch (EEG)**	 Bracelet Affectiva Q sensor (EDA)
Tobii TX300 	**Le Logiciel FaceReader 6.0** 

Figure 6. Les équipements de mesures des réactions de l'usager.

Parmi les indicateurs sur l’activité physiologique nous pouvons citer :

EDA (Electrodermal Activity) : cet indicateur est utilisé pour mesurer la conductivité de la peau en fonction de la transpiration (glandes sudoripares). La conductivité de la peau augmente quand la peau transpire. Plusieurs études ont montré qu’il existe une corrélation entre la conductivité de la peau et l’excitation (activation, stress).

HR (Heart Rate) : cet indicateur mesure le nombre de battements de cœur par minute. Elle est enregistrée à partir de l’activité électrique associée à la contraction des muscles du cœur. Il existe une corrélation positive entre le pic du rythme cardiaque et la valence.

RSP (Respiration) : cet indicateur est enregistré à partir de la contraction du diaphragme lors de la respiration. Selon quelques chercheurs l’intensité de l’émotion influence le rythme de la respiration (Prendinger et al., 2003).

Suivi oculaire : cet indicateur mesure le nombre de saccades et de fixations des yeux ainsi que la variation du diamètre de la pupille. Les saccades sont des mouvements rapides de l'œil qui dirigent le regard d'un champ visuel à un autre. Les fixations sont les moments entre les saccades où l’œil reste relativement immobile. Pour mesurer l’immersion, (Jennett et al., 2008) ont utilisé les réactions oculaires dans un environnement de jeux vidéo. Pour le suivi oculaire, nous avons utilisé Tobii Tx300. Il permet de tracer le regard avec différents taux d'échantillonnage (60, 120 ou 300 Hz), tolère le mouvement modérée de la tête de l’utilisateur, et peut fonctionner à une distance de jusqu'à un mètre. L’oculomètre mesure la position des yeux dans l'espace pour traduire ces informations en un point de regard dans la scène ou sur l'écran. Les utilisateurs doivent se soumettre à une procédure de calibration. Elle consiste à suivre avec les yeux plusieurs cibles prédéfinies (en pratique, cinq ou plus), généralement situés à des points clés tels que les coins de l'écran et son centre. Au cours de cette procédure, l'ordinateur recueille les données des points correspondants pour chaque cible, dans l'espace de l'écran et dans l'espace de la caméra de l’oculomètre, qui sont ensuite mis en correspondance pour traduire la position de l'œil dans la scène, typiquement en utilisant des modèles des yeux.

Expressions faciales : cet indicateur est enregistré à partir d'une webcam pour extraire les émotions de base à partir des expressions faciales (Ekman, 2005) des participants durant la session du jeu. Les émotions de bases seront identifiés en temps réel trame par trame à l'aide du logiciel d'analyse FaceReader 6.0. Le fichier de données résultant va contenir des valeurs de mesure sur les émotions suivante : la neutralité, la joie, la colère, la surprise, la peur, le dégoût.

EEG (Electroencephalography) : cet indicateur mesure l'activité électrique du cortex et de la surface du cerveau. Les signaux EEG sont repartis en différentes bandes de fréquences: delta (1-3 Hz), thêta (4-7 Hz), alpha (8-13 Hz), bêta (14 à 30 Hz) et gamma (31-42 Hz). Ces bandes peuvent être associées à des processus cognitifs spécifiques (Von Stein & Sarnthein, 2000) :

- Delta (sommeil, sommeil profond, transes): Ces rythmes se manifestent au cours du sommeil profond à mouvements oculaires rapides, mais aussi lors d'une souffrance grave du cerveau. Ces ondes sont plus localisées sur les lobes temporaux et en états subjectifs sur les lobes occipitaux.
- Thêta (relaxation, hypnose...): Lors d'un sommeil non-profond, ces rythmes Thêta jouent un rôle lors de l'apprentissage et de la consolidation de mémoires. Les ondes thêta sont reliées à notre subconscient, elles gouvernent la partie du mental entre le conscient et l'inconscient. Elles sont reliées à la mémoire et aux sensations.
- Alpha (relaxation, détente...): Présentes lors d'un état d'éveil de vigilance mais au repos, ces ondes sont associées à la coordination d'activité mentale et à l'apprentissage. Captées sur la partie postérieure de la tête dans la région occipitale, dans le cortex et dans sa bande périphérique. Les ondes Alpha sont un pont entre conscient et subconscient. On augmente la fréquence Alpha en fermant les yeux ou en respirant profondément et on la diminue par la pensée ou le calcul.
- Bêta (attention, pensée éveillée...): Ces rythmes sont associés à un état d'éveil mais avec engagement dans une tâche cognitive (p. ex. prendre une décision, résoudre un problème). Ces ondes sont plus captées sur les lobes temporaux, et sur les lobes occipitaux et frontaux du cerveau.
- Gamma (fonctions intellectuelles, conscience, perception...): Ces rythmes sont associés à la synchronisation de plusieurs régions du cerveau (traitement d'informations). Elles

sont également présentes lors d'état nécessitant un haut niveau d'attention ou de concentration. Une activité des ondes Gamma est associée à une bonne mémoire, tandis qu'une insuffisance crée des incapacités de se rappeler d'information et de manque de concentration.

Un index d'engagement (Chaouachi et al., 2010b; Pope, Bogart, & Bartolome, 1995a) peut être calculé à partir des valeurs de l'énergie moyenne de ces bandes de fréquences. Le cerveau est composé de différentes régions, chacune d'elles a ses fonctions émotionnelles et cognitives spécifiques. Par exemple, la peur induit une plus grande activation à la région frontale droite du cerveau, alors qu'une expression de joie induit une plus grande activation dans la région frontale gauche du cerveau (Avram et al., 2010).

2.5 L'adaptation dans les jeux affectifs

Les jeux affectifs sont des jeux vidéo qui s'adaptent selon l'état émotionnel du joueur estimé à partir des mesures affectives des capteurs pour ajuster dynamiquement le contenu jeu afin d'améliorer l'engagement, l'immersion, l'excitation et le défi dans le jeu (Lara-Cabrera & Camacho, 2019). Les jeux basés sur les mesures d'émotions sont dotés d'une boucle de rétroaction (Figure 7) qui permet l'adaptation des éléments du jeu en fonction des émotions détectées ou bien des signaux physiologiques captés.

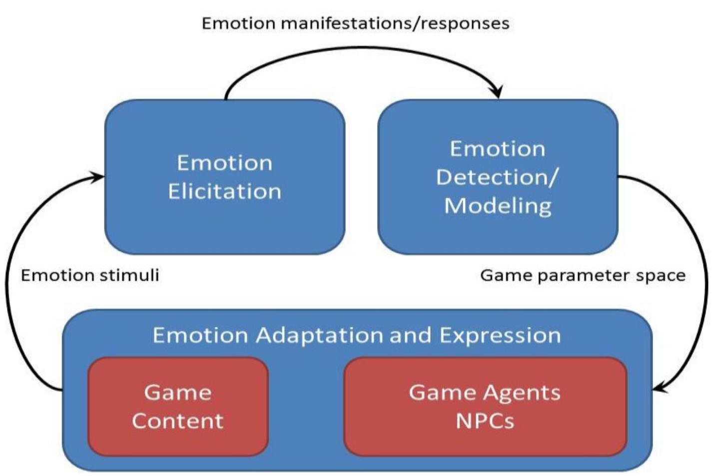


Figure 7. La boucle de rétroaction dans les jeux affectifs (Georgios N Yannakakis & Paiva, 2014).

Une boucle de rétroaction dans un jeu affectif comprendrait: l'acquisition multimodale d'émotions, la modélisation et l'identification des signaux collectés via apprentissage machine ou méthodes statistiques, adaptation du jeu par modification du contenu du jeu, en tenant compte de la force et du type de l'émotion reconnue (Georgios N Yannakakis & Paiva, 2014). Les jeux vidéo affectifs avec adaptation basée sur les émotions utilisent trois types d'adaptations dans la dynamique du jeu (Bontchev, 2016), à savoir:

- ajustement de tâches de jeu explicites, implicites ou dirigées par les joueurs;
- adaptation de la difficulté au niveau de compétence ou d'anxiété du joueur;
- ajustement des propriétés audiovisuelles telles que la lumière ambiante.

2.5.1 Technique d'adaptation automatique des tâches :

Les méthodes d'adaptation automatisée des tâches ont pour but de maintenir un niveau optimal de la charge de travail et de l'attention du joueur pendant l'exécution de la tâche du jeu. En effet, une **boucle de rétroaction négative** augmentant l'attribution des tâches lorsque les joueurs deviennent moins vigilants et inversement, Par exemple en adaptant l'aspect narratif dans le jeu (Rowe, Mott, & Lester, 2014) . En général, les tâches de jeu peuvent être classées en **trois types de tâches** : **explicites** (objectif et missions du jeu), **implicite** (éviter les échecs et accumuler le maximum de score) ou **dirigées par le joueur** (surtout pour les jeux sans narratif prédéfini, p. ex. créer et construire des objets dans Minecraft[5] et Fortnite[6]). La planification des tâches dans le jeu peut être contrôlée par les mesures physiologiques afin de favoriser ou d'atténuer des émotions spécifiques du joueur, par exemple en donnant des directives au joueur (Conati, Jaques, & Muir, 2013; Derbali et al., 2013).

2.5.2 Réglage du niveau de difficulté du jeu

Pour équilibrer dynamiquement le jeu, une boucle de rétroaction positive augmente la difficulté des tâches lorsque le joueur est plus performant et inversement. Les performances des joueurs permettent de réaliser directement les ajustements de la difficulté sans aucun

[5] https://minecraft.net/en-us/

[6] https://www.epicgames.com/fortnite/en-US/home

périphérique supplémentaire (Andrade et al., 2005). L'adaptation du niveau de difficulté selon la performance est très efficace dans les jeux thérapeutiques (Hocine et al., 2015).

De la même manière, (Abdessalem & Frasson, 2017) combinent le retour de frustration et d'excitation des joueurs pour l'ajustement dynamique de la difficulté en utilisant une boucle négative de rétroaction. Dans le même travail, les retours basés sur la frustration se sont avérés plus efficaces sur l'évaluation de jeu comme immersif que ceux basés sur les performances.

Les méthodes de réglage du niveau de difficulté du jeu utilisant seulement les données affectives peuvent ne pas être optimales. En fait, Plusieurs études (C. Liu et al., 2009; Sabourin & Lester, 2014) proposent des mécanismes d'ajustement de la difficulté en considérant « la performance du joueur, sa personnalité et le contexte du jeu » permettant d'avoir une expérience de jeu enrichissante.

En règle générale, les approches d'ajustement du niveau de difficulté peuvent être classées en trois types selon la méthode d'adaptation:

1) L'ajustement de la difficulté par la génération automatique de niveaux : utilise la technique de la génération procédurale de contenu (Procedural Content Generation - PCG) (par exemple : niveaux, quêtes, puzzles, textes) du jeu sans ou avec peu d'intervention humaine. Une revue définissant ces problèmes, leurs approches, et applications de la PCG dans la littérature peut être trouvée dans (Georgios N Yannakakis & Togelius, 2011) . La PCG pourrait être utilisée pour adapter le jeu au type de joueur, la production de contenu plus approprié, par exemple, ajuster la difficulté aux compétences du joueur, ou le type de défi aux préférences du joueur.
2) L'ajustement de la difficulté par modification de l'intelligence artificielle : par exemple, l'adaptation de l'intelligence des ennemis selon les compétences et les émotions du joueur génère du défi optimisé favorisant l'apprentissage (Hamari et al., 2016). Le comportement des PNJ (personnage non joueur) peut être contrôlé par des scripts dynamiques ou des méthodes d'apprentissage machine (Geogios N Yannakakis, 2012).
3) L'ajustement de la difficulté par modification du contenu de niveau : c'est-à-dire les éléments de jeu durant l'interaction du joueur en fonction de ses compétences. Par exemple, le système Hamlet (Hunicke & Chapman, 2004) définit des actions et des

règles pour l'inventaire du joueur dans le jeu Half Live[7] selon sa performance, ce qui a un effet direct sur la difficulté du jeu.

2.5.3 Adaptation des effets audio-visuels

Pour améliorer le *gameplay* et l'affichage d'un jeu d'horreur, des méthodes ont été proposé (Dekker & Champion, 2007) pour représenter les réponses physiologiques des joueurs aux propriétés du jeu audiovisuel. D'autre étude (Dekker & Champion, 2007) a proposé d'adapter la lumière ambiante dans les pièces d'un jeu vidéo selon des caractéristiques psychophysiologiques (la fréquence cardiaque et la conductance de la peau du joueur) en utilisant des algorithmes stochastiques. Diverses approches psychophysiologiques et pratiques de conception sonore ont été explorées (Garner, 2013) en vue de créer de plus grandes expériences émotionnelles dans un jeu adaptatif centré sur l'audio, en mettant un accent particulier sur les corrélations entre la peur et le son du jeu.

2.6 Le design de jeux vidéo

Selon la littérature, la conception du jeu implique deux perspectives: la perspective du designer et celle du joueur, comme dans le modèle MDA - Mécanique, dynamique et esthétique (Hunicke, LeBlanc, & Zubek, 2004). Les jeux peuvent aider les joueurs à se maîtriser et à acquérir des connaissances et des compétences liées aux situations de la vie quotidienne. Une telle pratique nécessite de l'attention, de l'intelligence et une résistance psychologique, le tout étant bien géré par le joueur. La conception du jeu doit être guidée par une description de l'expérience souhaitée du joueur dans le *gameplay*. En général, ce processus repose sur des documents de conception qui guident le projet du début à la fin. Dans ce qui suit nous présentons les trois modèles de design de jeu les plus connus, à savoir les modèles : MDA[8], DPE[9] et DDE[10].

---

[7] https://fr.wikipedia.org/wiki/Half-Life

[8] Mechanic, Dynamics, and Aesthetic

[9] Design, Play, and Experience

[10] Design, Dynamics, and Experience

2.6.1 Le Modèle MDA: Mechanic, Dynamics, and Aesthetic

Le modèle MDA (Hunicke et al., 2004) offre une approche formelle, contribuant ainsi à la compréhension du design des jeux et à définir le rapport entre le designer et le joueur (Figure 8). Le model MDA est compose de trois catégories :

- la mécanique (M) : elle contient une description des composants spécifiques du jeu en tant qu'actions, comportements et mécanismes de contrôle. Du point de vue du concepteur, les mécanismes créent le comportement dynamique du système de jeu et les expériences esthétiques du joueur.
- la dynamique (D) : elle comprend une description du comportement du système en relation avec la mécanique et les résultats durant le *gameplay*.
- l'esthétique (A - *Aesthetics*) : elle contient la description des réponses émotionnelles à évoquer chez le joueur pendant le jeu. Du point de vue du joueur, la couche esthétique définit l'apparence du jeu.

Selon ce modèle, le designer crée le jeu pour que le joueur puisse jouer, et tient donc compte de la perspective du joueur au cours du processus de design:

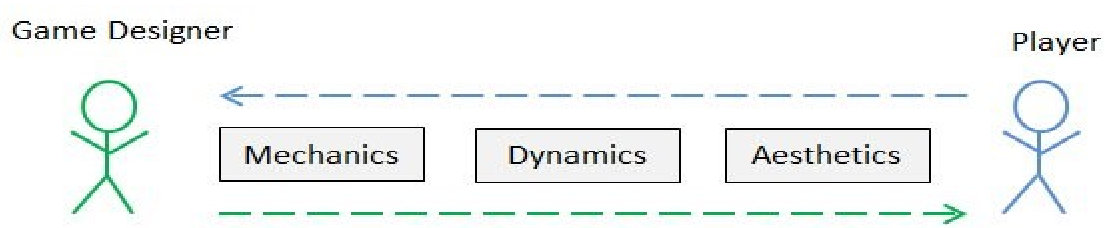


Figure 8. Le modèle MDA de Hunicke, LeBlanc & Zubek (2004) (König et al., 2017).

Toute modification mineure apportée à l'une de ces catégories peut avoir un effet progressif, car une catégorie peut affecter l'autre.

2.6.2 Le modèle DPE: Design, Play, and Experience

Le modèle DPE (Winn, 2009) a été proposé comme une extension du modèle MDA en préservant les perspectives du designer et du joueur et en rajoutant une flèche de retour entre les catégories principales *Design* (D) et *Experience* (E) pour représenter les itérations dans le processus de design (Figure 9). Ce modèle souligne le rôle de la technologie dans le processus de conception du jeu et dans l'expérience de l'utilisateur en la mettant comme couche de base de ce modèle.

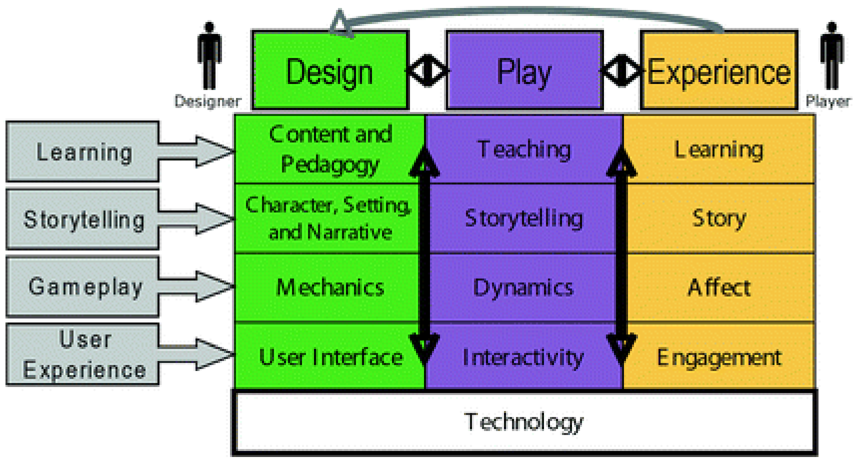


Figure 9. Le modèle DPE (Winn, 2009).

Ce modèle comporte quatre couches principales (Figure 9): Apprentissage, Narration, Gameplay et Expérience de l'utilisateur.

- La couche Apprentissage permet au designer de planifier comment intégrer le contenu pédagogique dans le jeu et quel contenu utiliser pour assurer l'apprentissage à partir de l'expérience globale.
- La couche Narration (*Storytelling*), qui présente deux contextes: (1) l'histoire du designer suscitant l'engagement du joueur et transmettant le contenu du jeu; et (2) l'histoire du joueur résultant de la narration projetée lors de l'interaction du joueur et à ses choix.
- La couche *Gameplay* définit les actions du joueur et leurs conséquences. Elle est composée de trois catégories (Mécanique, Dynamique et Affect) similaire au modèle MDA en changeant Esthétique par Affect. Dans cette couche, le designer définit les effets émotionnels à provoquer chez le joueur en tenant compte de tous les aspects du jeu.

- La couche Expérience de l'utilisateur a pour objectif de créer des moyens d'atteindre les résultats souhaités en impliquant le joueur et favoriser une immersion significative dans l'expérience de jeu.

Le modèle DPE met l'accent sur l'importance de la spécialisation des intervenants dans le projet de développement du jeu sérieux et suggère d'intégrer au moins trois spécialistes : le pédagogue, le spécialiste du contenu et le designer de jeux.

2.6.3 Le modèle DDE: Design, Dynamics, and Experience

Le modèle DDE (Walk, Görlich, & Barrett, 2017) a été conçu de façon élaborée pour couvrir l'ensemble du processus de développement de jeu, y compris les éléments de la production du jeu et le parcours éventuel du joueur. En préservant les perspectives du designer et du joueur, le modèle DDE est composé des catégories principales - Design, Dynamiques et Experience - sous forme de trois niveaux (Figure 10).

- Le niveau Design contient la catégorie mécanique qui détermine toute l'architecture du jeu, la technologie et d'autres éléments du système opérationnel du jeu. De même, le niveau design contient la catégorie Plan (*Blueprint*), qui définit conceptuellement le monde du jeu, et aussi la catégorie Interface, qui définit les éléments du système qui interagissent avec le joueur.
- Le niveau Dynamiques se focalise sur la définition du déroulement du jeu selon les mécanismes établis par la conception. Il considère l'imprévisibilité de la dynamique en raison des différents comportements que peut manifester les différents types d'acteurs.
- Le niveau Expérience, a été conçu plus en détail pour représenter en profondeur l'expérience de l'utilisateur. Ce niveau contient les catégories cervelet (*cerebellum*) qui détermine les émotions visées, cerveau (*cerebrum*) pour le parcours mental visé et la catégorie sens (*Senses*) pour les sens du corps humain ciblés. Ce niveau contient aussi la catégorie joueur-sujet (*Player-Subject*) reliée aux trois catégories précédentes pour refléter la subjectivité de l'‘expérience utilisateur d'une personne à l'autre.

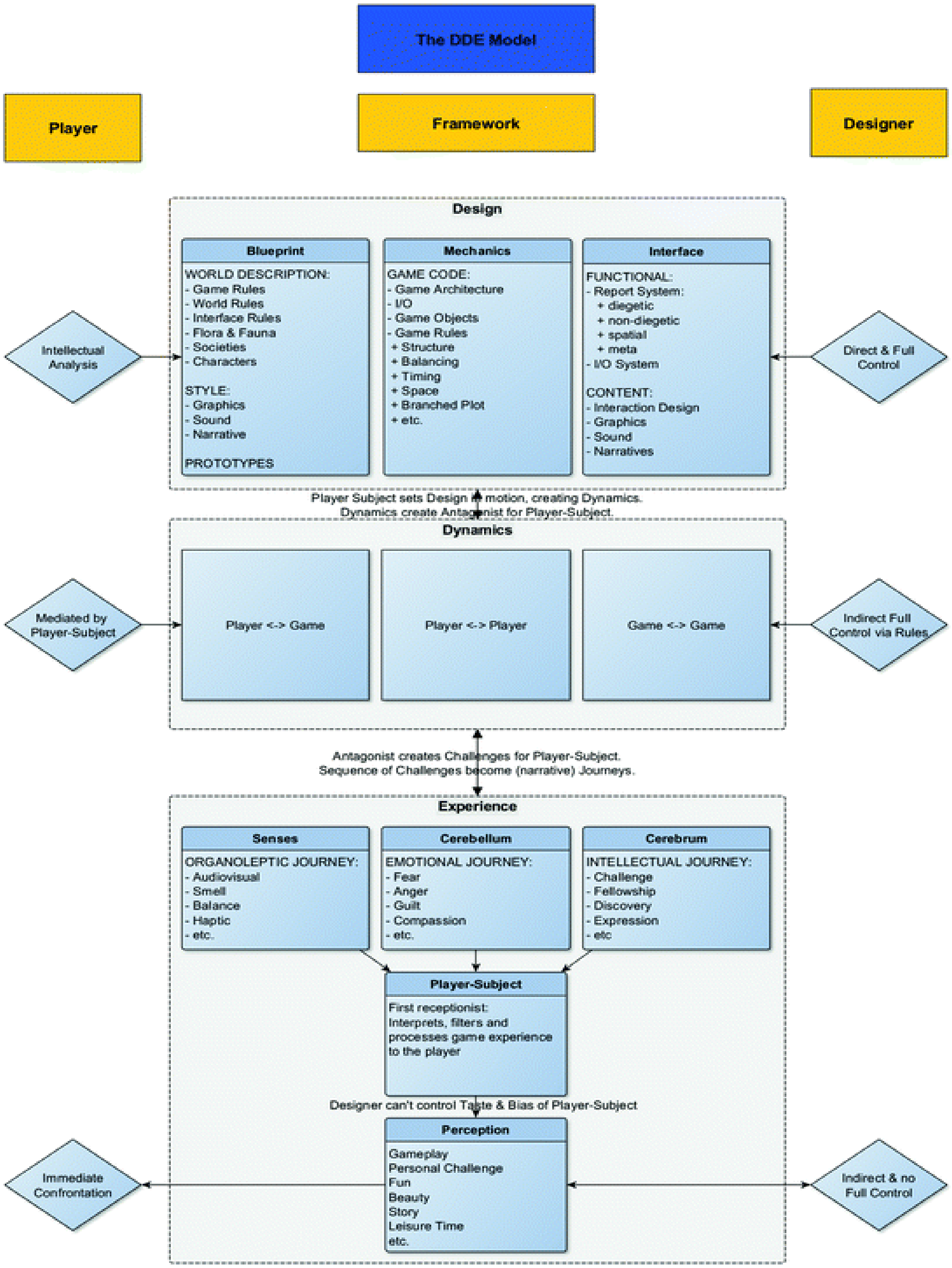


Figure 10. La synthèse du modèle DDE (Walk, Görlich, & Barrett, 2017).

2.7 Design de jeu émotionnellement intelligent

Le terme « émotionnellement intelligent » sous-entend l'intégration de techniques d'intelligence artificielle afin de concevoir des jeux plus adaptés selon les caractéristiques individuelles du joueur et plus adaptatifs à travers l'intégration et l'application de règles d'adaptation selon l'état émotionnel du joueur (détecté en temps-réel à travers les données physiologiques du joueur où ses expressions faciales par caméra).

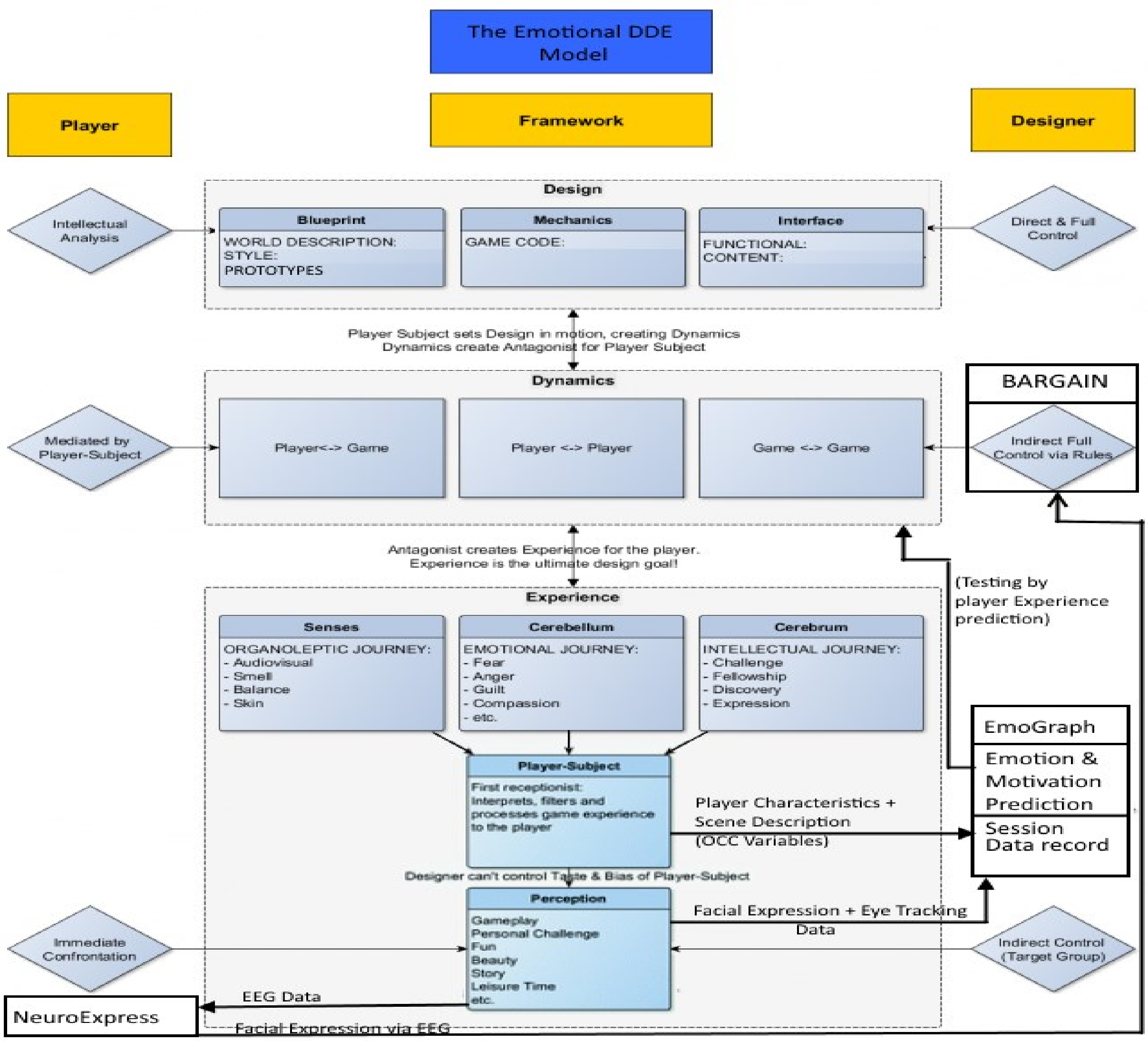


Figure 11. Le modèle DDE émotionnel.

Dans l'objectif d'avoir un design de jeux vidéo optimal avec moins de tests, nous pensons désormais intégrer des outils d'aide au designer pour la prédiction des émotions et de la motivation du joueur selon les objectifs de design de la scène de jeu et pour la définition et l'intégration de règles d'adaptation selon l'état émotionnel du joueur. Par ailleurs, l'intégration des techniques d'intelligence artificielle (apprentissage machine et apprentissage profond) peut faire avancer les modèles de design de jeu actuels et les rendre plus centrés sur l'émotion de l'utilisateur donc plus émotionnellement intelligents. Nous proposons ce modèle de design de jeu « *Emotional*-DDE » qui découle du modèle DDE mais intégrant **nos outils** de prédiction et d'adaptation émotionnels.

Comme présenté dans la figure ci-dessus (Figure 11), notre nouveau modèle de design de jeux émotionnellement intelligents « *Emotional*-DDE» intègre différents outils réalisés dans cette thèse. En effet, l'outil NeuroExpress est intégré du côté du joueur pour détecter ses expressions faciales à partir des données EEG lors de son interaction avec le jeu (c'est pour ça que la flèche « *EEG Data* » sort de la composante *Perception* dans la couche *Experience*). De plus les outils EmoGraph et BARGAIN sont intégrés dans la partie Designer puisque ce dernier va les utiliser dans la conception du jeu. En fait, le système Emograph contient deux modules : le premier module est pour la sauvegarde des données de la session de jeu à partir des *eye-tracking,* des expressions faciales de la composante « Perception » et les caractéristiques du joueur et la description de la scène de la composante « Player-subject ». Le deuxième module d'Emograph est pour la prédiction de l'émotion dominante du joueur et les orientations motivationnelles qui permet de donner une idée sur l'expérience de l'utilisateur selon les caractéristiques du joueur et la description de la scène. On pourrait le faire la prédiction de l'émotion dominante du joueur avant de construire la scène de jeu, en utilisant seulement la description de la scène par variables OCC et les caractéristiques du joueur. Le module de prédiction d'Emograph donne un feedback sur la dynamique du jeu sans avoir recours à faire des tests avec des vrais joueurs (la flèche sortante de « EmoGraph » vers la couche « Dynamics »). Selon les émotions prédites par Emograph dans la scène, le designer pourrait rajouter les règles adaptatives dans BARGAIN pour modifier la scène en tenant compte des émotions qui risquent d'être générées à l'intérieur de la scène de jeu. Lors du test du prototype, BARGAIN reçoit les mesures d'expression faciales à partir de NeuroExpress (la flèche entrant dans

BARGAIN) et calcule l'état émotionnel du joueur. Selon l'état émotionnel calculé, BARGAIN adapte les éléments du jeu de façon dynamique en appliquant les règles définies par le designer (la flèche sortante de « BARGAIN » vers la couche « Dynamics »).

2.8 Conclusion

Dans ce chapitre, nous avons exposé un état de l'art sur les modèles de design de jeu, les environnements virtuels et les méthodes d'adaptation afin d'optimiser l'expérience de l'utilisateur. Dans cette thèse, nous avons effectué des analyses affectives dans différents environnements virtuels tel que le débat en ligne (chat room), les jeux commerciaux tels que Outlast et Assasin's Creed, un jeu d'aventure Danger Island, et un jeu sérieux en réalité virtuelle Dilemmas_VR qui sont présentés dans les chapitres suivants.

Nous avons choisi d'utiliser ces différents environnements pour analyser les émotions des utilisateurs dans différents situations, de même pour construire des modèles émotionnels plus riches. De plus, nous avons choisi d'adopter une approche de mesure qui se base sur les données physiologiques pour faire de la prédiction des expressions faciales à partir des EEG et de la classification pour détecter l'émotion dominante du joueur et son orientation motivationnelle dans une scène de jeu. Plusieurs études (Chaouachi, Jraidi, & Frasson, 2015a; Ghali, Frasson, & Ouellet, 2016) ont montré que les mesures physiologiques basées sur les signaux cérébraux (EEG) sont efficaces puisqu'elles reflètent des états émotionnels et mentaux de l'usager tels que les émotions, l'engagement, l'attention, l'approche et l'évitement.

Ensuite, nous avons ciblé la question de la reconnaissance de l'état émotionnel du joueur lors de ses interactions dans un environnement de réalité virtuelle qui pose des défis techniques dans la capture de l'état émotionnel. En effet, dans un environnement de réalité virtuelle où le visage de l'usager est caché par le casque de réalité virtuelle, ce qui rend non-utilisable les systèmes de reconnaissance des expressions faciales via caméra. De plus, pour avoir un outil de diagnostic de scènes de jeux, il est nécessaire de prédire l'émotion dominante du joueur et ses orientations motivationnelles dans une scène de jeu. Selon les objectifs de la scène de jeu et les groupes de joueurs cibles. En s'inspirant des travaux de (Trabelsi & Frasson, 2010), nous avons utilisé des variables du modèle OCC pour décrire la scène du jeu. Nous avons alors construit des modèles de reconnaissance de l'émotion dominante ainsi que les buts motivationnels dans les scènes de jeu en validant avec les auto-réponses des joueurs dans les questionnaires à posteriori.

De plus, nous avons implémenté un outil qui permet au designer de définir des règles d'adaptations émotionnelles afin d'adapter le comportement des éléments de jeu selon l'état émotionnel du joueur. Le but est de donner au designer les outils nécessaires pour concevoir des jeux émotionnellement intelligents. Enfin, nous proposons dans cette thèse un modèle de design de jeux émotionnellement intelligents « *Emotional*-DDE» basé sur le modèle DDE de design de jeu en intégrant, dans ses couches, les différents outils réalisés dans cette thèse.

Chapitre 3 : Évaluation des émotions dans un environnement de débats en ligne

Dans ce chapitre, nous présentons une première étude réalisée dans le cadre d'un environnement de débat en ligne dans une plateforme[11] de discussions en ligne IRC[12]. Une session est composée de deux débats (2 sujets différents) d'une durée maximale de 20 minutes chacun. Nous avons sélectionné les sujets de débats parmi l'ensemble des débats populaires abordés dans les plateformes de forums de débat telles que iDebate[13] et DebateGraph[14]. Dans chaque session, il y a quatre participants et un modérateur. Au début du débat, le modérateur annonce le sujet à débattre et demande à chaque participant d'exprimer son opinion sur le sujet. Puis chacun des participants va discuter avec les autres participants afin de les convaincre de la validité de son point de vue. En cas de manque d'échanges actifs entre les participants, le modérateur intervient en proposant de nouveaux arguments afin de réanimer le débat. Chaque participant a été équipé d'un casque EEG et d'une caméra qui enregistre les expressions de son visage.

L'objectif général de l'étude est d'analyser la relation qui existe entre les émotions détectées chez les participants et le flux d'argumentation. L'idée est d'associer les arguments à l'engagement mental des participants détecté par le casque EEG et à leurs émotions détectées via l'outil de reconnaissance des expressions faciales par caméra (FaceReader). Plus précisément, à partir d'un sujet à débattre fourni par le modérateur, le but de l'expérience est de collecter les arguments proposés par les participants ainsi que les relations entre eux (Support/Attaque), et de les associer aux états d'engagement mental et aux émotions exprimées par les participants.

Lors du post-traitement des données collectées, nous avons synchronisé les arguments et leurs relations avec les mesures émotionnelles détectées. Ensuite, nous avons construit le

---

[11] http://webchat.freenode.net/

[12] Internet Relay Chat

[13] http://idebate.org/

[14] www.debategraph.org/

graphe d'argumentation bipolaire résultant pour chaque débat, de telle sorte que les graphes d'argumentation résultants soient étiquetés avec la source qui a proposé chaque argument et son état émotionnel. L'ensemble de données est composé de 598 arguments différents proposés par les participants à 12 débats différents. Pour chaque session, nous avons extrait pour chaque participant sa valeur d'engagement minimale (*min*), moyenne (*moy*) et maximale (*max*) afin de définir les trois niveaux d'engagements comme suit : « Bas » dans l'intervalle [min, (moy+min)/2[, « Moyen » dans l'intervalle [(moy+min)/2, (moy+max)/2[et « Haut » dans la plage [(moy+max)/2, max].

Dans cet article, nous voulons étudier la relation entre les arguments proposés par les participants dans un débat et leurs états émotionnels. Plus précisément, nous avons investigué les questions suivantes: (1) La polarité des arguments et les relations entre eux sont-elles corrélées à la polarité des émotions détectées?, (2) Quelle est la relation entre le nombre d'arguments (de support ou d'attaque) proposés dans un débat et le niveau d'engagement détecté chez les participants au débat?

L'analyse initiale des données recueillies, montre que les émotions les plus fréquentes sont la colère, puis le dégoût. Ceci peut être expliqué par la théorie de *negativity effect* (Rozin & Royzman, 2001), ce qui signifie que les émotions négatives dans un débat durent plus que les émotions positives. Nous avons également constaté qu'il existe une forte corrélation entre la colère et l'engagement. De plus, nous avons trouvé que le nombre d'attaques augmentait linéairement avec le dégoût, si le sujet du débat est conflictuel. Alors que si le sujet n'est pas conflictuel, nous avons une forte corrélation entre l'engagement et le nombre de support. Enfin, nous avons aussi noté que les participants les plus actifs dans un débat maintiennent un niveau fort d'engagement.

Le reste de ce chapitre est constitué de l'article intitulé « *Emotions in Argumentation: an Empirical Evaluation* » publié dans la conférence *International Joint Conferences on Artificial Intelligence, IJCAI 2015*. Nous rappelons que ma contribution essentielle consiste à la collecte des données, à l'analyse statistique de la relation les arguments, à l'interprétation des résultats des mesures émotionnelles et physiologiques et à la rédaction du papier.

Emotions in Argumentation: an Empirical Evaluation

Benlamine, M. S., Chaouachi, M., Villata, S., Cabrio, E., Frasson, C., & Gandon, F. (2015). Emotions in Argumentation: an Empirical Evaluation. *Proceedings of the Twenty- Fourth International Joint Conference on Artificial Intelligence, **IJCAI 2015***, Buenos Aires, Argentina, July 25-31, 2015. AAAI Press.

https://www.ijcai.org/Proceedings/15/Papers/029.pdf

Abstract. Argumentation is often seen as a mechanism to support different forms of reasoning such that decision-making and persuasion, but all these approaches assume a purely rational behavior of the involved actors. However, humans are proved to behave differently, mixing rational and emotional attitudes to guide their actions, and it has been claimed that there exists a strong connection between the argumentation process and the emotions felt by people involved in such process. In this paper, we assess this claim by means of an experiment: during several debates people's argumentation in plain English is connected and compared to the emotions automatically detected from the participants. Our results show a correspondence between emotions and argumentation elements, e.g., when in the argumentation two opposite opinions are conflicting this is reflected in a negative way on the debaters' emotions.

3.1 Introduction

The Web is becoming a hybrid space where men and machines interact. In this context, detecting and managing the emotional state of a user is important to allow artificial and human actors to adapt their reactions to others' emotional states. It is also a useful indicator for community managers, moderators and editors to help them in handling the communities and the content they produce. As a typical example, Wikipedia is managed by users and bots who constantly contribute, agree, disagree, debate and update the content of the encyclopedia. In this paper, we argue that in order to apply argumentation to scenarios like e-democracy and online debate systems, designers must take emotions into account. To efficiently manage and interact with such a hybrid society, we need to improve our means to understand and link the different dimensions of the exchanges (social interactions, textual content of the messages, dialogical structures of the interactions, emotional states of the participants, etc.). Beyond the challenges individually raised by each dimension, a key problem is to link these dimensions and their analysis together with the aim to detect, for instance, a debate turning into a flame war, a content reaching an agreement, a good or bad emotion spreading in a community.

In this paper, we aim to answer the following research question: *What is the connection between the arguments proposed by the participants of a debate and their emotional status*? Such question breaks down into the following sub-questions: *(1)* is the polarity of arguments and the relations among them correlated with the polarity of the detected emotions?, and *(2)* what is the relation between the kind and the amount of arguments proposed in a debate, and the mental engagement detected among the participants of the debate?

To answer these questions, we propose an empirical evaluation of the connection between argumentation and emotions. This paper describes an experiment with human participants which studies the correspondences between the arguments and their relations put forward during a debate, and the emotions detected by emotions recognition systems in the debaters. We design an experiment where 12 debates are addressed by 4 participants each. Participants argue in plain English proposing arguments that are in positive or negative relation with the arguments proposed by the other participants. During these debates, participants are equipped with emotions detection tools, recording their emotions. We hypothesize that negative relations among the arguments correspond to negative emotions felt by the participants proposing such arguments, and vice versa for the positive relation between arguments.

A key point in our work is that, up to our knowledge, no user experiment has been carried out yet to determine: *what is the connection between the argumentation addressed during a debate and the emotions emerging in the participants involved in such debate.* An important result is the development of a publicly available dataset capturing several debate situations, and annotating them with their argumentation structure and the emotional states automatically detected.

The paper is organized as follows. In Section 3.2 we describe the two main components of our framework (namely bipolar argumentation and emotions detection systems), and Section 3.3 describes the experimental protocol and the research hypotheses. In Section 3.4 we analyze our experimental results, and Section 3.5 compares this work with the relevant literature.

3.2 The Framework

In this section, we present the two main components involved in our experimental framework: *i)* bipolar argumentation theory, i.e., the formalism used to analyze the textual arguments retrieved from the debates (Section 3.2.1), and *ii)* the systems used to detect the degrees of attention and engagement of each participant involved in the debate as well as her facial emotions (Section 3.2.2).

3.2.1 Argumentation

Argumentation is the process of creating arguments for and against competing claims (Rahwan & Simari, 2009). What distinguishes argumentation-based discussions from other approaches is that opinions have to be supported by the arguments that justify, or oppose, them. This permits greater flexibility than in other decision-making and communication schemes since, for instance, it makes it possible to persuade other persons to change their view of a claim by identifying information or knowledge that is not being considered, or by introducing a new relevant factor in the middle of a negotiation, or to resolve an impasse.

Argumentation is the process by which arguments are constructed and handled. Thus argumentation means that arguments are compared, evaluated in some respect and judged in order to establish whether any of them are warranted. Roughly, each argument can be defined as a set of assumptions that, together with a conclusion, is obtained by a reasoning process. Argumentation as an exchange of pieces of information and reasoning about them involves groups of actors, human or artificial. We can assume that each argument has a proponent, the person who puts forward the argument, and an audience, the person who receives the argument. In our framework, we rely on abstract bipolar argumentation (Cayrol & Lagasquie-Schiex, 2013; Dung, 1995) where we do not distinguish the internal structure of the argument (i.e., premises, conclusion), but we consider each argument proposed by the participants in the debate as a unique element, then analyzing the relation it has with the other pieces of information put forward in the debate. In particular, in bipolar argumentation two kinds of relations among arguments are considered: the *support relation*, i.e., a positive relation between arguments, and the *attack relation*, i.e., a negative relation between arguments.

3. 2.2 Emotion Detection

Emotions play an important role in decision making (Quartz, 2009), creative thinking, inspiration, as well as concentration and motivation. During conversations, emotions are expressed by participants according to their beliefs and viewpoints with respect to the opinions put forward by the other participants. In the argumentation process, how they feel about the others' point of view influences their reply, being it a support or an attack. To capture these different emotional reactions through automatic emotions recognition systems, cameras and/or physiological sensors (e.g., EDA[15], EEG[16], EMG[17]) can be used. Many researches propose to apply multimodal techniques, i.e. to combine different media in order to improve the automatic emotions recognition (Calvo & D'Mello, 2010; Jraidi, Chaouachi, & Frasson, 2013). For these reasons, in our experiments we have used webcams for facial expressions analysis relying on

[15] Electrodermal Activity: electrical changes measured at the surface of the skin.

[16] Electroencephalography: recording of electrical activity along the scalp.

[17] Electromyography: recording the electrical activity produced by skeletal muscles.

the FACEREADER 6.0 software[18], and physiological sensors (EEG) to assess and monitor users' cognitive states in real-time (Chaouachi et al., 2010b).

Emotiv EPOC EEG headset. It has been used to record physiological data during the debate sessions. It contains 14 electrodes spatially organized according to International 10 - 20 system[19], moist with a saline solution (contact lens cleaning solution). As shown in Figure 12, the electrodes are placed at AF3, AF4, F3, F4, F7, F8, FC5, FC6, P7, P8, T7, T8, O1, O2, and two other reference sensors are placed behind the ears. The generated data are in (μV) with a 128Hz sampling rate. The signal frequencies are between 0.2 and 60Hz.

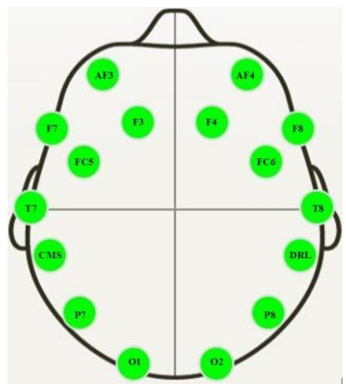


Figure 12. Emotiv Headset sensors/data channels placement.

Computing Engagement Index. The engagement index used in our study comes from the work of (Pope et al., 1995a) at the National Aeronautics and Space Administration (NASA). This work is based on neuroscientific research on attention and vigilance (Lubar, 1991). It was found that the user's performance improved when this index is used as a criterion for switching between manual and automated piloting mode (Freeman et al., 2000; Pope et al., 1995a). Many studies showed the usefulness of integrating this index in many fields like eLearning and games to assess user's cognitive states. This index is computed from three EEG frequency bands: ϴ (4 - 8Hz), α (8 - 13Hz) and β (13 - 22Hz) as follows: $Eng = \frac{\theta}{(\alpha + \beta)}$. This index is computed each second from the EEG signal. To reduce its fluctuation, we use a moving average on a 40-second

---

[18] http://www.noldus.com/human-behavior-research/products/facereader

[19] International 10 - 20 system is an internationally recognized method to describe and apply the location of scalp electrodes in the context of an EEG test or experiment.

mobile window. Thus, the value of the index at time *t* corresponds to the total average of the ratios calculated on a period of 40 seconds preceding *t*. The extraction of the Θ, α and β frequency bands is performed by multiplying every second of the EEG signal by a Hamming window (to reduce the spectral leakage) and applying a Fast Fourier Transform (FFT). As the Emotiv headset measures 14 regions at the same time, we use a combined value of Θ, α and β frequency bands by summing their values over all the measured regions. To examine participants' engagement, we extract their minimum, average and maximum values during the debate, and we use such values to identify the range of engagement (High, Medium, and Low) of every participant.

Facial expressions analysis. The primary emotion facial expressions (Ekman, 2005) from the user's emotional reactions are identified using real-time frame-by-frame analysis software FaceReader 6.0 via a webcam. FaceReader infers the emotional state by extracting and classifying in real-time 500 key points in facial muscles of the target face. These key points are provided as input to a neural network trained on a dataset of 10000 manually annotated images corresponding to these six basic emotions: *Happy*, *Sad*, *Angry*, *Surprised*, *Scared* and *Disgusted*. In addition to these emotions, the resulting file contains the *Valence*[20], and the Arousal of emotion as well as the probability of the Neutral state. In this study, to align the dominant emotion state occurring at every second, we compute the average (10 values/sec) for each column according to the camera frame rate[21]. We extract the most dominant emotion having the maximum value, pleased/unpleased valence depending on positive/negative values, and the active/ inactive arousal by comparing the obtained values to 0.5.

3.3 The Experiment

This section details the experimental session we set up to analyze the relation between emotions and the argumentation process: we detail the protocol we have defined to guide the experimental

[20] The valence refers to the degree of pleasantness of an expressed emotion. A positive valence corresponds to an emotion with pleasant character and a negative valence to an unpleasant one.

[21] We used cameras of 10 frames/sec.

setting (Section 3.3.1), and the resulting datasets (Section 3.3.2). Finally, we specify the hypotheses we aim to verify in this experiment (Section 3.3.3).

3.3.1 Protocol

The general goal of the experimental session is to study the relation (if any) holding between the emotions detected in the participants and the argumentation flow. The idea is to associate arguments and the relations among them to the participants' mental engagement detected by the EEG headset and the facial emotions detected via the Face Emotion Recognition tool. More precisely, starting from an issue to be discussed provided by the moderators, the aim of the experiment is to collect the arguments proposed by the participants as well as the relations among them, and to associate such arguments/relations to the mental engagement states and to the facial emotions expressed by the participants. During a post-processing phase on the collected data, we synchronize the arguments and the relations put forward by the different participants at instant t with the emotional indexes we retrieved. Finally, we build the resulting bipolar argumentation graph for each debate, such that the resulting argumentation graphs are labelled with the source who has proposed each argument, and the emotional state of each participant at the time of the introduction of the argument in the discussion.

The first point to clarify in this experimental setting is the terminology. In this experiment, an argument is each single piece of text that is proposed by the participants of the debate. Typically, arguments have the goal to promote the opinion of the debater in the debate. Thus, an *opinion* in our setting represents the overall opinion of the debater about the issue to be debated, i.e., "Ban animal testing". The opinion is promoted in the debate through arguments, that will support (if the opinions converge) or attack (otherwise) the arguments proposed in the debate by the other participants.

The experiment involves two kinds of persons:

- Participant: she is expected to provide her own opinion about the issue of the debate proposed by the moderators, and to argue with the other participants in order to convince them (in case of initial disagreement) about the goodness of her viewpoint[22].
- Moderator: she is expected to propose the initial issue to be discussed to the participants. In case of lack of active exchanges among the participants, the moderator is in charge of proposing pro and con arguments (with respect to the main issue) to reactivate the discussion.

The experimental setting of each debate is conceived as follows: there are 4 participants for each discussion group (each participant is placed far from the other participants, even if they are in the same room), and 2 moderators located in another room with respect to the participants. The moderators interact with the participants uniquely through the debate platform. The language used for debating is English. In order to provide an easy-to-use debate platform to the participants, without requiring from them any background knowledge, we decide to rely on a simple IRC network[23] as debate platform. The debate is anonymous and participants are visible to each other with their nicknames, e.g., *participant1*, while the moderators are visualized as *moderator1* and *moderator2*. Each participant has been provided with 1 laptop device equipped with internet access and a camera used to detect facial emotions. Moreover, each participant has been equipped with an EEG headset to detect engagement index. Each moderator used only a laptop.

The procedure we followed for each debate is:

- Participants' familiarization with the debate platform;
- The debate - participants take part into two debates each, about two different topics for a maximum of about 20 minutes each:
  - The moderator(s) provides the debaters with the topic to be discussed;

[22] Note that in this experimental scenario we do not evaluate the connection between emotions and persuasive argumentation. This analysis is left for future research.

[23] http://webchat.freenode.net/

- The moderator(s) asks each participant to provide a general statement about his/her opinion concerning the topic;
- Participants expose their opinion to the others;
- Participants are asked to comment on the opinions expressed by the other participants;
- If needed (no active debate among the participants), the moderator posts an argument and asks for comments from the participants;

- Debriefing: each participant is asked to complete a short questionnaire about his/her experience in the debate[24].

The measured variables in the debate are: engagement (measurement tool: EEG headset), and the following emotions: Neutral, Happy, Sad, Angry, Surprised, Scared and Disgusted (measurement tool: FaceReader).

The post-processing phase of the experimental session involved *(i)* the detection of the support and attack relations among the arguments proposed in each discussion, following the methodology described in Section 3.3.2, and *(ii)* the synchronization of the argumentation (i.e., the arguments/relations proposed at time *t*) with the emotional indexes retrieved at time *t* using the EEG headset and FaceReader.

Participants. The experiment was distributed over 6 sessions of 4 participants each; the first session was discarded due to a technical problem while collecting data. We had a total of 20 participants (7 women, 13 men), whose age range was from 22 to 35 years. All of them were students in a North American university, and all of them had good computer skills. Since not all of them were native English speakers, the use of the Google translate service was allowed. They have all signed an ethical agreement before proceeding to the experiment.

3.3.2 Dataset

In this section we describe the dataset of textual arguments we have created from the debates among the participants. The dataset is composed of three main layers: *(i)* the basic annotation of the arguments proposed in each debate (i.e. the annotation in xml of the debate flow downloaded from the debate platform); *(ii)* the annotation of the relations of support and attack

[24] Such material is available at http://bit.ly/DebriefingData

among the arguments; and *(iii)* starting from the basic annotation of the arguments, the annotation of each argument with the emotions felt by each participant involved in the debate. The basic dataset is composed of 598 different arguments proposed by the participants in 12 different debates. The debated issues and the number of arguments for each debate are reported in Table 2. We selected the topics of the debates among the set of popular discussions addressed in online debate platforms like iDebate[25] and DebateGraph[26]. The annotation (in xml) of this dataset is as follows: we have assigned to each debate a unique numerical id, and for each argument proposed in the debate we assign an id and we annotate who was the participant putting this argument on the table, and in which time interval the argument has been proposed. An example of basic annotation is provided below:

<debate id="1" title="Ban_Animal_Testing">

<argument id="1" debate_id="1" participant="mod" time-from="19:26" time-to="19:27">Welcome to the first debate! The topic of the first debate is that animal testing should be banned.</argument>

<argument id="3" debate_id="1" participant="2" time-from="20:06" time-to="20:06">If we don’t use animals in these testing, what could we use?</argument>

</debate>

The second level of our dataset consists in the annotation of arguments pairs with the relation holding between them, i.e., support or attack. To create the dataset, for each debate of our experiment we apply the following procedure, validated in (Cabrio & Villata, 2013):

1. the main issue (i.e., the issue of the debate proposed by the moderator) is considered as the starting argument;
2. each opinion is extracted and considered as an argument;
3. since attack and support are binary relations, the arguments are coupled with:
   a) the starting argument, or

[25] http://idebate.org/

[26] www.debategraph.org/

b) other arguments in the same discussion to which the most recent argument refers (e.g., when an argument proposed by a certain user supports or attacks an argument previously expressed by another user);

4. the resulting pairs of arguments are then tagged with the appropriate relation, i.e., attack or support.

To show a step-by-step application of the procedure, let us consider the debated issue Ban Animal Testing. At step 1, we consider the issue of the debate proposed by the moderator as the starting argument (a):

(a) *The topic of the first debate is that animal testing should be banned*.

Then, at step 2, we extract all the users' opinions concerning this issue (both pro and con), e.g., (b), (c) and (d):

(b) *I don't think the animal testing should be banned, but researchers should reduce the pain to the animal*.

(c) *I totally agree with that*.

(d) *I think that using animals for different kind of experience is the only way to test the accuracy of the method or drugs. I cannot see any difference between using animals for this kind of purpose and eating their meat*.

(e) *Animals are not able to express the result of the medical treatment but humans can*.

At step 3a we couple the arguments (b) and (d) with the starting issue since they are directly linked with it, and at step 3b we couple argument (c) with argument (b), and argument (e) with argument (d) since they follow one another in the discussion. At step 4, the resulting pairs of arguments are then tagged by one annotator with the appropriate relation, i.e.: (b) *attacks* (a), (d) *attacks* (a), (c) *supports* (b) and (e) *attacks* (d). For the purpose of validating our hypotheses, we decide to not annotate the supports/attacks between arguments proposed by the same participant (e.g., situations where participants are contradicting themselves). Note that this does not mean that we assumed that such situations do not arise: no restriction was imposed to the participants of the debates, so situations where a participant attacked/supported her own arguments are represented in our dataset. We just decided to not annotate such cases in the dataset of argument pairs, as it was not necessary for verifying our assumptions.

To assess the validity of the annotation task and the reliability of the obtained dataset, the same annotation task has been independently carried out also by a second annotator, so as to compute inter-annotator agreement. It has been calculated on a sample of 100 argument pairs (randomly extracted). The complete percentage agreement on the full sample amounts to 91%. The statistical measure usually used in NLP to calculate the inter-rater agreement for categorical items is Cohen's kappa coefficient (Carletta, 1996), that is generally thought to be a more robust measure than simple percent agreement calculation since κ takes into account the agreement occurring by chance. More specifically, Cohen's kappa measures the agreement between two raters who each classifies N items into C mutually exclusive categories. The equation for κ is:

$$\kappa = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(\mathrm{e})}$$

where Pr(a) is the relative observed agreement among raters, and Pr(e) is the hypothetical probability of chance agreement, using the observed data to calculate the probabilities of each observer randomly saying each category. If the raters are in complete agreement then $\kappa = 1$. If there is no agreement among the raters other than what would be expected by chance (as defined by Pr(e)), $\kappa = 0$. For NLP tasks, the inter-annotator agreement is considered as significant when $\kappa > 0.6$. Applying such formula to our data, the interannotator agreement results in $\kappa = 0.82$. As a rule of thumb, this is a satisfactory agreement, therefore we consider these annotated datasets as reliable (i.e., our *goldstandard* dataset where arguments are associated to participants' emotions detected by EEG/FaceReader) to be exploited during the experimental phase.

Table 2. The textual dataset of the experiment.

Dataset				
Topic	**#arg**	**#pair**	**#att**	**#sup**
BAN ANIMAL TESTING	49	28	18	10
GO NUCLEAR	40	24	15	9
HOUSEWIVES SHOULD BE PAID	42	18	11	7
RELIGION DOES MORE HARM THAN GOOD	46	23	11	12
ADVERTISING IS HARMFUL	71	16	6	10
BULLIES ARE LEGALLY RESPONSIBLE	71	12	3	9
DISTRIBUTE CONDOMS IN SCHOOLS	68	27	11	16
ENCOURAGE FEWER PEOPLE TO GO TO THE UNIVERSITY	55	14	7	7
FEAR GOVERNMENT POWER OVER INTERNET	41	32	18	14
BAN PARTIAL BIRTH ABORTIONS	41	26	13	11

USE RACIAL PROFILING FOR AIRPORT SECURITY	31	10	1	9
CNABIS SHOULD BE LEGALIZED	43	33	20	13
TOTAL	598	263	136	127

Table 2 reports on the number of arguments and pairs we extracted applying the methodology described before to all the mentioned topics. In total, our dataset contains 598 different arguments and 263 argument pairs (127 expressing the support relation among the involved arguments, and 136 expressing the attack relation among the involved arguments).

The final dataset adds to all previously annotated information the player characteristics (gender, age and personality type), FaceReader data (dominant emotion, Valence (pleasant/unpleased) and Arousal (activated/ inactivated)), and EEG data (Mental Engagement levels)[27]. A correlation matrix has been generated to identify the correlations between arguments and emotions in the debates, and a data analysis is performed to determine the proportions of emotions for all participants. We consider the obtained dataset as representative of human debates in a non-controlled setting, and therefore we consider it as the reference dataset to carry out our empirical study. An example, from the debate about the topic "Religion does more harm than good" where arguments are annotated with emotions (i.e., the third layer of the annotation of the textual arguments we retrieved), is as follows:

<argument id="30" debate_id="4" participant="4" time-from="20:43" time-to="20:43" emotion_p1="neutral" emotion_p2="neutral" emotion_p3="neutral" emotion_p4="neutral">

Indeed but there exist some advocates of the devil like Bernard Levi who is decomposing arabic countries.

</argument>

<argument id="31" debate_id="4" participant="1" time-from="20:43" time-to="20:43" emotion_p1="angry" emotion_p2="neutral" emotion_p3="angry" emotion_p4="disgusted">

I don't totally agree with you Participant2: science and religion don't explain each other, they tend to explain the world but in two different ways.

</argument>

[27] The datasets of textual arguments are available at http://bit.ly/TextualArgumentsDataset.

<argument id="32" debate_id="4" participant="3" time-from="20:44" time-to="20:44" emotion_p1="angry" emotion_p2="happy" emotion_p3="surprised" emotion_p4="angry">
Participant4: for recent wars ok but what about wars happened 3 or 4 centuries ago?
</argument>

3.3.3 Hypotheses

This experiment aims to verify the link between the emotions detected on the participants of the debate, and the arguments and their relations proposed in the debate. Our hypotheses therefore revolve around the assumption that the participants' emotions arise out of the arguments they propose in the debate:

H1: There are some emotional and behavioral trends that can be extracted from a set of debates.

H2: The number and the strength of arguments, attacks and supports exchanged between the debaters are correlated with particular emotions throughout the debates.

H3: The number of expressed arguments is connected to the degree of mental engagement and social interactions.

3.4 Results

In order to verify these hypotheses, we first computed the mean percentage of appearance of each basic emotion across the 20 participants. Results show (with 95% confidence interval) that the most frequent emotion expressed by participants was anger, with a mean appearance frequency ranging from 8.15% to 15.6% of the times. The second most frequent emotion was another negative emotion, namely disgust, which was present 7.52% to 14.8% of the times. The overall appearance frequency of other emotions was very low. For example, the frequency of appearance of happiness was below 1%. Even if this result might be surprising at a first glance, this trend can be justified by a phenomenon called negativity effect (Rozin & Royzman, 2001). This means that negative emotions have generally more impact on a person's behavior and cognition than positive ones. So, negative emotions like anger and disgust have a tendency to last in time more than positive emotions like happiness.

With regard to the mental engagement, participants show in general a high level of attention and vigilance in 70.2% to 87.7% of the times. This high level of engagement is also correlated with appearance of anger (r=0.306), where r refers to the Pearson product-moment correlation coefficient. It is a standard measure of the linear correlation between two variables X and Y, giving a value between [1,-1], where 1 is a total positive correlation, 0 means no correlation, and -1 is a total negative correlation. This trend confirms that, in such context, participants may be thwarted by the other participants' arguments or attacks, thus the level of engagement tends to be high as more attention is allowed to evaluate the other arguments or to formulate rebutting or offensive arguments. Thus, our experiments confirm behavioral trends as expected by the first hypothesis.

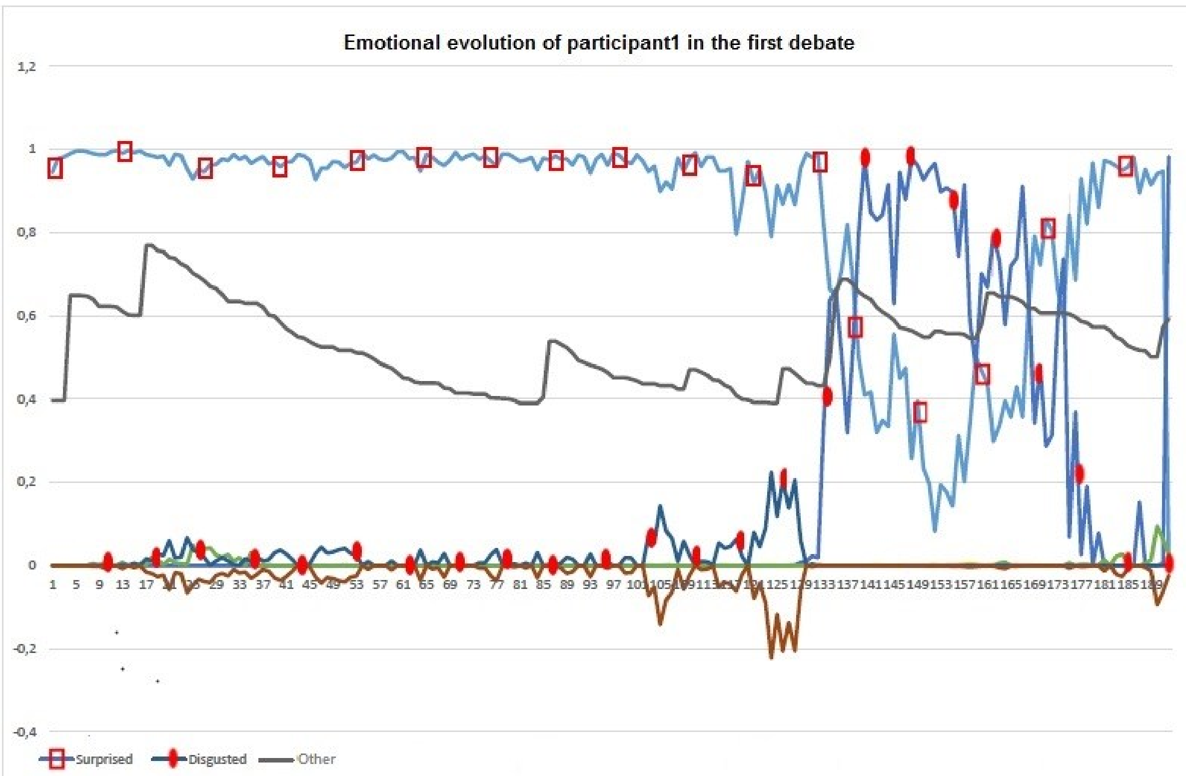


Figure 13. Emotional evolution of participant 1 in debate 1.

Figure 13 shows an evolution of the first participant's emotions at the beginning of the first debate. The most significant lines of emotions are surprise and disgust (respectively, the line with squares and the line with circles). The participant is initially surprised by the discussion (and so mentally engaged) and then, after the debate starts, this surprise switches suddenly into disgust, due to the impact of the rejection of one of her arguments; the bottom line with circles grows and replaces the surprise as the participant is now actively engaged in an opposed argument (thus confirming our hypothesis 2). Finally, the participant is calming down. In this line, Table 3 highlights that we have a strong correlation (r= 0.83) in Session 2 showing that the number of attacks provided in the debate increased linearly with the manifestation of more disgust emotion.

Table 3. Correlation table for Session 2 (debated topics: *Advertising is harmful* and *bullies are legally responsible*.

	NB ARG	**ATTACK**	**SUPPORT**
Pleasant	0,0962	0,1328	-0,0332
Unpleasant	-0,0962	-0,1328	0,0332
High ENG	-0,0718	**-0,6705**	0,2459
Low ENG	-0,2448	0,2115	-0,1063
Neutral	0,0378	0,6173	-0,1138
Disgusted	-0,0580	**-0,4367**	-0,3621
Scared	0,1396	-0,0952	**0,5755**
Angry	-0,1018	-0,4386	0,0582

In the second part of our study, we were interested in analyzing how emotions correlate with the number of attacks, supports and arguments. We have generated a correlation matrix to identify the existent correlations between arguments and emotions in debates. Main results show that the number of arguments tends to decrease linearly with manifestations of sadness (r=-0.25). So when the participants start to feel unpleasant emotions, such as sadness, the number of arguments decreases, showing a less positive social behavior[28] and a tendency to retreat into herself.

[28] By positive social behavior, we mean that a participant aims at sharing her arguments with the other participants. This attitude is mitigated if unpleasant emotions start to be felt by the participant.

Table 4. Correlation table for Session 3 (debated topics: *Distribute condoms at schools* and *Encourage fewer people to go to the university*).

	NB ARG	ATTACK	SUPPORT
Pleasant	**0,7067**	-0,3383	-0,3800
Unpleasant	-0,7067	0,3383	0,3800
High ENG	**-0,6903**	-0,3699	-0,1117
Low ENG	-0,1705	**0,5337**	-0,0615
Neutral	0,8887	-0,0895	-0,3739
Disgusted	0,1017	**0,8379**	**0,5227**
Scared	0,2606	-0,4132	**-0,7107**
Angry	**-0,7384**	-0,5072	-0,0937

This negative correlation between the number of arguments and sadness even reaches very high level in certain debates (i.e. a mean correlation r= -0.70 is registered in the two debates of the second session). Another negative linear relationship is registered with regard to the number of attacks and the anger expressed by the participant (r=0.22). Participants who tend to attack the others in the debate are less angry than those whose number of attacks is smaller. Table 4 shows the correlation table for Session 3. The analysis of the results we obtained shows the occurrence of strong correlations between emotions and attacks / media / number of arguments in some discussions, but not in others. This is an interesting index to investigate in future work.

Table 5. General correlation table of the results.

	NB ARG	ATTACK	SUPPORT
Pleasant	**0,1534**	0,0134	-0,0493
Unpleasant	-0,1534	-0,0134	0,0493
High ENG	-0,0246	-0,0437	**0,3185**
Low ENG	**0,2054**	0,1147	0,1592
Neutral	0,0505	0,1221	-0,2542
Disgusted	-0,0177	-0,0240	**0,2996**
Scared	-0,0278	0,0297	-0,2358
Angry	0,0344	**-0,2206**	0,0782

Table 5 shows the most significant correlations we detected. For instance, the number of supports provided in the debate increased linearly with the manifestation of high levels of mental engagement (r=0.31). This trend is more pronounced when the debate does not trigger controversies and conflicts between the participants. For example, in the debate *Encourage fewer people to go to university*, all the participants shared the same opinion (against the main issue as formulated by the moderator) and engaged to support each other's arguments. The correlation between the number of supports and the engagement was very high (r=0.80) in this debate. The number of attacks is more related to low engagement. The moderator can provide more supporting arguments to balance participants' engagement, and if the attacks are increasing, that means participants tend to disengage. The experiments show that participants maintaining high levels of vigilance are the most participative in the debate and resulted in a more positive social behavior (thus confirming our hypothesis 3).

3.5 Related Work

Cerutti et al. (Cerutti, Tintarev, & Oren, 2014) propose an empirical experiment with humans in the argumentation theory area. However, the goal of this work is different from our one, emotions are not considered and their aim is to show a correspondence between the acceptability of arguments by human subjects and the acceptability prescribed by the formal theory in argumentation. Rahwan et al. (Rahwan et al., 2010) study whether the meaning assigned to the notion of reinstatement in abstract argumentation theory is perceived in the same way by humans. They propose to the participants of the experiment a number of natural language texts where reinstatement holds, and then asked them to evaluate the arguments. Also in this case, the purpose of the work differs from our one, and emotions are not considered at all.

Emotions are considered, instead, by (Nawwab, Dunne, & Bench-Capon, 2010) that propose to couple the model of emotions introduced by (A Ortony, Clore, & Collins, 1988) in an argumentation-based decision making scenario. They show how emotions, e.g., gratitude and displeasure, impact on the practical reasoning mechanisms. A similar work has been proposed by (Martinez & Simari, 2012) where emotions are exploited by agents to produce a line of reasoning according to the evolution of its own emotional state. Finally, (Lloyd-Kelly & Wyner, 2011) propose emotional argumentation schemes to capture forms of reasoning involving emotions. All these works differ from our approach since they do not address an empirical evaluation of their models, and emotions are not detected from humans. Several works in philosophy and linguistics have studied the link between emotions and natural argumentation, like (Carofiglio & de Rosis, 2003; Fulkerson, 1993; Gilbert, 1995). These works analyze the connection of emotions and the different kind of argumentation that can be addressed. The difference with our approach is that they do not verify their theories empirically, on actual emotions extracted from people involved in an argumentation task. A particularly interesting case is that of the connection between persuasive argumentation and emotions, studied for instance by (DeSteno et al., 2004).

3.6 Conclusions

In this paper, we presented an investigation into the link between the argumentation people address when they debate with each other, and the emotions they feel during these debates. We conducted an experiment aimed at verifying our hypotheses about the correlation between the positive/ negative emotions emerging when positive/negative relations among the arguments are put forward in the debate. The results suggest that there exist clear trends that can be extracted from emotions analysis. Moreover, we also provide the first open dataset and gold standard to compare and analyze emotion detection in an argumentation session.

Several lines of research have to be considered as future work. First, we intend to study the link between emotions and persuasive argumentation. This issue has already been tackled in a number of works in the literature (e.g., (DeSteno et al., 2004)), but no empirical evaluation has been addressed yet. Second, we aim to study how emotion persistence influence the attitude of the debates: this kind of experiment has to be repeated a number of times in order to verify whether positive/ negative emotions before the debate influence the new interactions. Third, we plan to add a further step, namely to study how sentiment analysis techniques are able to automatically detect the polarity of the arguments proposed by the debaters, and how they are correlated with the detected emotions. Moreover, we plan to study emotions propagation among the debaters, and to verify whether the emotion can be seen as a predictor of the solidity of an argument, e.g., if I write an argument when I am angry I may make wrong judgments.

Acknowledgment. The authors acknowledge support of the SEEMPAD associate team project (http://project.inria.fr/seempad/).

Note: This work has been supported from the Canadian side by the FRQNT (Fonds de Recherche du Québec Nature et Technologie) and NSERC (National Science and Engineering Research Council).

Chapitre 4 : Évaluation de la persuasion dans un environnement de débats en ligne

Dans le chapitre précédent, nous avons présenté notre première étude sur les états émotionnels et mentaux et leurs relations avec les arguments dans un environnement de débat en ligne. Ce chapitre est une extension et continuation de ces travaux. Cette fois-ci, nous analysons l'effet des stratégies d'argumentation sur les émotions et l'engagement des participants et leurs persuasions (changement d'attitude). Nous avons toujours une session de deux débats, portant sur deux sujets différents, d'une durée maximale de 20 minutes par débat. Nous avons présélectionné les participants selon leurs réponses à un questionnaire en ligne sur leurs opinions initiales à propos de 10 sujets. Nous avons essayé d'avoir des débats équilibrés avec deux participants pour le sujet de débat et deux contre. Chaque participant était équipé d'un casque d'électroencéphalographie (EEG) pour détecter l'engagement mental et de caméras pour détecter les émotions du visage. Dans chaque débat, il y a un participant qui joue le rôle de persuadeur, appelé le PP (participant persuadeur) dans la suite du document. Le PP adopte et maintient un point de vue prédéfini dans le débat (c'est-à-dire, favorable ou défavorable), ainsi qu'une stratégie de persuasion (c'est-à-dire, *Logos*, *Pathos* ou *Ethos*). Les arguments proposés par le PP étaient prédéfinis selon les trois stratégies de persuasion avec légère adaptation à la conversation. PP n'essaye pas de dominer le débat pour apparaître aux autres participants comme un simple autre participant.

Dans cette étude, nous voulons analyser l'impact des stratégies de persuasion (*Logos*, *Pathos* et *Ethos*) sur les états mentaux et les émotions des participants du débat. Plus précisément, nous avons investigué les questions suivantes: (1) Quel est l'effet des stratégies de persuasion (Ethos, Logos et Pathos) sur les émotions négatives et l'engagement selon la position finale du participant (Supporter, Opposant ou Neutre)?, (2) Quel est l'impact des stratégies de persuasion (Ethos, Logos, pathos) sur l'activation des lobes cérébraux (Frontal, Pariétal, Temporal et Occipital)? Et (3) Quelle stratégie de persuasion génère plus de support que d'attaque dans le débat et provoque un changement d'opinion pour les participants?

Pour répondre à ces questions, nous avons procédé en trois étapes : (1) collecter les données des participants pour les différents senseurs (Caméra et EEG), (2) structurer les débats en trois phases (Introduction, Argumentation et Conclusion), synchroniser les arguments avec les mesures affectives et annoter les relations entre les arguments (support/attaque), (3) analyser statistiquement les résultats. En s'inspirant de la structure des conversations en pragmatique (initiation, maintenance et terminaison (Kellermann et al., 1989)), nous avons structuré le débat en trois phases: l'introduction où le PP exprime sa propre opinion sur le sujet du débat; l'argumentation qui inclut la reformulation, la réfutation et l'apport de nouvelles idées en fonction de la stratégie adoptée par le PP; la conclusion où le PP rappelle sa position et son avis final. Pour la synchronisation des données, nous avons calculé les valeurs d'émotion moyennes des 10 secondes après chaque proposition d'argument. Pour l'annotation, chaque argument a été annoté avec l'identificateur de débat, l'id de l'argument, le participant et ses relations avec les autres arguments. Au total, 791 arguments et 162 paires d'arguments (74 liées par une attaque et 88 par un support) ont été annotées. En plus de l'engagement, nous avons pris en compte les scores de colère dans l'analyse des résultats car il s'agissait de l'émotion la plus prédominante lors des débats.

Dans un premier temps, pour vérifier l'impact des stratégies de persuasion (Ethos, Logos, pathos) sur les mesures de colère et d'engagement, nous avons effectué une analyse ANOVA à mesures répétées sur les trois phases du débat (Introduction, Argumentation et Conclusion). Dans cette analyse de variance de ces mesures, nous avons étudié leurs variations selon la position finale du participant (Supporter, Opposant ou Neutre). Nous avons comparé la variation des deux mesures (Engagement et Colère) selon la phase du débat et la position finale du participant pour chacune des stratégies de persuasion. Nous avons constaté que la stratégie de persuasion qui a le niveau d'engagement le plus élevé, est la stratégie Logos pour les participants avec position finale neutre (resté indécis tout au long du débat). Les analyses de variances ont montré aussi que l'engagement et la colère dépendent de la stratégie de persuasion et de la position finale du participant.

Dans un deuxième temps, nous avons étudié l'effet des stratégies de persuasion (Ethos, Logos, pathos) sur l'activation des différents lobes (Frontal, Pariétal, Temporal et Occipital), nous avons effectué une analyse ANOVA à mesures répétées sur les trois phases du débat (Introduction, Argumentation et Conclusion). Nous avons comparé la variation des mesures d'Engagement moyen dans les quatre lobes du cerveau (Frontal, Pariétal, Temporal et Occipital) selon la phase du débat pour chacune des stratégies de persuasion. Les analyses de variances ont montré aussi que la mesure d'engagement dans les lobes du cerveau dépend de la stratégie de persuasion. En effet, nous avons trouvé que pour les stratégies Logos et Ethos, le lobe du cerveau le plus activé est le pariétal alors que pour la stratégie Pathos c'est le lobe frontal.

Dans un troisième temps, nous avons analysé pour chaque stratégie de persuasion le pourcentage d'arguments de support et d'attaque et aussi le pourcentage de participants qui ont changé d'opinions par comparaison avant et après le débat. Nous avons constaté que la stratégie de persuasion qui a le pourcentage de support le plus élevé et le pourcentage de changement d'avis le plus élevé, est la stratégie Pathos.

Le reste de ce chapitre est constitué de l'article intitulé « *Assessing Persuasion in Argumentation through Emotions and Mental States* » publié à la conférence International Florida Artificial Intelligence Research Society conference, Flairs 2018. La première auteure a contribué à la préparation des arguments selon la stratégie de persuasion prédéfinie par débat, à l'annotation des arguments selon le modèle d'argumentation, l'analyse de la relation entre les stratégies de persuasion le nombre de support et d'attaque dans un débat et le changement d'opinion et à la rédaction. Nous rappelons que ma contribution essentielle consiste à la conception de l'expérimentation, à la collecte des données, à l'analyse de l'effet des stratégies de persuasion sur la variation des mesures émotionnelles et physiologiques et à la rédaction du papier.

Assessing Persuasion in Argumentation through Emotions and Mental States

Villata, S., **Benlamine**, M. S., Cabrio, E., Frasson, C., & Gandon, F. (2018). Assessing persuasion in argumentation through emotion and mental states. The *31st Florida Artificial Intelligence Research Society Conference, **FLAIRS 2018***, Melbourne, Florida, USA, May 21-23, 2018. AAAI Press.

https://aaai.org/ocs/index.php/FLAIRS/FLAIRS18/paper/view/17677/16868

Abstract. Argumentative persuasion usually employs one of the three persuasion strategies: Ethos, Pathos or Logos. Several approaches have been proposed to model persuasive agents, however, none of them explored how the choice of a strategy impacts the mental states of the debaters and the argumentation process. We conducted a field experiment with real debaters to assess the impact of the mental engagement and emotions of the participants, as well as of the persuasiveness power of the arguments exchanged during the debate. Our results show that the Pathos strategy is the most effective in terms of mental engagement.

4.1 Introduction

In everyday life situations like online discussions and political debates, "the aim of persuasion is for the persuader to change the mind of the persuadee" (Hunter, 2016). This process, called *persuasive argumentation*, may employ different strategies. In the Ethos strategy, persuasion relies on the authority of the persuader with respect to the topic of the debate. The Logos strategy is grounded on logical arguments leading to a sound inference process to derive conclusions, while the Pathos strategy solicits the emotions of the interlocutors to generate empathy. These strategies have been used to define formal models of persuasion, e.g., (Hunter, 2016), to be employed by intelligent agents to persuade the others to change their beliefs. However, analyzing how these strategies are perceived by humans when they argue, and what is the impact of these strategies on the humans' mental states like *engagement* and *emotions* has not been explored. Yet, this would be of valuable importance for argumentative agents to be able to apply persuasion strategies as humans do, resulting in more effective interactions with people.

In this paper, we answer the following research question: *what is the impact of persuasion strategies on the mental states and emotions of the debaters?* To answer, we conducted a field experiment with users, starting from three hypotheses to be validated. We raised a number of debates in which, together with the participants of the experiment, a persuader was involved to convince the other participants about the goodness of her viewpoint, applying one of the three persuasion strategies. The persuader is a person who has been provided with particular argumentation frameworks but appears to the other participants as just another participant, e.g., she does not dominate the debate. Every participant was equipped with an Electroencephalography (EEG) Headset to detect mental engagement, and cameras to detect facial emotions. The collected data was synchronized to assess the validity of our hypotheses. Results highlight the higher persuasion impact of the Pathos strategy.

4.2 Preliminaries

4.2.1 Argumentative persuasion.

In computational models of argument (Rahwan & Simari, 2009) , arguments are linked to each other by *attacks*, indicating that an argument is incompatible with another one, and *supports*, indicating an argument provides some backing to another. Three kinds of argumentative persuasion exist: *Ethos*, *Logos*, and *Pathos* (Ross, 2010). Ethos deals with the character of the speaker, whose intent is to appear credible. The main influencing factors for Ethos encompass elements such as vocabulary, and social aspects like rank or popularity. Logos is the appeal to logical reason: the speaker wants to present an argument that appears to be sound to the audience. Pathos encompasses the emotional influence on the audience.

4.2.2 Mental states and emotions.

To assess both participants' mental condition and their involvement in the argumentation, we adopt the engagement index (Chaouachi & Frasson, 2012b). Engagement is defined as the mental vigilance and alertness while accomplishing a task (Berka et al., 2004). This index was first defined in (Pope et al., 1995a), and relies on neuroscientific research on attention and vigilance. It is computed from three EEG bands: Θ (4 - 8Hz), α (8 - 13Hz) and β (13 - 22Hz), and it obeys to this equation (Chaouachi & Frasson, 2012a): $Engagement = \frac{\theta}{(\alpha + \beta)}$. In this paper, we investigate also the distribution of engagement among the brain lobes (Teplan, 2002; Vuilleumier, 2005): the Frontal lobe has two key functions, i.e., controlling motor activities (including speech), and human "executive functions" (e.g., planning, reasoning, making decision); the Temporal lobe controls visual and auditory memories; the Parietal lobe is responsible for processing sensory information, comprehending oral and writing language, and controlling working memory; the Occipital lobe is responsible for vision. We consider the brain lobe reaction to an argument within 10 sec. to characterize the persuasive strategy effect. Emotions have an important role in decision making and can manifest with regard to three levels, namely, *experiential*, *behavioral* and *physiological*. For example, during conversations, when someone attacks an argument, she could experience the anger emotion, her behavioral reaction is show by the angry facial expression or aggravated voice tone, and the physiological response consists in an increasing heart rate. To improve the emotion recognition accuracy, multimodal techniques by combining different sensors to capture these different emotional reactions are used. We combined physiological sensors (EEG) with facial expression analysis system

(FaceReader 6.1)[29]. By analyzing the user's face streamed via webcam, the FaceReader software is able to recognize six basic emotions: happy, sad, angry, surprised, scared and disgusted. The FaceReader model reaches 87% accuracy by extracting and classifying in real-time 500 key points in facial muscles. As output, FaceReader provides the probability of the presence of these six emotions, as well as the probability of the neutral state.

4.3 Experimental setting

The goal of our experiment was to investigate how the argumentative persuasion process in debates is affected by the mental states and emotions of the participants, and vice-versa. In each debate, besides the participants equipped with the EEG Emotiv EPOC devices, there is a participant who plays the role of the *persuader*, called the PP in the remainder of the paper. The PP adopts and maintains a predefined viewpoint in the debate (i.e., pro or con), together with an argumentation strategy (i.e., Logos, Pathos or Ethos). PP intends to persuade other debaters of her viewpoint on the debated issue. The goal is to evaluate the following hypotheses:

- H1: Argumentation strategies trigger negative emotions and engagement having an impact on the persuasion.
- H2: Specific brain lobes are activated when a Logos or an Ethos argument is proposed by the PP, while other lobes are solicited when the PP puts forward a Pathos argument.
- H3: Pathos arguments activate a higher empathy, triggering a number of arguments put forward by the other participants to support PP's arguments. Pathos arguments have a more effective persuasive power in the debate.

4.3.1 Participants and roles

4 participants aged from 19 to 45 were involved in each of the 5 debate sessions, and each participant received a compensation of 20$ at the end of the session. In total, we collected data from 20 participants (7 women, 13 men). The size of the experiment is driven by the complexity of the experimental setting (devices, protocol). Debaters were preselected after filling an online form that collects their initial opinions about all the debate subjects, data is anonymized and

[29] www.noldus.com/human-behavior-research/products/facereader

kept confidential. This step was necessary to ensure possibly conflicting initial opinions in the debates. The ideal configuration includes 2 participants in favor and 2 against the debated topic. When not possible, a random assignment has been carried out. Each participant was kept separate from the others to avoid interactions out of the debate platform. In addition to the four participants and the PP, a moderator who proposes the debated issue and solicits unresponsive participants participated too. Each group of participants was involved in two debates. All participants (including the PP) were identified in the debate platform through a nickname. The PP cannot be identified by her nickname. No personal information about participants was disclosed during the debates.

4.3.2 Protocol

Phase 0: Participants fill in the self-reporting questionnaire about their initial opinions on the debate topics. They are associated to the debate sessions.

Phase 1: Familiarization of the participants with the Internet Relay Chat debate platform, the EEG headset, the camera for emotion recognition, and signature of a consent form.

Phase 2: The debate starts. Participants are involved in two debates for a maximum of 20 minutes each. The moderator provides the debaters with the topic to be discussed, and asks each participant to provide a general statement about her opinion on the topic. Each participant writes her viewpoint to the others, then the others are asked to comment on the expressed opinions. The PP plays the predefined persuasion strategy to convince the others with a different opinion, meaning that all arguments put forward by the persuader apply only the selected strategy. No turn taking was applied. Participants were free to propose their arguments, and the PP participates in the debate with the same amount of arguments as the other participants. The debaters were free to put forward generic arguments about the debated topic, or to explicitly refer to the other participants' argument to attack or support them. Arguments proposed by the PP were pre-instantiated arguments retrieved on online debate platforms[30], and categorized with the three persuasion strategies we identified. These arguments allowed us to provide a fixed stimulus in the debate. When necessary, the PP slightly adapted the pre-defined argument to precisely refer to another participant's argument, e.g., "I don't agree with you Participant1

[30] www.debate.org/ , www.createdebate.com/

because predefined argument". After about 15 minutes of debate, the moderator asked to provide their final viewpoint on the topic, and the debate is closed. Strategies have not been randomized. For each debate session, the PP applies the logos strategy for one debate, and either Pathos or Ethos for the second debate to compare for each set of debaters a more rational strategy (i.e., Logos) vs a more empathic one (either Ethos or Pathos). The contingency table below (Table 6) shows the correlation of the strategy adopted by the persuader and her stance in the 10 debates[31].
Phase 3: Participants are asked to fill a second self-reporting questionnaire on their experience in the debate.

Table 6. Persuader's opinions and strategies.

Strategy	***Pro***	***Con***	***Total by Strategy***
Pathos	0	3	3
Logos	4	1	5
Ethos	1	1	2
Total By Stance	5	5	10

4.3.3 Post-processing phase

We synchronized the textual argument collected during the debates, with the engagement index and the emotions. We are aware that field experiments, as the one proposed in this paper, suffer from the possibility of contamination, and we agree about the fact that experimental conditions can be controlled with more precision in a constrained experimental setting. However, field experiments have the advantage that outcomes are observed in a natural setting rather than in a contrived environment, thus showing higher external validity than "laboratory" experiments. For instance, the reader may argue about our choice of an experimental setting where 5 persons are involved at the same time, instead of a more controlled setting with a 1:1 face-to-face exchange. However, our interest is not in studying the effect of a single strategy on a single

[31] The Pathos strategy has not been used with a Pro stance because i) we had 6 debate sessions but the EEG data of the first session, where we considered Pathos/Pro, was corrupted, and ii) the stance depends also on the arguments used on the debate platforms we collected to construct PP's ones.

person with respect to a single dialogue move, but in considering a more realistic setting where several persons interact, like on social media.

4.3.4 Dataset

Two annotation tasks have been carried out offline on the collected data[32] by two annotators. Each argument is annotated with debate identifier, argument identifier, participant, and timestamp. In total, 791 arguments, and 162 argument pairs (74 linked by an attack and 88 by a support) were annotated. We computed the inter-annotator agreement for the relation annotation task on 1/3 of the pairs of the dataset (54 randomly extracted pairs), obtaining a satisfactory agreement: $\kappa = 0.83$.

4.4 Experimental results

This section reports on the obtained results for our hypotheses. We divided the debate into three phases: the introduction (INTRO) where the PP states her own opinion on the topic of the debate; the argumentation (ARG) includes the reformulation, the refutation and the contribution of new ideas according to the strategy adopted by the PP; the conclusion (CONC) where the PP recalls her position and final opinion. This structure is inspired from the conversation structure in pragmatics, where conversations have a linear structure, i.e., initiation, maintenance and termination (Kellermann et al., 1989). For data synchronization, we considered the participants' physiological reactions during 10 seconds after each intervention of the PP (Lee & Hsieh, 2014), and we computed the average emotion values of the 10 seconds after each argument proposal. We considered the anger scores in the result analysis because it was the most predominant emotion during the debates (Rozin & Royzman, 2001).

Table 7. Experiments finding at a glance.

There is a significant correlation between the persuasion strategy and the participants' emotions	NO	H1
Engagement in supporters and anger in opponents grow in an inversely proportional way	YES	H1

[32] The corpus is available at https://goo.gl/xSykTi.

Logos activates language comprehension and situations correlation	YES	H2
Logos activates planning and decision making	NO	H2
Ethos leads to the higher percentage of attacks with regard to PP's arguments	YES	H3
Pathos leads to the higher percentage of supports with regard to PP's arguments	YES	H3

H1 - Persuasion vs. emotions and engagement

In this first hypothesis, we verified for each strategy, the means of anger generated throughout the different phases of the debate. To verify the impact of anger and engagement on persuasion, we ran a repeated ANOVA measure. As within-subjects factors, we consider the debate phases (INTRO, ARG, CONC). As between-subjects factors, we consider *PP_ strategy* (Ethos, Logos, Pathos), measure (anger, engagement), and participant's final position (Neutral, Opponent, Supporter). We validate the repeated ANOVA measures with Mauchly's test (Mauchly, 1940) for sphericity on the dependent variable *Deb_phases* (sig=.013) (we assess the significance of the corresponding F with Greenhouse and Geisser 's correction (Greenhouse & Geisser, 1959)). For the within-subject effect test, we have a significant effect of debate phases and *PP_strategy* on measuring (engagement and anger) with p=0.016 and F(8.857, 113.372)=2.405. The between-subject effects results show that there are significant main effects of the *PP_strategy* * *Final_Position*, F(8, 64)=2.178, p=0.041, meaning a significant effect of the persuasion strategy, anger and engagement on persuasion. Figure 14 presents the corresponding engagement to compare the effect of emotions on the engagement. Note that if anger decreases, the engagement increases in all persuasion strategies.

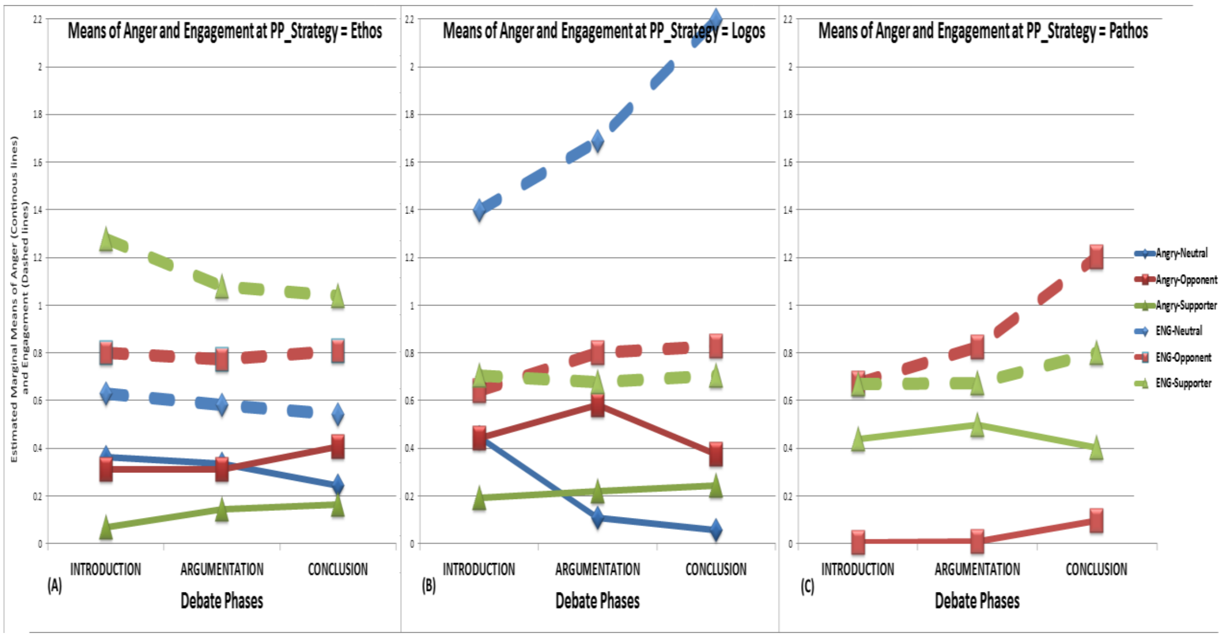


Figure 14. Means of anger (continuous lines) and engagement (dashed lines) (y axis) by debates' phases (x axis) for the different persuasion strategies. Blue, red and green colors correspond, respectively, to the participants' final position (Neutral, Opponent, and Supporter) to PP's opinion.

For the Logos strategy (Figure 14-B), participants who stayed *Neutral* all over the debates had low negative emotions and their engagement was high. So participants who have not decided about the PP's opinion were more engaged in looking for logical reasons to support opinions. This can be interpreted as follows: neutral participants follow the arguments deployed by Logos and show a high engagement in trying to be persuaded. The opponents show a clear increase of negative emotions and loss of engagement. They are more engaged in the ARG phase in refuting the PP's arguments (*emotional resistance*) whereas the supporters were less engaged because they already accepted PP's logic. Hence, for the Logos strategy, neutral participants show decreasing negative emotions and engagement growth, whereas opponents are mostly subject to negative emotions and disengaged to follow the logical reasoning.

For the Ethos strategy (Figure 14-A), opponents rejected the credibility of the PP and were not engaged in following her opinion. Their position does not change during the debates end where the negative emotion is higher. The neutrals were less engaged throughout the debate phases compared to the other participants. This can be due to the lack of interest in the subject of the

debate and even disengagement in taking a position face to an expert opinion. We may notice that the supporters' engagement is higher in the INTRO phase, and continues to decrease at the ARG and CONC phases while their negative emotion is the lowest through the debate phases compared to other participants, indicating their satisfaction towards the expert's opinion.
For the Pathos strategy (Figure 14-C), there are no neutral participants. We have opponents with increasing engagement related to the resistance to the emotional examples proposed by the PP. They were suppressing their negative emotion elicited by the Pathos strategy so their anger is low. Supporters were affected by Pathos, so their negative emotions are higher and their engagement is lower compared to the opponents because of the emotional effect of this strategy.

H2 - Brain solicitation vs. strategies

To verify the second hypothesis, we compute the differences in terms of engagement of each brain region for each participant, running a repeated ANOVA measure. The goal is to measure the effect of persuasion strategies on the engagement of each participant, considering both the different brain lobes that are activated, and the debate phases (the latter is the *within-subject factor*). As *between-subjects factors*, we consider the strategies and the brain lobes. Considering the resulting correlations among the strategy applied and the brain lobes activated in the participants in the different phases of the debate, we found $F(1.243, 30.683)=4.495$ and $p=0.027$. The factor Deb phases has a significant effect on the participant's engagement. We also have a significant interaction of the factors *Deb_phases* * *PP_strategy* with $F(2.486, 30.683)=4.059$ and $p=0.012$, meaning a significant effect on engagement[33]. The between-subject effects results show that there is a significant main effect of *PP_strategy* on the engagement, $F(2, 148)=3.885$, $p=0.023$.

[33] Complete SPSS's results: http://bit.ly/2nmbygV .

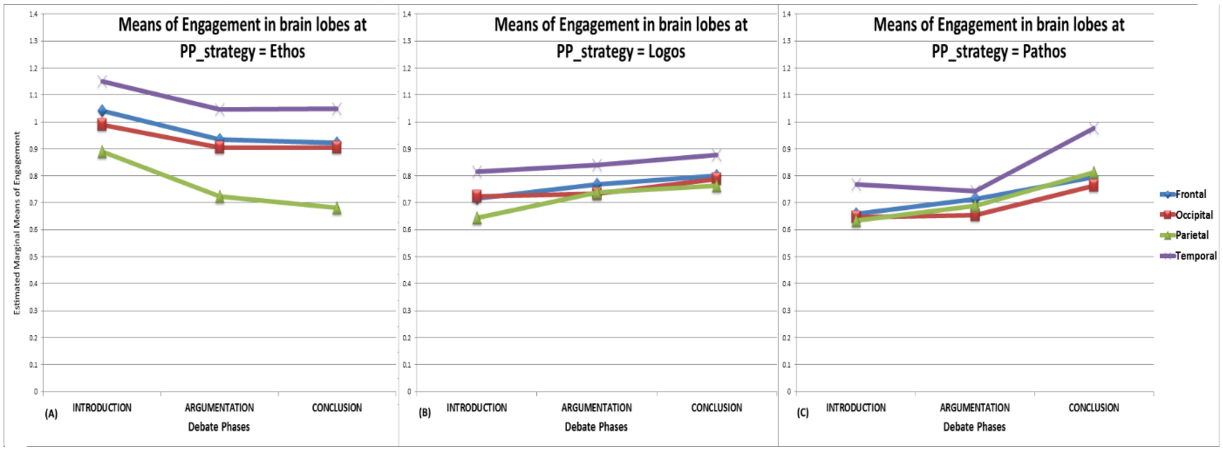


Figure 15. Estimated marginal means of engagements (y axis) in brain lobes by debates' phases (x axis) for the different persuasion strategies. Blue, red, green and violet lines correspond to the Frontal, Occipital, Parietal and Temporal brain lobes.

For the Logos strategy, the most activated brain region is the parietal. Figure 15-(B) shows that there is a significant difference between the INTRO and ARG phases for the parietal, which is the most activated lobe. By looking at the simple effect comparison, we found that the only significant mean difference is of parietal engagement between the ARG and INTRO phases with the Logos strategy (Mean difference= .115, p = .019). This result was unexpected, as we know that the frontal lobe is normally in charge of the planning, and rational decisions. By analyzing Logos arguments, we find that the PP used examples to justify her point of view, and imagination, residing in the parietal lobe, was triggered.

For the Ethos strategy, we have found that the parietal region was also activated. Looking at Figure 15-(A), we see that the engagement in the parietal is high in the INTRO phase and decreases in the ARG phase. By looking at the simple effect comparison, we found that the only significant mean difference is of parietal engagement between the ARG and INTRO phases with the Ethos strategy (Mean difference = -.174, p = .024). For the CONC phase, the engagement remains similar to the ARG phase both with the Logos and Ethos strategies. Engagement is related to the resistance towards the persuader's arguments: the more there is a resistance, the more there is engagement. For the Ethos strategy, as the PP is assimilated to an expert, the engagement is decreasing in the ARG phase. Parietal lobes play a role in interpreting sensory information and orientation, meaning that the participant tries to establish new rules to take decisions. Recent studies discuss the correlation between this region and the process of decision making (Huk & Meister, 2012), and other studies have shown the role of right temporal-parietal junction for thinking about thoughts, e.g., people's belief, desires and emotions (Saxe & Wexler, 2005).

For the Pathos strategy, the PP tried to induce empathy in participants. This resulted in the generation of strong emotions, and the circuit of emotions starts from the frontal to reach, through the cingulate Cortex, amygdala and hippocampus in the limbic system. The most important difference of engagement between the INTRO and ARG phases is indeed in the frontal lobe (see Figure 15-(C)). In the simple effect analysis, the mean difference of the frontal engagement between INTRO and ARG with the Pathos strategy is the most important compared to the other brain lobes, even if it is not statistically significant (Mean diff.=0.61, p = 0.332).

H3 - Pathos persuasiveness

We hypothesize (H3) that the Pathos strategy impacts more than the other strategies in terms of persuasive power, and consequently it gathers more support towards the PP's arguments than the others. Table 8 reports about the changes of opinion of participants by comparing their initial opinion, and the final opinion after the debate. Since self-reporting is not predictive (Stock, Guerini, & Pianesi, 2016), the table reports also about participants who have changed their opinions but did not disclose this change in the questionnaire.

Table 8. Participants' changes of opinion. Y: an opinion change occurred; N: no change; underlined: change from neutral; italic: a change not reported by the participant (detected by comparing his initial and after-debate opinions).

Debate	Strategy	PP position	P1	P2	P3	P4
DeathPenalty	Pathos	Con	<u>Y</u>	N	N	<u>Y</u>
Torture	Logos	Pro	N	Y	N	*Y*
Suicide	Ethos	Pro	N	N	N	*Y*
Profiling	Logos	Con	N	N	<u>Y</u>	*Y*
Nuclear	Logos	Pro	N	N	N	Y
Religion	Pathos	Con	N	N	*Y*	Y
Vaccins	Logos	Pro	N	N	N	N
GunRights	Ethos	Con	N	N	N	Y
Schools	Logos	Pro	N	N	*Y*	N
Organs	Pathos	Con	N	N	*Y*	*Y*

To verify this hypothesis, we first need to normalize the number of attacks and supports for each debate with regard to the different strategies. Figure 16 shows that the number of attacks and supports significantly changes depending on the strategy employed by the PP: Ethos is the strategy leading to the higher percentage of attacks in the argumentation, much more than the Logos and the Pathos strategies, while Pathos is the strategy leading to the higher percentage of supports with regard to the arguments proposed by the PP. Logos is in-between, as it is the most balanced strategy with regard to the percentage of attacks and supports. These results confirmed from the argumentation perspective what we already observed in H1 and H2: Pathos leads to the higher empathy leading to more supports than the other strategies. Note that these supports come even from those participants who do not agree with the PP, but they "*cannot*" attack the Pathos arguments she proposes, so they tend to agree on minor points related to the main topic. Ethos leads to more attacks than Logos: this can be explained by the fact that when an Ethos argument is proposed, the other participants do not evaluate the source as reliable, and tend to attack these arguments asking for evidences. Given that participants do not know each other, this behavior makes sense as authority is assessed by reputation and recommendation, and not only by claims.

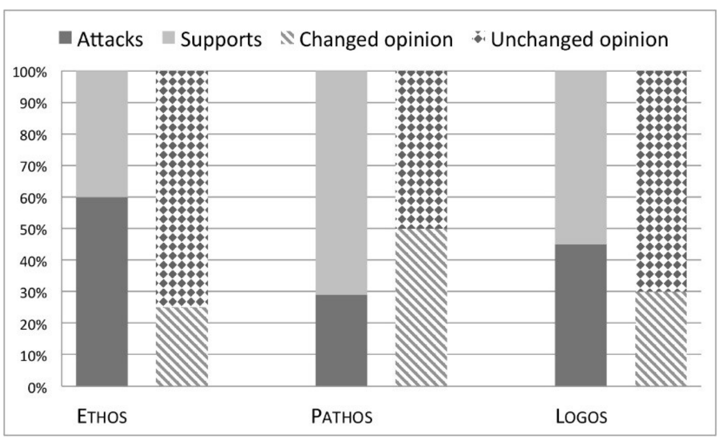


Figure 16. Percentage of attacks and supports for and against PP's arguments (1st columns), and percentage of participants with changed/unchanged opinion (2nd columns).

The validation of H3 is confirmed by analyzing the percentage of participants who changed/did not change their opinions with regard to the persuasion strategies (see Figure 16). On the one side, Pathos is the most effective strategy with regard to the percentage of participants who actually changed their mind after the debate, in line with the fact that participants tend to support

Pathos arguments in the debate. On the other side, Logos and particularly Ethos are the less effective strategies with few participants persuaded by the PP.

4.5 Related work

Very few approaches in persuasive argumentation involve humans in the loop. Among them, (Rosenfeld & Kraus, 2016) evaluate a methodology for human persuasion through argumentative dialogs, with human users. The huge difference with regard to (Rosenfeld & Kraus, 2016) is that we do not analyze the argumentation style, but we capture the emotions and mental states directly on human participants through sensors. In (S. Benlamine et al., 2015), we studied the connections between emotions and argumentation, but we do not consider persuasion. In (M. S. Benlamine et al., 2017), we studied the correlation of the engagement index in brain hemispheres with the persuasion strategies. The difference with H2 is twofold: i) here, we provide a more fine grained analysis of the correlation of the engagement with regard to the four lobes instead of the left and right sides, ii) we concentrate on the correlation with the persuasion strategies, while in (M. S. Benlamine et al., 2017) we correlated with the neutral vs. opinionated (pro/con) stance of the participants. To the best of our knowledge, in neuroscience (Cacioppo, Cacioppo, & Petty, 2018), no other work investigates the correlation between persuasive argumentation and mental states captured from users' brain through sensors. Usually, these factors are studied based on questionnaires with the participants.

4.6 Conclusions

The main contributions of this paper are: *i)* the first field experiment to study the correlation of persuasion strategies, argumentation and emotions using EEG headsets and cameras, *ii)* an annotated dataset of arguments characterized by a persuasion strategy, and *iii)* the first steps towards the definition of human-like empathic argumentative agents. The analysis of the results allowed us to highlight some drawbacks of our experimental setting to be addressed: *i)* more fine grained persuasion strategies may be considered, as these categories are highly general and sometimes difficult to be evaluated; *ii)* the strategies adopted by the other participants should be taken into account to expand the scenario (here, to overcome this issue, we consider them as random and we focus on the punctual reactions of the participants to PP's arguments); *iii)* the binary variable (pro/con) expressing the stance of the participants with regard to the debated issue may not fully capture the effect of a strategy, so allowing the expression of degrees of pro/con could be preferable.

Acknowledgments. The authors acknowledge support of the SEEMPAD project (http://project.inria.fr/seempad/), the FRQNT (Fonds de Recherche du Québec Nature et Technologie) and NSERC (National Science and Engineering Research Council).

Chapitre 5 : Évaluation des émotions dans un environnement de jeu vidéo

Nous continuons à analyser les émotions et l'engagement dans les environnements virtuels, mais pour ce chapitre et le suivant, l'étude est réalisée dans le cadre d'un environnement de jeu vidéo. Dans cette étude, nous analysons les réactions affectives des utilisateurs dans un jeu d'aventure en 3D à l'aide d'un casque EEG et un menu d'auto-évaluation de leurs émotions au cours du jeu. Nous avons comparé les résultats en fonction de leur sexe et de leur niveau d'expérience de jeux (*Gamer*/*non-Gamer*).

L'objectif de notre recherche est de répondre aux questions suivantes: (1) Quelle proportion d'émotions positives/négatives les joueurs ont-ils ressentie au cours de leur interaction avec le jeu? (2) Existe-t-il des différences significatives entre les émotions ressenties par les joueurs masculins et féminins et / ou les *Gamers* par rapport aux *non-Gamers*?

Au cours des expériences, nous avons recruté vingt participants pour jouer à Danger Island développée dans unity3d par Samira Bouslimi dans le cadre de sa maîtrise. La mission du participant était de se sauver d'une île dangereuse dans laquelle son hélicoptère a atterri à cause de manque de carburant. Le joueur doit trouver un ou deux bidons de carburant, selon ce qu'il a choisi soit *Gamer* ou *non-Gamer* dans le menu au début du jeu, et de revenir à l'hélicoptère. Armé d'une arme automatique, le joueur va devoir s'échapper de différents ennemis dans l'île. Pendant le jeu, le participant a été invité à faire part de son état émotionnel à partir d'un menu permanent. Ainsi, lorsqu'il décide que son état émotionnel a changé, le participant appuie sur Echap pour mettre le jeu en pause, faire un choix et revenir au jeu. Après l'expérience, le moniteur ouvre le ficher journal (*log-file*) contenant les événements du jeu et les émotions auto-déclarées pour le montrer au joueur qui confirme ses émotions auto-déclarées et donne éventuellement des émotions supplémentaires pour certains événements pour lesquels il a oublié de faire un rapport pendant le jeu.

Dans cette expérience, nous avons choisi de combiner deux mesures: l'indice d'engagement (à partir des données EEG) et les émotions rapportées dans le fichier log lors de l'auto-évaluation au cours du jeu. L'excitation était l'état émotionnel le plus souvent rapportés, suivi de la joie, tandis que l'ennui et la tristesse étaient les états émotionnels les plus rarement rapportés au cours du jeu. La différence entre les sexes concernait la peur et la surprise. Les femmes ont déclaré avoir plus la peur et les hommes plus la surprise.

Afin de comparer efficacement les états émotionnels des joueurs nous avons classé les émotions en quatre groupes (Jason M Harley, Bouchet, & Azevedo, 2013): Activation positive (engagement, flux, excitation et joie), Désactivation positive (calme), Valence négative (frustration, peur, ennui, tristesse et colère) et non-valence (confusion et surprise). Les *Gamers* masculins ont déclaré avoir vécu la plus grande proportion d'émotions du groupe Activation positive et le moins de Désactivation positive (calme). Les *non-Gamers* ont plus rapporté l'excitation pendant le jeu et des faibles niveaux d'émotions négatives. Les femmes *non-Gamers* ont déclaré avoir beaucoup plus peur que les hommes, selon le résultat du t-Test.

Les données EEG résultantes ont été échantillonnées à une fréquence de 128 Hz. Nous avons appliqué une transformation rapide de Fourier (FFT) sur les données EEG de chaque électrode pour extraire une puissance spectrale (en μV^2) toutes les secondes à l'aide du logiciel Acqknowledge[34] Biopack avec une fenêtre de Hamming pour éliminer les discontinuités[35] n'existant pas dans le signal d'origine qui sont dues à la périodisation de la FFT .

Comme résultats, nous avons trouvé que les joueurs (masculins) ont connu la plus grande proportion du groupe d'émotions Activation positive et un haut niveau d'engagement. Les femmes ont déclaré avoir plus de peur et de surprise. Les *non-Gamers* ont déclaré apprécier le jeu plus que les *Gamers*, ce qui était probablement dû au fait qu'il s'agissait d'un prototype plutôt que d'un jeu commercial raffiné. L'EEG peut être utilisé efficacement pour mesurer l'engagement des joueurs, puisqu'il correspond aux émotions déclarées par les participants.

Le reste de ce chapitre est constitué de l'article intitulé « *Toward Brain-based Gaming: Measuring Engagement During Gameplay*» publié à la conférence world conference on educational media and technology, Edmedia 2015. Nous rappelons que ma contribution essentielle consiste à la conception de l'expérimentation, à la collecte des données, au

---

[34] Acqknowledge Biopack: https://www.biopac.com/product/acqknowledge-software/

[35] http://www.iro.umontreal.ca/~mignotte/IFT3205/Chapitre_IFT3205_AnayseSpectrale_slides/slide009.html

développement du menu d’auto-évaluation au cours du jeu ainsi que le traitement et l’analyse de données physiologiques et à la rédaction du papier.

Toward Brain-based Gaming: Measuring Engagement During Gameplay

Benlamine, M. S., Bouslimi, S., Harley, J. M., Frasson, C., & Dufresne, A. (2015). Brain-based gaming: measuring engagement during gameplay. *Proceedings of the 9th world Conference on Educational Media and technology,* ***EDMEDIA-2015***, (pp. 717-722), Montreal, Quebec, Canada, June 22-25, 2015. Association for the Advancement of Computing in Education (AACE).

https://www.learntechlib.org/primary/d/151340/

Abstract: In this study we analyzed electroencephalography (EEG) and online self-report emotion data from 20 students during their interactions with a 3D adventure game developed using the Unity engine. Analyses of the data revealed that EEG can be effectively used to measure learners' engagement, that this metric corresponds to participants' levels of self-reported emotions, and that directional differences exist between gamer and non-gamer men, as well as women that warrant future study.

5.1 Introduction

Educational and commercial learning environments are blending together with the increasing awareness of the importance of supporting student engagement and other positive affective states. Moreover, a recent review by Harley and Azevedo (Jason M Harley & Azevedo, 2014) revealed that amongst agent-based (computer-based) learning environments those with successfully implemented gaming characteristics elicited the highest proportion of positive activating emotions. Given the relationship between learning and emotions it is critical that educational researchers advance an understanding of the discrete gameplay mechanisms and characteristics that foster student engagement, such as narrative (Sabourin & Lester, 2014), in order to implement empirically-validated features into state-of-the-art learning systems. A second challenge associated with designing and evaluating computer-based learning environments with regard to emotion is exploiting the diverse, yet complex technologies available to researchers. One of the most promising means of detecting and measuring emotions lies in the use of physiological measures which can measure learners' emotions without requiring them to report how they are feeling (Calvo & D'Mello, 2010; Jason Matthew Harley, 2016). In this paper we advance the state of the art with an electroencephalography (EEG) headset, capable of measuring engagement, while using a non-invasive self-report method to provide participants with an opportunity to report how they feel while playing without being forced to interrupt their gaming session.

This paper tackles a third challenge related to games and emotions: gender differences. It is a widely held belief that men tend to play more games, especially Triple-A title adventure games, than women. In this study we recruited both men and women and investigated whether there were observable differences in how they responded emotionally to the game. Our research questions were the following: What proportion of positive emotions did learners' experience over-all during their play-through of the game? Did significant differences exist between the emotions experienced by male and female players and/or gamers and non-gamers?

5.2 The Study

Twenty participants (7 women, 13 men) participated in the experiment which involved them playing Escape from Danger Island, a game developed by the authors. Participants' age ranged was from 24 to 35 and all were recruited via a departmental list serve from the Computer Science department of one large North American University. 20% of the participants identified themselves as regular gamers. In Escape from Danger Island (see Figure 17), gamers played a 3D, top-view avatar named Clara who became stranded on an island inhabited by menacing creatures after her helicopter ran out of fuel and forced her land. In order to survive she has to locate gas on the uncharted island and return with it to her helicopter. Over the course of the game the player is confronted by zombies, automatic machines guns, and wild animals, all of which are hostile. In addition to these numerous threats, players must also locate the fuel and return to the helicopter before their time runs out.

Prior to playing the game, participants were explained that their mission would be to: (1) find and return to the helicopter with two cans of fuel, if they identified as a gamer, or (2): to find and return with one can of fuel if they identified as a non-gamer. The emotion self-reporting tool was also explained to them in addition to other interface and control mechanisms necessary for them to navigate through and interact with the 3D world. Upon arrival, participants were greeted by the experimenter, filled out a consent form and put on the EEG headset with experimenter's assistance. The headset was readjusted until the electrodes had sufficient contact, indicated by a green light in the Emotiv Control Panel. Players had 10 minutes to complete their in-game objectives.

Figure 17. Clara in action.

Electroencephalography (EEG) data were recorded during the game session with an Emotiv EEG headset. The headset contains 14 electrodes spatially organized using the International 10–20 system. The electrodes are located at AF3, AF4, F3, F4, F7, F8, FC5, FC6, P7, P8, T7, T8, O1, O2 with two other reference sensors (placed behind the ears; see Figure 18). Data is generated in µVolts with a sampling rate of 128Hz. The signal's frequencies are between 0.2 and 60Hz. The Emotiv headset records the electrical neural activity in the brain. The resulting EEG data was sampled at a rate of 128 Hz. We applied a Fast Fourier Transformation (FFT) on the EEG data of each electrode to extract spectral power (in μV^2) for every second (Bin powers) using the Acqknowledge Biopack software with a hamming window to smoothen the signal. Average powers are computed for each frequencies band: delta (1–3 Hz), theta (4–7 Hz), alpha (8–13 Hz), beta (14–30 Hz), and gamma (31–42 Hz). These bands can be associated to specific cognitive processes (Von Stein & Sarnthein, 2000). The engagement index (Chaouachi et al., 2010b; Freeman et al., 1999; Pope et al., 1995a) was computed for every second using the EEG bands data as: Beta Power / (Alpha power + Theta power). In order to examine learners' levels of engagement their mean and maximum levels during gameplay were extracted across all 14 electrodes. The average of these values was then calculated so as to partially mitigate the influence of outlying scores. Mean max and minimum values were then used to identify the range of engagement participants experienced while playing the game. Mean and standard deviation values presented in the results correspond to the average standardized levels of engagement across the gaming session.

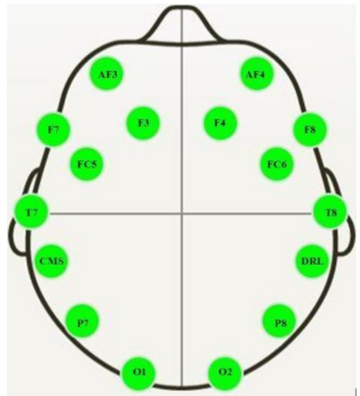


Figure 18. Sensor/data channels placement of the Emotiv Headset.

During the game the participant was invited to report his or her emotional state in a window in the upper right corner at any time during the experiment. Participants reported their emotional state by selecting the emotion that best corresponded to their current state. They were encouraged to report their emotional state whenever they felt it change, but no prompts to do so were provided in order to avoid interrupting their gaming session. Participants were able to pause the game when they wanted to report an emotional state. Whenever a self-reported emotional state was made it was logged and time-stamped in the system file (i.e., logfile). Learners were asked to select an emotional state from a list of twelve (see Table 9) and update the self-report window whenever they experienced a change in their emotional state. In order to account for different session lengths and sampling (participants emotional states were recorded in log files whenever an event occurred; e.g., bear attack) we transformed the frequency data into proportions. This was accomplished by summing the number of times each of the twelve states were reported in the log file and dividing by the total number of emotional states reported in the log file (whether or not a participant reported a change). In order to better compare learners' experience of target emotional states (e.g., engagement, flow, excitement, and joy), we grouped these states positively-valenced, activating states together and compared them to positively-valenced de-activating (calm), negatively valenced (frustration, fear, boredom, sadness, and anger), and non-valenced states (confusion and surprise;(Jason M Harley et al., 2015)). Non-valenced emotional states are expected to occur during learning and gaming experiences and do not necessarily constitute undesirable states (D'Mello et al., 2014).

5.3 Findings

Table 9 reveals that learners' reported experiencing the greatest proportion of positively-valenced activating emotions and the fewest positively-valenced de-activating emotions (relief). Excitement was the most commonly reported emotional state, followed by joy, while boredom and sadness were the most seldom reported emotional states during gameplay.

Table 9. Descriptive statistics for proportions of emotions of all participants.

Labels	Emotion	*Mean Proportion (SD)*		
		Male (*N* = 12)	Female (*N* = 7)	Over-all (*N* = 19)
Discrete	Engagement	9.72 (11.35)	13.55 (12.03)	11.13 (11.43)
	Flow	2.02 (4.41)	1.43 (2.44)	1.80 (3.74)
	Excitement	29.04 (19.34)	21.41 (18.66)	26.23 (18.95)
	Joy	16.52 (12.82)	10.81 (9.50)	14.42 (11.77)
	Surprise	15.37 (15.40)	5.63 (4.80)	11.78 (13.26)
	Frustration	7.99 (8.64)	10.00 (7.69)	8.73 (8.15)
	Fear	1.77 (3.62)	17.32 (25.02)	7.50 (16.61)
	Confusion	1.46 (3.43)	2.69 (2.67)	1.91 (3.15)
	Boredom	0.45 (1.56)	0.00 (0.00)	0.28 (1.24)
	Sadness	0.30 (1.03)	1.06 (2.80)	0.58 (1.85)
	Anger	9.48 (11.93)	5.87 (11.10)	8.15 (11.45)
	Calm	5.89 (8.89)	10.25 (8.33)	7.49 (8.73)
Grouped by dimension	Positive Valence Activating	57.30 (15.77)	47.20 (21.58)	53.58 (18.23)
	Positive Valence De-activating	5.89 (8.89)	10.25 (8.33)	7.49 (8.73)
	Negative Valence	19.99 (14.22)	34.23 (23.52)	25.24 (18.93)
	No Valence	16.83 (15.67)	8.32 (4.27)	13.69 (13.19)

Of the independent sample t-tests run for the twelve emotional states and four grouped emotional states, one significant result emerged: women reported experiencing significantly more fear than men, $t(17) = -2.16$, $p < .05$. This result should be interpreted with caution, however, as Levene's test for equality of variances was violated. In examining the descriptive statistics we found that men reported more positively-valenced activating emotions and non-valenced emotions, and fewer negatively valenced and positively-valenced, de-activating emotions. Amongst the discrete emotions, the greatest difference between genders was for fear and surprise. Women reported experiencing more of former, and men more of the later. In order to determine whether some of the directional differences could have been due to four of the male participants being gamers (and none of the female participants) we examined the descriptive statistics of male and female non-gamers as well as the male gamers in Table 10. We did not conduct inferential between-group comparisons because of the smaller sample sizes. Findings revealed that non-gamer men reported the greatest proportion of target emotions during the game (due mostly to higher levels of excitement) and the lowest levels of negative emotions.

Table 10. Descriptive statistics for proportions of emotions of male and female non-gamers and male gamers.

Labels	Emotion	*Mean Proportion (SD)*		
		Non-Gamers		Gamers
		Male ($N = 8$)	Female ($N = 7$)	Male ($N = 4$)
Discrete	Engagement	11.55 (13.03)	13.55 (12.03)	6.03 (6.99)
	Flow	1.79 (5.05)	1.43 (2.44)	2.50 (3.38)
	Excitement	35.06 (19.98)	21.41 (18.66)	17.00 (12.28)
	Joy	13.30 (12.14)	10.81 (9.50)	22.95 (13.29)
	Surprise	15.01 (17.62)	5.63 (4.80)	16.09 (11.97)
	Frustration	5.46 (6.37)	10.00 (7.69)	13.03 (11.33)
	Fear	2.66 (4.24)	17.32 (25.02)	0.00 (0.00)
	Confusion	1.19 (3.37)	2.69 (2.67)	2.00 (4.00)
	Boredom	0.68 (1.91)	0.00 (0.00)	0.00 (0.00)
	Sadness	0.44 (1.26)	1.06 (2.80)	0.00 (0.00)
	Anger	9.77 (11.12)	5.87 (11.10)	8.89 (15.24)
	Calm	3.08 (7.37)	10.25 (8.33)	11.50 (10.00)
Grouped by dimension	Positive Valence Activating	61.70 (18.00)	47.20 (21.58)	48.49 (1.12)
	Positive Valence De-activating	3.08 (7.37)	10.25 (8.33)	11.50 (10.01)
	Negative Valence	19.02 (15.51)	34.23 (23.52)	21.92 (13.15)
	No Valence	16.20 (17.53)	8.32 (4.27)	18.09 (13.19)

Table 11 presents learners' average standardized levels of engagement from the EEG headsets. Of the twenty participants, 15 of them had engagement levels classified as high (e.g., over 50%), whereas only three had low levels (the remaining two had an average level of engagement at exactly 50%).

Table 11. Descriptive statistics for levels of engagement across gaming sessions. Each column with numbers and demographics.

	Participant level of Engagement																			
Gender	M	F	F	M	M	F	M	M	M	M	M	F	M	F	M	M	M	F	M	F
Gamer	N	N	N	N	N	N	N	N	N	N	Y	N	Y	N	N	N	Y	N	Y	N
Level	H	H	L	H	M	L	H	L	H	M	H	H	H	H	H	H	H	H	H	H
Mean	59	70	45	62	50	28	51	49	57	50	53	51	51	64	60	64	61	55	68	63
SD	02	02	03	02	01	05	04	07	03	03	05	02	05	03	04	02	03	03	03	07

Table 12 presents the frequencies of gamer and male and female non-gamers by level of EEG engagement.

Table 12. EEG Engagement levels by Gamers and Non-Gamers.

Level	Non-Gamer		Gamer
	Male (N = 9)	Female (N = 7)	Male (N = 4)
High	6	5	4
Medium	2	0	0
Low	1	2	0

5.4 Conclusion

In summary, our results revealed that learners experienced the greatest proportion of self-reported positive activating emotions (relative to other emotional states) and high levels of engagement (compared to low levels). This finding suggests that the adventure features of the game (3D exploration, combat) were sufficiently implemented for the short gaming session to be enjoyed and immersive. Although valid significant differences did not exist between genders, clear differences emerged which may have been significant had the sample size been larger and the variances similar, in particular for fear and surprise which women reported experiencing more of. This could have implications for design features for participants' gender, where potentially fear-eliciting elements, like enemies jumping out at you from hiding places, could be toned down or clues given.

With regard to gamers vs. non-gamers, non-gamer men reported enjoying the game more than gamers, which was likely due to the game being a prototype rather than a polished commercial title. None-the-less, both groups (as well as women) were highly engaged by the game. This study and preliminary analyses of the data has demonstrated that EEG can be effectively used to measure learners' engagement, that this metric corresponds to participants' self-reported emotions, and that directional differences exist between gamer and non-gamer men, as well as women that warrant future study.

Acknowledgements. The research presented in this paper has been supported by funding awarded to the fourth author from the Natural Sciences and Engineering Research Council of Canada (NSERC). This research has also been supported by a postdoctoral fellowship awarded to the third author from the Fonds de recherche société et culture du Québec (FQRSC).

Chapitre 6 : Reconnaissance des émotions à partir des signaux physiologiques

Dans le chapitre précédent, nous avons présenté notre première étude sur les états émotionnels et mentaux dans un environnement de jeu vidéo. Le travail présenté dans ce chapitre fait partie de la même expérience, puisque avant que les joueurs ne commencent le jeu, nous avons demandé qu'ils regardent quelques images et indiquent les émotions ressenties à propos de ces images. Dans ce chapitre, l'étude présentée porte sur la reconnaissance des expressions faciales seulement à partir des données physiologiques (EEG). Les systèmes de reconnaissance des expressions faciales utilisent généralement la caméra. Mais les caméras ne conviennent pas dans de nombreuses situations, par exemple dans les environnements obscurs, ou quand le visage de l'utilisateur n'est pas dans le champ de la caméra, ou lors de l'usage de lunettes de réalité augmentée (Hololens, Google glaces, Meta,…) ou de casque de réalité virtuelle (Occulus, carte Google board, …).

Dans cette étude, nous répondons aux questions suivantes: (1) Peut-on prédire les expressions faciales à partir de données physiologiques telles que l'encéphalographie EEG? (2) Comment pouvons-nous construire un ensemble de données efficace pour l'entraînement des algorithmes d'apprentissage machine? (3) Quelle est l'exactitude du modèle de prédiction des expressions faciales à partir des signaux EEG? Les expressions faciales reflètent notre état émotionnel actuel: colère, joie, tristesse, peur, surprise, dégoût ou mépris. L'objectif de notre étude est de prédire les expressions faciales uniquement à partir de signaux physiologiques du cerveau (EEG).

Au cours de l'expérience, nous avons demandé aux participants de regarder les stimuli d'images de l'IAPS - International Affective Picture System (Lang, Bradley, & Cuthbert, 2008) et de rapporter l'émotion ressentie dans un questionnaire à l'écran. Vingt participants de l'Université de Montréal ont participé à cette expérience. Nous avons utilisé trente images d'IAPS regroupées dans huit émotions. Nous avons recueilli des données de l'encéphalographie EEG (Emotiv Epoch) et du système de détection des expressions faciales par caméra (FACET).

Dans cette étude, nous avons extrait dix-sept caractéristiques (12 temporelles et 5 spectrales) à partir de chacun des 14 senseurs du casque Emotiv EEG pour construire notre vecteur de caractéristiques de 239 dimensions. Nous avons ensuite entraîné des algorithmes d'apprentissage automatique en utilisant les données d'expression faciale comme *ground-truth* pour construire des modèles de régression pour prédire l'intensité des expressions faciales à partir des signaux EEG. Trois algorithmes d'apprentissage automatique ont été utilisés pour prédire les valeurs numériques de chaque catégorie d'émotions (kNN, Random Forest, l'arbre de décision rapide) avec une validation croisée par 10 flots dans la phase de test. Random Forest a dépassé les autres algorithmes avec un coefficient de corrélation plus élevé (arrivant à 92%) et des taux d'erreur plus faibles pour toutes les catégories d'émotions. Nous avons aussi utilisé des méthodes de sélection des caractéristiques pour avoir un vecteur de caractéristiques de 24 dimensions pour chacune des catégories d'émotion. À partir de ces vecteurs nous avons identifié les senseurs du casque EEG les plus importants dans la reconnaissance des expressions faciales. Notre approche fournit un moyen simple et fiable pour la détection des réactions émotionnelles de l'utilisateur. Cette même approche été utilisée pour développer une application temps-reel pour la reconnaissance des expressions faciales à partir des EEG, intitulée « NeuroExpress ». Les modèles utilisés dans cette application sont basés sur les données de toute l'expérience sur les images IAPS et le jeu « Danger Island ». Durant l'expérimentation, nous avons enregistré les expressions faciales des participants et leurs signaux cérébraux EEG. L'application NeuroExpress permet de mesurer les réactions émotionnelles de l'utilisateur dans son interaction avec n'importe quel media mais elle est plus utile dans les environnements de réalité virtuelle (VR) ou augmenté (AR) où le visage de l'usager est caché par le casque VR/AR.

Le reste de ce chapitre est constitué de l'article intitulé « *Physiology-based recognition of Facial micro-expressions using EEG and identification of the relevant sensors by emotion*» publié à la conférence conference on physiological computing systems, PhyCs 2016. Nous rappelons que ma contribution essentielle consiste à la conception de l'expérimentation, à la collecte des données, à s'assurer d'avoir un dataset fiable à partir des données recueillies et à appliquer les algorithmes d'apprentissage machine pour construire les modèles de prédiction des expressions faciales à partir des signaux EEG et à la rédaction du papier.

Physiology-based recognition of Facial micro-expressions using EEG and identification of the relevant sensors by emotion

Benlamine M.S., Chaouachi M., Frasson C., & Dufresne A. (2016). Physiology-based recognition of Facial expressions using EEG and identification of the relevant sensors by emotion. The *3rd International Conference on physiological computing systems,* ***PhyCs-2016***, Lisbon, Portugal, July 27-28, 2016. INSTICC / Springer International Publishing.

http://www.scitepress.org/Papers/2016/60027/60027.pdf

Abstract

In this paper, we present a novel work about predicting the facial expressions from physiological signals of the brain. The main contributions of this paper are twofold. a) Investigation of the predictability of facial micro-expressions from EEG. b) Identification of the relevant features to the prediction. To reach our objectives, an experiment was conducted and we have proceeded in three steps: i) We recorded facial expressions and the corresponding EEG signals of participant while he/she is looking at pictures stimuli from the IAPS (International Affective Picture System). ii) We fed machine learning algorithms with time-domain and frequency-domain features of one second EEG signals with also the corresponding facial expression data as ground truth in the training phase. iii) Using the trained classifiers, we predict facial emotional reactions without the need to a camera. Our method leads us to very promising results since we have reached high accuracy. It also provides an additional important result by locating which electrodes can be used to characterize specific emotion. This system will be particularly useful to evaluate emotional reactions in virtual reality environments where the user is wearing VR headset that hides the face and makes the traditional webcam facial expression detectors obsolete.

6.1 Introduction

6.1.1 Context and motivation

Affective Computing (Picard, 1997) is the general frame-work that considers emotions in human computer interaction. In particular, the overall objective is to make computers able to perceive, feel and express emotions. However an important goal remains to detect human emotions. Several studies have been successfully conducted to detect emotions using models that track facial expressions with camera or webcam with, for instance, CERT (Littlewort et al., 2011), or FaceReader (Lewinski, den Uyl, & Butler, 2014) …etc. The obtained results in these studies showed a high accuracy that has never been reached with other approaches using physiological data.

The joint efforts of researchers in machine learning, affective computing, physiological computing (Fairclough, 2010) and neuroscience are producing innovative methods for emotional and affective states recognition by analyzing data collected with subjective methods (self-report, expert annotation) or objective methods (log files, Kinect, camera, eye-tracking and electrophysiological sensors: EDA, HR, EMG, EEG, Respiration rate,…) .

Thanks also to breakthrough advancements in computer vision, facial expressions detection technology has reached commercial-level maturity and has become more common, e.g. with Kinect 2 in Xbox, software like Facereader, iMotions FACET, and NVISO. However, so far, all the focus has been on external assessment methods and to the best of our knowledge, no attempt has ever been made at detecting facial micro-expressions from EEG signal. A micro-expression (Ekman, 2007) is a brief spontaneous facial expression, unconscious (involuntary) and hard to hide as it lasts between 1/24 and 1/15 of a second. Because of their short duration, micro-expressions are identifiable only by trained peoples or in videos where the person's face is recorded. Software like FACET and FaceReader analyses videos frame by frame to extract the micro-expressions.

Micro-expressions are important because they give the spontaneous emotions of the users, which can be detected using facial expression software. However, it is not always possible to record the person's face; for example with low luminosity or when the person is moving his face or when he is using VR headset to interact with an immersive virtual reality environment. In such situations, physiological measures like EEG represent a promising alternative that could potentially solve these problems. Moreover EEG devices are being increasingly used as they present a practical low cost solution and help building interesting accurate models to track and assess users' states. EEG features can improve recognition of affect and facial expressions.

Recent studies in the field of neuroscience have found a relation between neural activations and emotions using the technique of Functional Magnetic Resonance Imaging, or fMRI. For example, (Kassam et al., 2013) show, through an experiment with 10 participants using fMRI, that there are consistent patterns of neural activation for all emotion categories within and between participants. In the meanwhile, several works has shown that emotional states can be recognized from EEG signal with reasonable accuracy (AlZoubi, Calvo, & Stevens, 2009; Chaouachi et al., 2010b; Chaouachi & Frasson, 2012a; Chaouachi et al., 2015a; Heraz & Frasson, 2007; Jraidi et al., 2013; Y. Liu, Sourina, & Nguyen, 2011; Lu et al., 2015). So, it seems sensible then to consider cerebral activity as input for detecting facial expressions rather than user's face images or videos. All this leads us to believe that a correlation between EEG features and facial micro-expressions should be investigated.

6.1.2 Objectives

Knowing now the importance of micro-expressions for emotion detection, we aim to build a predictive model of these emotional micro-expressions from EEG signals. With such a model, it will be possible to predict spontaneous facial expressions having as input cerebral activity signal using Emotiv Headset. More precisely, we aim to answer two questions: (1) How well can we predict facial expressions from the brain signals of participants? (2) Which EEG features are important for the prediction?

To reach these objectives, we will proceed according to the following steps: 1) measuring emotional reactions of a user confronted to specific emotional pictures using FACET system, 2) measuring the corresponding EEG signals and train machine learning models that correlates the micro-expressions with the EEG signals, and 3) predict the emotion only from the model and check the accuracy of the model.

The organization of this paper is as follows: section 2 presents the design of our experimental approach and methodology for the EEG-based facial expressions recognition. We show how we can evaluate emotions using a well-known set of emotional pictures. We detail the material used for the experimentation and the different measures obtained. Results and discussions are presented in section 3. We finally show how our model can predict the micro-expressions from EEG signals. We conclude this study with a presentation of the scope of use of this model as well as our future work.

6.2 Method

6.2.1 Participants

Twenty participants (7 women, 13 men) were recruited for our study. Their ages ranged from 21 to 35 years and all of them came from a North American University. All participants were right handed and comparable in age, sex and education.

6.2.2 IAPS pictures selection

We selected 30 pictures from IAPS (International Affective Picture System) (Lang, Bradley, & Cuthbert, 2008) with regard to their affective ratings (valence, arousal) after consulting the IAPS documentation. The selected pictures are well distributed throughout the space (Figure 19). We grouped those pictures in 8 emotion categories: Joy, Calm, Excitement, Engagement, Frustration, Anger, Sadness, and Surprise.

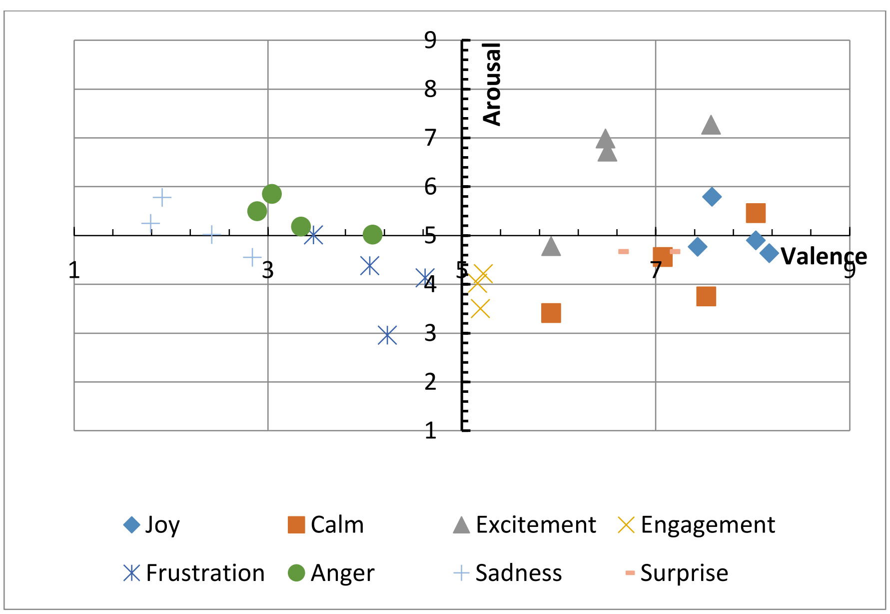


Figure 19. IAPS pictures' affective ratings distribution and their categories of emotion.

6.2.3 Experimental procedure

The participants were received and introduced to our laboratory. A consent form was given to the participant to read and sign it in order to register his agreement about doing the experiment. They were seated in front of a computer and a webcam. The participant's chair was adjusted to maintain good view of their faces to the webcam.

The experiment design was synchronized by iMotions Attention Tool which is a software platform that permit multi-sensors study (eye tracking, facial expressions detection, EEG, GSR …) with automatic synchronization. We have recorded data of facial expressions using FACET module and EEG using Emotiv Epoch headset. IAPS Pictures with the same emotion category were displayed 6 seconds one by one, separated by a noise screen of 5 seconds to accomplish a relaxation phase before the projection of the next picture. After that, a form appears asking the user to choose one from the eight emotion categories that best represent his global feeling about the previewed pictures. The same procedure was repeated for each of the eight emotion categories. The goal of this form is to have the user's subjective confirmation of the emotion he/she supposed to feel seeing the pictures. The chart below (Figure 20) shows the percent of the self-reported emotion categories by the IAPS pictures groups. We configured FACET to analyze frame by frame the videos of the user's face to extract his/her micro-expressions.

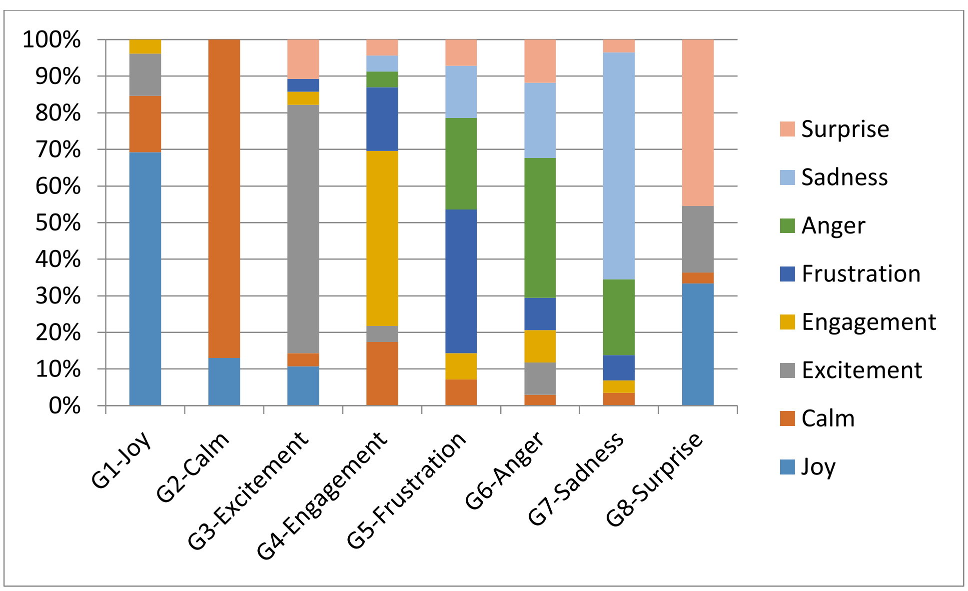


Figure 20. Self-reported emotions by pictures groups.

6.2.4 Experimental procedure

In this study, we have used the following tools.

6.2.4.1 iMotions FACET module

The FACET module detects and tracks facial expressions of primary emotions using real-time frame-by-frame analysis of the emotional responses of users via a webcam. A commercial webcam is used for the user face recording (Webcam Logitech Pro C920). The captured image is 1920 x 1080 pixels with 24-bit RGB colours, acquired at 6 frames/sec. The FACET module is the commercial version of CERT (Computer Expression Recognition Toolbox) (Littlewort et al., 2011) which is a robust facial micro-expressions recognition software with an accuracy that reaches 94% (Emotient, 2015). The resulted data file contains two columns (Evidence and Intensity) for-each of the following categories: Joy, Anger, Surprise, Fear, Contempt, Disgust, Sadness, Neutral, Positive Valence, and Negative Valence.

6.2.4.2 Emotiv EPOCH EEG

Physiological data were recorded during the session using the Emotiv EEG headset. The headset contains 14 electrodes that must be moist with a saline solution (Contact lens cleaning solution). The electrodes are spatially organized with respect to the International 10–20 system. They are located at AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4 (see Figure 21) with two other reference nodes (that would be placed behind the ears). The generated data are in µVolt with a sampling rate of 128 Samples/sec. The signal's frequencies are between 0.2 and 60Hz.

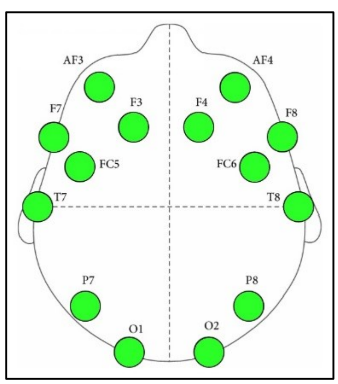


Figure 21. Data channels placement of the Emotiv Headset.

6.3 Data Analysis

6.3.1 Facial Expressions Data

Taking the webcam stream as input, the FACET system classifies each frame and provides two values for each emotion category, namely: Evidence and Intensity. The Intensity number is a value between 0 and 1, which denotes the estimation by expert human coders of the intensity of an expression. While the Evidence number is a value between -3 and 3 that represents the presence of an expression. For an easier interpretation, the Evidence number has been transformed into emotion probability between 0 and 1, using this formula (as in FACET guide) (Facet, 2013):

$$P_{Emotion} = \frac{1}{1+10^{-Evidence}} \qquad (1)$$

We computed the probability of each emotion in our dataset. These probabilities will be considered as ground truth in the models training phase.

6.3.2 Dataset creation

Building the dataset is an important process that has a big impact in the robustness and the accuracy of the resulting machine learning models.

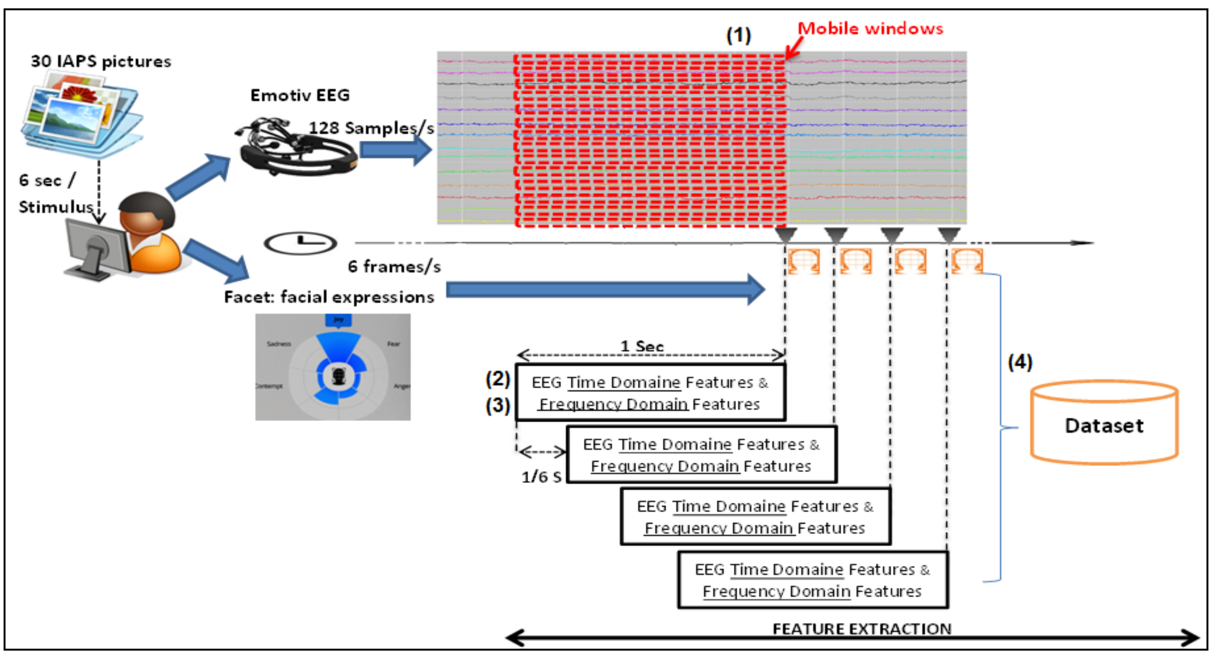


Figure 22. Pipeline of dataset construction for EEG-based Facial Expressions recognition.

The figure above (Figure 22) illustrates the entire pipeline that we have designed for the construction of our dataset for EEG-based Facial Expressions recognition. We developed a java program that uses 14 first-in first-out queues of size 128 of reals as mobile windows of 1 second EEG data from each electrode (each window contains 128 samples). (1) For each FACET frame time (every 1/6 sec) the program reads the content of each window and (2) performs statistical analysis to extract time-domain features and (3) spectral analysis to extract frequency-domain features. (4) The program records in a separate CSV file the FACET frame Time, the computed Time-domain and Frequency-domain features of each electrode (Table 13) and the probability value of each emotion category.

The frequency-domain features are computed by applying FFT on the 128 samples contained in each window for each FACET Frame Time. By using the Fast Fourier transform (FFT), we calculated the magnitudes (in μV^2) in frequency domain for the corresponding frequency bands (delta [1–4 Hz[, theta [4–8 Hz[, alpha [8–12 Hz[, beta [12–25 Hz[, and gamma [25–40 Hz[ ) with the application of hamming window to smooth the signal. The FACET frame rate is 6 frames per second, so each window receives, every 1/6 sec, 22 new EEG values (128/6=21.33 ~ 22).

Table 13. Computed features from EEG Signals.

Frequency-domain EEG Features (5 Features)	delta [1–4 Hz[, theta [4–8 Hz[, alpha [8–12 Hz[, beta [12–25 Hz[, and gamma [25–40 Hz[
Time-domain EEG Features (12 Features)	Mean, Standard Error, Median, Mode, Standard Deviation, Sample Variance, Kurtosis, Skewness, Range, Minimum, Maximum and Sum

The time-domain features (Table 13) are also computed based on each window that contains 128 samples for each FACET Frame time. Therefore, the used epoch length in this study is 128 samples. The twelve features of the EEG signal indicated above are extracted in the time domain. For each FACET Frame time, we computed from each window: Minimum, Maximum, Mean and Sum values. The Range represents the difference between the Minimum and Maximum. The Mode is the most commonly occurring value. The Variance measures the spread between the values in the window and the Mean. A variance of zero indicates that all the values are identical. The standard deviation is the square root of the variance. The standard Error is the standard deviation divided by the square root of the window size.

Kurtosis is a descriptor of the shape of a distribution which represents the distribution's width of peak. A Gaussian distribution has a kurtosis of zero; a flatter distribution has a negative kurtosis, and a more peaked distribution has a positive kurtosis. Skewness is a measure of the asymmetry of a distribution relative to its mean; a distribution can be negatively skewed when the left tail is longer or positively skewed when the right tail is longer, and a symmetrical distribution has a skewness of zero.

We have not used a window containing average values from all electrodes because we assumed that each emotion has its own activated area in the brain (Kassam et al., 2013). The Machine learning algorithms will select features from the suitable electrodes for the prediction of specific emotion category values.

The total number of computed features from all the electrodes is 238 (17 features per electrode: 5 frequency-domain features + 12 time-domain features). The collected dataset contains 21553 data points (1078 data point per participant; 36 data point per stimulus). We have created 10 CSV files where we put together all the extracted EEG Features and one emotion category extracted from FACET data. Each file was entered as an input to the WEKA machine learning toolkit (Hall et al., 2009) for generating EEG-based regression model to predict the values of one emotion category.

6.4 Data Analysis

For every one of the 10 emotion categories, a regression model was generated. Three machine learning algorithms were used to predict the numeric values of each emotion category; namely IBk (K-nearest neighbors' classifier), Random Forest (classifier constructing a forest of random trees) and RepTree (Fast decision tree learner).We used 10 fold validation in our test phase. For the prediction of the emotion classes' values, we have chosen dependent criteria: Correlation Coefficient (CC), Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) to evaluate the goodness of different machine learning algorithms.

Table 14. Comparison between machine learning methods.

Emotion	IBk			Random Forest			RepTree		
	CC	MAE	RMSE	CC	MAE	RMSE	CC	MAE	RMSE
Joy	0.85	0.02	0.05	0.91	0.02	0.05	0.72	0.03	0.07
Anger	0.88	0.05	0.09	**0.92**	0.05	0.08	0.83	0.07	0.10
Surprise	0.85	0.02	0.04	0.90	0.02	0.03	0.76	0.02	0.05
Fear	0.89	0.04	0.08	**0.92**	0.05	0.07	0.78	0.07	0.10
Neutral	0.87	0.05	0.10	0.91	0.06	0.09	0.74	0.08	0.14
Contempt	0.79	0.03	0.07	0.85	0.04	0.06	0.63	0.05	0.08
Disgust	0.88	0.03	0.06	**0.92**	0.03	0.05	0.63	0.05	0.08
Sadness	0.89	0.03	0.06	**0.92**	0.03	0.06	0.83	0.04	0.08
Positive	0.85	0.05	0.11	0.91	0.06	0.09	0.76	0.08	0.13
Negative	0.86	0.10	0.16	0.90	0.11	0.14	0.80	0.13	0.18

Compared to IBk (k=1 neighbor) and RepTree methods, Random Forest obtains higher correlation coefficient and lower error rates such as MAE and RMSE for all emotion categories, as illustrated in Table 14.

In order to find the optimal EEG features for each emotion category, we used a feature selection method called ReliefF over Random forest algorithm to choose an optimal feature set of size 24 (10% of the initial features set). The experimental results performed on the different emotion categories are presented in Table 15.

Table 15. Random forest models results with reliefF feature selection method.

Emotions	R.F. (selection of 24 Attributes)			Emotions	R.F. (selection of 24 Attributes)		
	CC	MAE	RMSE		CC	MAE	RMSE
Joy	0.8336	0.0266	0.0585	**Contempt**	0.8266	0.0374	0.0607
Anger	0.8888	0.0612	0.0872	**Disgust**	0.9196	0.0307	0.0483
Surprise	0.8272	0.0219	0.0428	**Sadness**	0.8854	0.0401	0.0640
Fear	0.8949	0.0516	0.0744	**Positive**	0.8647	0.0655	0.1012

Neutral	0.8854	0.0648	0.0971	**Negative**	0.8878	0.1093	0.1439

In the experiments, a 10-fold cross validation method was used for evaluation. It is notable that Random Forest method performs well even with 24 features. The 24 selected EEG features by emotion category are presented in Table 16.

Table 16. The selected 24 attributes by ReliefF over Random Forest for each emotion category.

Emotion	**The selected 24 attributes** (10% of features set)	Emotion	**The selected 24 attributes** (10% of features set)
Joy	P8_Mode, T8_Mode, P8_Range, T8_Minimum, T8_Range, P8_Maximum, T8_Maximum, T8_Standard_Deviation, T8_Standard_Error, P8_Minimum, T8_Median, P8_Median, T8_Sum, T8_Mean, FC6_Standard_Error, FC6_Standard_Deviation, FC6_Minimum, P8_Standard_Error, P8_Standard_Deviation, T8_Delta, P8_Mean, P8_Sum, T8_Beta, P7_Range	**Contempt**	T8_Mode, P8_Mode, T8_Minimum, FC6_Mode, F7_Minimum, FC6_Standard_Error, FC6_Standard_Deviation, FC6_Minimum, FC6_Range, FC5_Minimum, T7_Delta, FC5_Mode, T8_Maximum, F8_Mode, T8_Range, F7_Range, F7_Mode, FC6_Delta, F8_Range, FC5_Standard_Error, FC5_Standard_Deviation, F7_Maximum, T7_Beta, F8_Minimum
Anger	P7_Range, P8_Range, P8_Maximum, P7_Standard_Deviation, P7_Standard_Error, P8_Mode, P8_Minimum, P7_Maximum, AF4_Range, P8_Median, AF3_Mode, F3_Range, T7_Skewness, P7_Gamma, P8_Standard_Error, P8_Standard_Deviation, P7_Median, P7_Beta, P7_Alpha, P8_Sum, P8_Mean, AF3_Maximum, P7_Mode, P7_Delta	**Disgust**	P8_Maximum, P8_Mode, P8_Range, P7_Range, P8_Minimum, P8_Median, P8_Mean, P8_Sum, P8_Standard_Deviation, P8_Standard_Error, O1_Minimum, AF4_Range, FC5_Minimum, AF4_Minimum, F7_Minimum, P7_Standard_Deviation, P7_Standard_Error, P7_Maximum, F8_Maximum, FC5_Standard_Error, FC5_Standard_Deviation, T7_Skewness, AF4_Standard_Deviation, AF4_Standard_Error
Surprise	P7_Mode, P7_Range, P7_Maximum, P8_Skewness, P7_Minimum, P7_Standard_Error, P7_Standard_Deviation, P7_Kurtosis, P7_Median, P7_Sum, P7_Mean, P8_Kurtosis, P7_Skewness, T7_Sample_Variance, T7_Mode, P8_Minimum, T7_Standard_Deviation, T7_Standard_Error, T7_Kurtosis, P7_Sample_Variance, O1_Skewness, O2_Kurtosis, F8_Kurtosis, O1_Kurtosis	**Sadness**	P8_Range, P8_Maximum, P8_Mode, P8_Minimum, P7_Range, P8_Standard_Deviation, P8_Standard_Error, P8_Median, P8_Beta, P8_Theta, P8_Sum, P8_Mean, P8_Alpha, P7_Kurtosis, P8_Gamma, P7_Maximum, AF3_Range, P7_Standard_Deviation, P7_Standard_Error, P8_Delta, AF3_Minimum, AF3_Maximum, P8_Sample_Variance, P8_Skewness
Fear	FC6_Standard_Error, FC6_Standard_Deviation, FC6_Minimum, P8_Range, P8_Minimum, FC5_Minimum, FC6_Range, P8_Maximum, P7_Range, T8_Minimum, P8_Median, FC5_Standard_Error, FC5_Standard_Deviation, P8_Mean, P8_Sum, P8_Standard_Deviation,	**Positive**	P8_Mode, T8_Mode, T8_Minimum, T8_Range, P8_Range, P8_Maximum, T8_Maximum, P8_Minimum, T8_Standard_Deviation, T8_Standard_Error, FC6_Standard_Deviation, FC6_Standard_Error, P7_Range, T8_Median, FC6_Range, T8_Sum, T8_Mean, FC6_Minimum, T8_Beta, FC6_Mode,

	P8_Standard_Error, FC6_Delta, T7_Delta, FC6_Sample_Variance, FC6_Mode, P8_Mode, T7_Skewness, P7_Skewness		P8_Median, P8_Standard_Error, P8_Standard_Deviation, T8_Delta
Neutral	P8_Mode, P8_Range, P8_Maximum, T8_Mode, P8_Minimum, T8_Minimum, T8_Range, T8_Maximum, P8_Median, P8_Mean, P8_Sum, P8_Standard_Error, P8_Standard_Deviation, P7_Range, FC5_Minimum, T8_Standard_Deviation, T8_Standard_Error, FC6_Mode, FC6_Standard_Deviation, FC6_Standard_Error, FC6_Range, FC6_Minimum, P8_Theta, FC5_Standard_Error	**Negative**	P8_Range, P8_Maximum, P8_Minimum, P7_Range, F7_Minimum, P8_Median, P8_Sum, P8_Mean, P8_Standard_Error, P8_Standard_Deviation, F3_Range, T8_Minimum, T8_Range, F7_Standard_Error, F7_Standard_Deviation, F7_Range, P7_Gamma, AF4_Range, FC5_Mode, P7_Standard_Error, P7_Standard_Deviation, FC5_Minimum, P7_Maximum, FC6_Standard_Deviation

From the results in Table 16, we have identified the most suitable EEG electrodes by emotion category as illustrated in Table 17.

Table 17. Random Forest selected EEG electrodes by emotion category.

Emotion	Selected Electrodes
Joy	P8, T8, FC6
Anger	P7, P8, AF4, AF3, F3, T7
Surprise	P7, P8, T7, O1, O2, F8
Fear	FC6, P8, FC5, FC6, P7, T8, T7
Neutral	P8, T8, P7, FC5, FC6
Contempt	T8, P8, FC6, FC5, T7, F8, F7
Disgust	P8, P7, O1, AF4, FC5, F7, F8, T7
Sadness	P8, P7, AF3
Positive	P8, T8, FC6, P7
Negative	P8, P7, F7, F3, T8, AF4, FC5, FC6

These results are very important, since this is the first time we identify the most reliable sensors for each emotion category. In previous study, Liu and colleagues (Y. Liu et al., 2011) implemented a real-time EEG-based emotion recognition algorithm based on fractal dimension calculation of six different emotions using only AF3, F4 and FC6 electrodes. Our proposed model has better accuracy, more adaptability to all users and several advantages besides. In fact, the identification of the most active electrodes detected by our model gives us a deep understanding of how the brain reacted with regard to emotional elicitation. Furthermore, we note that the P8 is a common electrode for all emotion categories. The P8 sensor position is localized on the parietal lobe of the right cerebral hemisphere of the brain. This result is consistent with the study of Sarkheil and his colleagues (Sarkheil et al., 2013), and confirms the role of the right IPL (Inferior Parietal Lobule) in decoding dynamic facial expressions. So, the right IPL is not only involved in decoding the others' facial expressions but also in generating our own facial expressions. The same consistency holds for the F4 sensor that was completely excluded by our model and F3 sensor that was only used to detect Anger and Negative emotions. In fact, the two sensors are located in the frontal lobe which is, according to the study of Lough et al. (Lough et al., 2006), related to the modification of emotions to generally fit socially acceptable norms.

With these results, our EEG-based facial expressions prediction approach provides a simple and reliable way to capture the emotional reactions of the user that can be used in learning, games, neurofeedback, and VR environments.

6.5 Conclusion

This work shows that user's facial expressions are predictable from EEG physiological signals. The emotion recognition problem is considered from regression perspective taking as ground truth the detected facial expressions' data from webcam-based facial expressions recognition software (FACET). The experiment results revealed that facial expressions can be recognizable from specific EEG sensors by the mean of specific time-domain and frequency-domain features. Our experimental design and features construction method has enhanced the physiological emotions recognition accuracy reaching performances similar to computer vision technics. This finding suggests that the used EEG features were sufficiently implemented for the prediction of facial expressions from EEG with high accuracy. This accuracy is compared with FACET output and not against the self-reported emotional state of the user which is out of the scope of this current study and would be an interesting direction for further work with larger sample size. With the advances in the technology of EEG and appearance of new EEG headsets with dry sensors and wireless transmission of physiological data to mobile applications (Chi et al., 2012; Samsung, 2015), emotions assessment with EEG equipment will be more common in our daily life. Therefore, we are considering the integration of our models in a virtual reality environment to test their performances in real-time conditions and detect the user's facial reactions even with VR headset that hides his face.

ACKNOWLEDGEMENTS

The research presented in this paper has been supported by funding awarded by the Natural Sciences and Engineering Research Council of Canada (NSERC) and Beam Me Up Games.

Chapitre 7 : Reconnaissance des émotions dans un environnement de jeu vidéo

Dans les quatre chapitres précédents, nous avons présenté notre première contribution qui consiste à analyser les émotions et l'engagement dans différents environnements virtuels tels que les environnements de débat en ligne et les environnements de jeu vidéo et à faire la reconnaissance des émotions (expressions faciales) seulement à partir des signaux physiologiques (EEG). Ce chapitre et le suivant reflètent notre deuxième contribution dans les environnements de jeux vidéo qui est la réalisation d'un système permettant d'analyser et de classifier l'émotion dominante et la motivation dans les scènes de jeu en fonction des caractéristiques du joueur (traits de personnalité et type de joueur) et de la description de la scène de jeu (par des variables du modèle OCC).

Ce système a été développé en python par René Doumbouya dans le cadre de sa maîtrise. Ce système d'analyse multimodale de sessions de jeu, nommé « EmoGraph », intègre les données venant de plusieurs sources telles que les états mentaux extraits du casque Emotiv (EEG), les zones d'intérêts (AOI) des scènes de jeu venant du traceur de regard Tobii TX300 et la mesure des expressions faciales.

Lors des expérimentations, nous avons utilisé un jeu commercial d'horreur, nommé « Outlast » comme stimulus d'interaction. Dans cet environnement, le joueur prend le rôle d'un journaliste qui explore un hôpital psychiatrique hanté ou il rencontre plusieurs personnages horrifiques, il ne peut que s'enfuir ou se cacher, sa mission est d'investiguer et de documenter les expériences inhumaines subies par les patients pendant la guerre. Cet environnement stimule chez le joueur la peur, la curiosité et l'immersion à travers son aspect 3D et les aspects de recherche d'indices des évènements dans différentes scènes de l'environnement du jeu. En utilisant le jeu « Outlast », nous avons réalisé une étude expérimentale où nous avons collecté les données de 20 participants, de l'Université de Montréal qui ont l'habitude de jouer à des jeux vidéo. L'expérimentation est de durée moyenne de 45 minutes par participant pour un total de trois semaines.

L’outil EmoGraph permet d’identifier l’émotion dominante du joueur en l'associant à l'évènement dans la scène et de prédire ses émotions pour de nouvelles scènes. Il est important d'intégrer les théories sur les émotions, ainsi que les théories sur les différences individuelles dans le design de jeux. Il est plus avantageux de centrer le design des jeux vidéo sur les émotions des joueurs. En fait, les actions et les sentiments des joueurs ne sont pas les mêmes, même si nous les plaçons dans la même situation. Par conséquent, il est intéressant dans nos recherches de comprendre: «Quels sont les éléments précis de l'interface que le joueur regarde et avec lesquels il interagit?» Cette question concerne l'émotion dominante du joueur dans une scène de jeu.

C'est pourquoi nous avons proposé une méthode de sélection de l’émotion dominante dans la scène à partir des expressions faciales et un questionnaire après le jeu et entrainé des algorithmes d'apprentissage machine pour la prédiction de l’émotion dominante. Après la session de jeu, les participants ont répondu à un questionnaire en sélectionnant l’émotion ressentie par scène de jeu. Ces réponses ont été prises en considération dans la sélection de l’émotion dominante par scène de jeu. De même nous avons proposé une formalisation de la description de scène de jeu en utilisant des variables du modèle OCC (A Ortony et al., 1988), ce qui revient à faire une représentation de la scène par une combinaison de valeurs numériques. Le modèle OCC est composé de trois types de variables : globaux, centraux et locaux. Les variables globales sont incluses dans toutes les situations, tandis que les variables centrales et locales sont spécifiques à certaines situations caractérisant leur contenu informationnel. Ainsi, nous pouvons représenter formellement une scène de jeu avec un vecteur de 16 valeurs numériques. Grâce à sa généralité, la représentation modèle de la scène OCC reste applicable aux scènes de jeu à des fins d’apprentissage ou de divertissement.

Dans cette étude, nous utilisons un module de reconnaissance de l'expression faciale combiné à l'électroencéphalogramme (EEG) et aux systèmes de suivi du regard. Avec les données EEG on a calculé le ratio du FAA (*Frontal Alpha Assymetry*) pour déterminer si le comportement est relié à l’approche ou l’évitement (Coan & Allen, 2003). Nous avons aussi collecté les données sur la personnalité en utilisant le questionnaire Big Five (John & Srivastava, 1999). Le modèle utilise un vecteur de caractéristiques à 27 dimensions (comme variables d’entrée : 16 variables du modèle OCC, 4 variables sociodémographiques (sexe, âge, appartenance ethnique et catégorie de joueur), 5 valeurs de trait de personnalité, l'approche/évitement et comme résultat de sortie: l’émotion dominante dans la scène).

Nous avons utilisé deux algorithmes d’apprentissage machine (kNN et Random forest) pour faire la prédiction. Nous avons utilisé deux approches pour prédire l’émotion dominante (Approche individuelle et l’approche générale). Pour l’approche individuelle, le modèle formé est spécifique au participant (arrive à 90% de precision avec RF) ou à la scène (avec une précision de % avec RF) avec une validation croisée Leave-one-out (LOOCV). Pour l’approche générale, on utilise un seul dataset contenant 335 exemples et une méthode de validation croisée par 10-fold. Par rapport à KNN, Random Forest réalise la meilleure performance avec une précision moyenne de 96% pour l’approche générale.

Le reste de ce chapitre est constitué de l'article intitulé « *Game Scenes Evaluation and Player’s Dominant Emotion Prediction* » publié dans la conférence International Conference on Intelligent Tutoring Systems, ITS 2018. Nous rappelons que ma contribution essentielle consiste à la conception de l'expérimentation, à utiliser des variables du modèle OCC pour représenter la scène de jeu, à déterminer le vecteur d’entrainement pour la tache de prédiction et à la rédaction du papier.

Game Scenes Evaluation and Player's Dominant Emotion Prediction

Doumbouya R., **Benlamine**, M.S., Dufresne, A., & Frasson, C. (2018). Game Scenes Evaluation and Player's Dominant Emotion Prediction. T*he 14th International Conference on Intelligent Tutoring Systems* ***ITS-2018***, Montreal, Quebec, Canada, June 13-15, 2018. Springer International publishing.

https://link.springer.com/chapter/10.1007/978-3-319-91464-0_6

Abstract. In this paper, we present a solution for computer assisted emotional analysis of game session. The proposed approach combines eye movements and facial expressions to annotate the perceived game objects with the expressed dominate emotions. Moreover, our system EMOGRAPH (Emotional Graph) gives easy access to information about user experience and predicts player's emotions. The prediction mainly uses both subjective measures through questionnaire and objective measures through brain wave activity (electroencephalography - EEG) combined with eye tracking data. EMOGRAPH's method was experimented on 21 participants playing horror game "*Outlast*". Our results show the effectiveness of our method in the identification of the emotions and their triggers. We also present our emotion prediction approach using game scene's design goal (defined by OCC variables from the model of emotions' cognitive evaluation of Ortony, Clore and Collins (Andrew Ortony et al., 1990)) to annotate the player's situation in a scene and machine learning algorithms. The prediction results are promising and would widen possibilities in game design.

7.1 Introduction

Nowadays, thanks to technological advances, video games continue to grow and become based on more sophisticated rules and more realistic elements to captivate and engage players during interactions. To advance this field, video game research should focus more on user experience by becoming more aware of the player's emotions. From a human-centred computing perspective, the end-user's emotional reaction within learning/gaming environments is a critical challenge to design for, with empirical evidence already showing that not all the used practices are necessarily effective from a learning/gaming perspective (Jason Matthew Harley, 2016). It is important to analyze the emotional dynamics in video games because it enhances the player's feeling of presence and immerses him in an intense emotions experience more than other media (Abdessalem & Frasson, 2017; Hemenover & Bowman, 2018; Villani et al., 2018).

With scientific and technological progress, detecting emotional reaction is now more accessible via tools based on ocular devices (webcam, infrared camera, Kinect, eye tracker …) or physiological sensors (EDA, HR, EMG, EEG, Respiration rate …). With these tools, it is possible to capture physiological data and recognize player's emotions. In the literature, emotions generation can be described by three components: cognition, physiological processes and interactions between them (Gratch, Marsella, & Petta, 2009). Nevertheless, the most applied theories are of cognitive evaluation as reported by Scherer (Scherer, 1986) and Lazarus (Lazarus, 1991). These theories focus on the cognitive perception of situations or events generating emotional responses. Several researchers suggest that emotional reactions are regulated by two basic motivational systems: Avoidance system and Approach system (Elliot & Covington, 2001). Moreover, many studies (Amodio et al., 2008; Davidson, 2004) associate the Frontal alpha asymmetry (FAA) EEG measure with Approach and Avoidance tendencies and individual differences in personality and also other studies (Derbali & Frasson, 2010; Derbali et al., 2013) underscore the role of prefrontal cortex in emotion process.

In this paper, we aim to answer the following research questions: (1) what is the connection between the game elements (or scenes) perceived by the participant and his emotional state? And (2) can we predict the generated dominant emotion from a scene based on the characteristics of the player, his Approach/Avoidance behavior and the design objective of the scene (the emotions targeted by the designer).

To answer these questions, we propose an evaluation system using eye tracking data and facial expressions to make the association between game elements and the dominant emotion. This paper describes our resulting system that matches between the scenes during a game session, and the players' emotion. We designed experiments with 21 participants who played a commercial horror game: *Outlast*. During the game session, the participant was equipped with webcam based facial expressions detection tool, recording his emotions and an eye-tracker recording his gaze on the screen. To address the prediction issue, we used a theoretical model of cognitive evaluation of emotions Ortony, Clore and Collins (OCC) (Andrew Ortony et al., 1990) for the annotation of the player' situation in the scene enriched with the player characteristic (socio-demographic information, and personality trait) as an input to the Machine Learning algorithm to predict the player's dominant emotion when interacting with a game scene.

7.2 Background on video games and emotion analysis

The reactions of a person playing are different from one game to another and the design of video games is able to influence the affective state of the player (Kors et al., 2016; L. E. Nacke et al., 2011; Quick et al., 2012; Sweetser, Johnson, & Wyeth, 2012).

7.2.1 Game elements and generated emotions

It's more advantageous to make video game design more centered on players' emotions. In fact, the players' actions and feelings are not the same even if we put them in the same game situation. Player's emotions need to be analyzed when interacting with game elements (aesthetics, visual and audio elements) during a game session. Therefore, it is interesting in our research to understand: "*what are the precise elements in the interface that the gamer is watching and interacting with*". For that reason, we have opted to use eye-tracking techniques in this study.

7.2.2 Measuring emotions

Used techniques can be categorized on subjective and objective measures.

- Subjective measures consist of self-evaluation of the person's emotions; the users report their own emotions. Different types of subjective measures can be used open-ended, multiple-choice, and Likert-scales items in surveys. As a famous example, we can cite the Self-Assessment Manikin (Bradley & Lang, 1994).

- Objective measures consist of capturing and analyzing the signals coming from the player's body and face. Different tools can be used such as cerebral activity EEG (M. Benlamine et al., 2016), skin conductance (Stemmler, 2004) or facial expressions recognition (Littlewort et al., 2011).

7.2.3 Learning with intense emotions games

Learning is a process that always involves emotional component because it is mostly treated in the limbic system (especially Hippocampus, Thalamus, Hypothalamus and Amygdala) (LeDoux, 1992; Zirbel, 2014). In pedagogy, teachers are encouraged to instill an environment that promotes positive emotions to activate the hippocampus (for information processing and transfer to the prefrontal cortex) rather than to create an environment of fear and stress - thus activating the amygdala (Defensive survival circuit) (LeDoux, 2017; Zirbel, 2014). In games (e.g. MMORPG, adventure and horror games), in situations like facing hard enemies, the use of violent messages to push the player to the limit may lead to the success of the team. Teacher sometimes requires laying down strict guidelines which is pedagogically not advised if used directly in the classroom. So game agents playing the role of teacher can use some directive language (maybe sarcastic) to push students reach challenging levels in a game where they have fictive pseudonyms (to make things not personal). This method may lead the students to adopt the group's objective and do more practice to meet that objective (guided by the teacher) because in such environment players accept to be exposed in such situations and may enjoy it (Lin, Wu, & Tao, 2017; Lynch & Martins, 2015). This field is less explored in educational research and needs more attention.

7.3 Experimental settings

7.3.1 Participants

This study involves 21 participants (12 males; 9 female), aged between 18 and 35 years from a North American university. We have discarded 2 participants due to technical problem while collecting data. We categorized the players according to the number of hours of play per week (5 extreme players, 6 intermediates and 8 novices).

7.3.2 The game - Outlast

In this study, participants were asked to play the first level of a horror commercial game named *Outlast* (developed by Red Barrels Games). Outlast allows players to embody Miles Upshur, an investigative reporter sent to find the truth about the company Murkoff Corporation. As a reporter, the player must find a way to enter the asylum before starting his investigation. The only means of the player's survival are: "flee or hide" armed only with his camera. The player is placed in strange game situations that influence his emotional reactions. The participants were required to play the first stage of the game.

7.3.3 Experiment and equipment

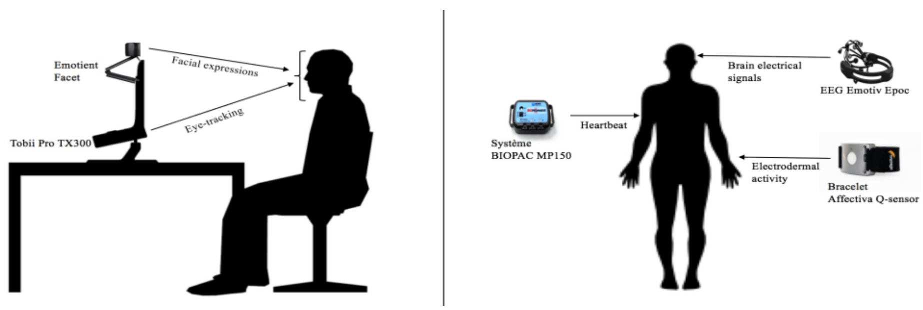


Figure 23. The experimental design.

The experimental scheme is presented in Figure 23 showing the positions of the sensors placed on the body (right) and those integrated in the experimental computer (left). In this study, we use facial expression recognition module combined with electroencephalogram (EEG) and the eye-tracking systems.

7.3.4 Measures

Different studies have used several methods, subjective or objective, in order to evaluate the emotion in game scenes.

7.3.4.1 Subjective measures

To gather as much information as possible to complete our study, forms were submitted to the participants before and after the game. Two forms were filled before session: the first questionnaire for collecting socio-demographic data (age, gender and ethnicity), school level and hours of play per week and the second was the "Big Five" questionnaire (Goldberg, 1992) for the assessment of the participant's personality traits. In the post-test, we used the immersion and flow questionnaire. The questionnaire was adapted from the GameFlow questionnaire (Sweetser et al., 2012). In addition, participants were asked to answer a final questionnaire about their emotions felt during stages of the game.

7.3.4.2 Objective measures

We collected multimodal data from the player's body and face to analyze his affective and mental state. In this study, we are using among these measures: EEG data to compute the *Frontal Alpha Asymmetry* (FAA), facial expressions to extract dominant emotion and eye-tracking data to detect the scenes visualization time and duration.

Frontal Alpha Asymmetry

Neuroscientists have found that higher alpha power (8-12 Hz) of the left compared to the right frontal brain is related to positive feelings (Coan & Allen, 2003).

Facial expression analysis

The iMotions FACET software generates numerical values for each basic emotion. These values scaled between [-5, 5] are the log likelihood of the presence of an emotion. For example, a joy value of 4 means a big smile and any human expert would see that. A joy value of 0 means that the observed expression is equally categorized by human experts as ‘’*being*’’ and ‘’*not being*’’ joyful.

Eye-tracking analysis

We used the software iMotions to replay and annotate chronologically the participants’ game session by defining Areas of Interest (AOI) in order to get gaze statistics by AOI. Participants may take different durations for each scene during their course in the game. Using hit time and the time spent metrics, we identified the time and duration that the player spent for each scene.

7.4 Method

Multimodal analysis was performed using several sources of information to determine players' emotional reactions. This require the use of statistical analyzes and AI techniques to construct the player's dominant emotion model.

7.4.1 Dominant emotion extraction

Even if the player has a fixed path in the game, the participants don't look at all the AOIs. The participant's dominant emotion is identified using the following steps:

- Every time window of 500 milliseconds (Libet, 2006), calculate the median of each emotion and assign 1 to the emotion if its median exceeds a relative threshold otherwise 0.
- For each AOI, sum the binary scores reported for each emotion.
- Select the emotion with the highest score: Multiple emotions can occur in the same time window, we attribute the score " 1 " to the emotion whose median is the highest. In the case of having equal emotional counts for an AOI, we choose as dominant emotion from the player emotion self-report.

At the end, EMOGRAPH stores, in a MySQL database, information about the participant, visualized AOIs, and the corresponding dominant emotion. This database is used by the player's experience visualization and the emotion prediction modules.

7.4.2 The FAA computation

The frontal asymmetry index was computed from raw frontal EEG data using electrodes F3/F7 and F4/F8.We calculated FAA using the formula below:

$$FAA = \log\left(\frac{Alpha\ Power_{Right} - Alpha\ Power_{Left}}{Alpha\ Power_{Right} + Alpha\ Power_{Left}}\right)$$

Higher scores on this asymmetry index indicate greater relative left hemisphere activation which means that the player's behavior in the scene is APPROACH otherwise it is AVOIDANCE.

7.4.3 OCC game scene representation

To predict the player dominant emotion for new game scenes, we need first to characterize the game scenes according to the designer perception and goals using variables from OCC model

(Andrew Ortony et al., 1990). Cognitive variables characterize a person's interpretation of a situation. In fact, the emotional response depends on the situation interpretation as desirable or undesirable, expected or unexpected, etc. The OCC model has its own descriptive variables divided into 3 categories: global, central and local variables. Global variables are included in all situations, whereas the central and local variables are specific to a certain situation characterizing their informational content.

Table 18. Global, central and local variables and their associated values.

	Global variables		Central variables			Local variables										
Evaluation Variable	Surprise	Sense of reality	Desirability	Approval	Attraction	Desirability by other	Esteem for other	Merit for other	Likelihood	Realization	Effort	Agent	Power of the link	Deviation	Disposition	Familiarity
Values	0 (False) or 1 (True)	0 (False) or 1 (True)	-1, -0.5, 0, 0.5, 1	-1, -0.5, 0, 0.5, 1	-1, -0.5, 0, 0.5, 1	-1, -0.5, 0, 0.5, 1	-1, -0.5, 0, 0.5, 1	-1, -0.5, 0, 0.5, 1	-1, -0.5, 0, 0.5, 1	-1, -0.5, 0, 0.5, 1	-1, -0.5, 0, 0.5, 1	0 (other) or 1 (self)	0, 0.5, 1	-1, -0.5, 0, 0.5, 1	0, 0.5, 1	-1, -0.5, 0, 0.5, 1

This model identifies the involved variables and allows the description of a game scene by precise variables summarized in table above (see Table 18). The scene OCC representation is applicable to all kind of games intended to learning or entertainment.

7.4.3 The dominant emotion prediction

In this section we describe our approach to training and evaluating classifiers for the task of detecting the player dominant emotion given formal description of a scene, Approach/Avoidance player behavior and his demographical data.

7.4.3.1 Training set construction.

In order to train a scene's dominant emotion predictive model, we used a training set containing scene descriptions, participant's socio-demographic information, personality traits, Approach/Avoidance behavior in scene and also the participant's dominant emotion. In our study, we targeted two objectives for predicting the dominant emotion of:

1. An existing player having as input description variables of a new scene defined by the designer.
2. A new player (defined by his socio-demographic information, player category, personality trait and Approach/Avoidance behavior) in front of an existing scene.

For each of the approaches, the models use 27-dimensions vector: 16 variables from the OCC model, 4 socio-demographic variables (gender, age, ethnicity, and player category), 5 personality trait variables, 1 variable for Approach/Avoidance and the dominant emotion. The approaches have been developed with Python language and scikit-learn library.

Individual approach. The prediction is based on the emotional reactions of individuals during interaction with game scene. The trained model is individual, meaning that it is specific to the participant or the scene depending on the objective of the model. We have trained and validated our models using the leave-one-out cross-validation (LOOCV) method because of the size of the individual dataset.

General approach. The prediction is based on a unique dataset (contains 335 examples) gathering all the emotional responses of participants in all scenes. We have trained and validated our model using the k-fold cross-validation method. We found the optimal number of blocks with a grid search.

7.5 Results

7.5.1 Game analysis results

Area Of Interest	Dominant emotion	Emotion values	Pic	Time of pic	Probabilities	Time of happening
Entrance gate	anger	2,49098444	2,7644091	0:02:51.777000	0,996781781	0:02:59.248000
First garden	anger	2,59793206	2,7852449	0:03:35.985000	0,997482478	0:04:58.985000
Closed door of the first garde	anger	1,8406873	2,009673	0:04:00.384000	0,985773766	0:04:03.376000
Small gap	surprise	1,8360396	2,0368406	0:06:27.649000	0,985622904	0:06:33.641000
Second garden	surprise	2,2435746	2,8181016	0:06:40.080000	0,994325157	0:07:49.076000
Ladder	joy	1,11826675	1,94129675	0:07:12.080000	0,929228898	0:07:12.580000
Dark living room	surprise	1,5929086	2,0025576	0:08:19.924000	0,975103287	0:08:27.424000
Living room 2 (Tv)	surprise	2,1482006	2,3136126	0:08:52.721000	0,992941328	0:09:07.221000
Bloody corridor	surprise	2,0635346	2,2020056	0:09:41.046000	0,991434955	0:09:45.513000
Office 1	anger	2,3163624	2,3163624	0:09:45.248000	0,995196625	0:09:48.241000
Kitchen	fear	2,373933	2,375858	0:10:30.481000	0,995790457	0:10:48.981000
Bloody aeration pipe	joy	2,29445975	2,58997975	0:10:33.313000	0,99494942	0:10:36.313000
Aeration pipe	surprise	2,2290766	2,4041826	0:10:56.819000	0,994133647	0:10:57.819000
Library	anger	2,47712497	2,8704365	0:11:44.028000	0,996677769	0:12:07.013000
Hanging corpse	surprise	1,9114076	2,5582846	0:11:42.785000	0,987885679	0:11:43.285000
Speaking corpse	surprise	2,7199446	3,2812686	0:13:19.840000	0,998097921	0:13:27.317000
Monster	joy	0,23861945	0,23861945	0:14:50.912000	0,634006026	0:14:51.912000
NPC	sadness	1,38494255	1,8665364	0:15:13.153000	0,960416249	0:15:18.152000
Blood in the hall	sadness	1,9187056	2,1609914	0:15:39.041000	0,988085145	0:15:40.030000

Figure 24 Dominant emotions for participant P21

Here above, the Figure 24 shows the dominant emotions of participant (P21) obtained from the EMOGRAPH system. On this example it is a woman, beginner in video games. We see also the moment of appearance of the peak. For example, the orange colored line shows that the dominant emotion determined for the stage "Speaking corpse" is the surprise and the relative time of occurrence of the peak is at 00:13:19:8400 from the beginning of the game session' video. A screenshot of participant P21 at the "speaking corpse" step, illustrated by Figure 25, is representing the orange line in Figure 24. On this capture, we note the expression of surprise on the face of the participant when seeing the hanged corpse marked by his gaze. At the bottom, on the right, the time of the video zoomed and circled in red is equivalent to that recorded by EMOGRAPH (Figure 25). EMOGRAPH offers the visualization of the dominant emotion frame extracted from the game session video using the peak time. These game screen-captures help game designer to better analyze the player experience.

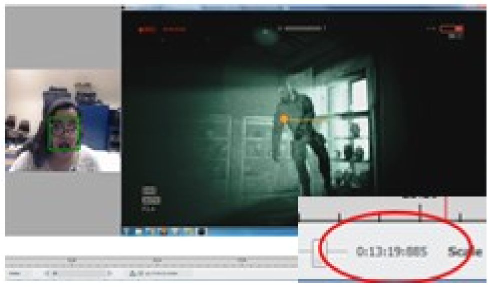


Figure 25. Game screenshot of participant P21.

7.5.2 EMOGRAPH: System interface.

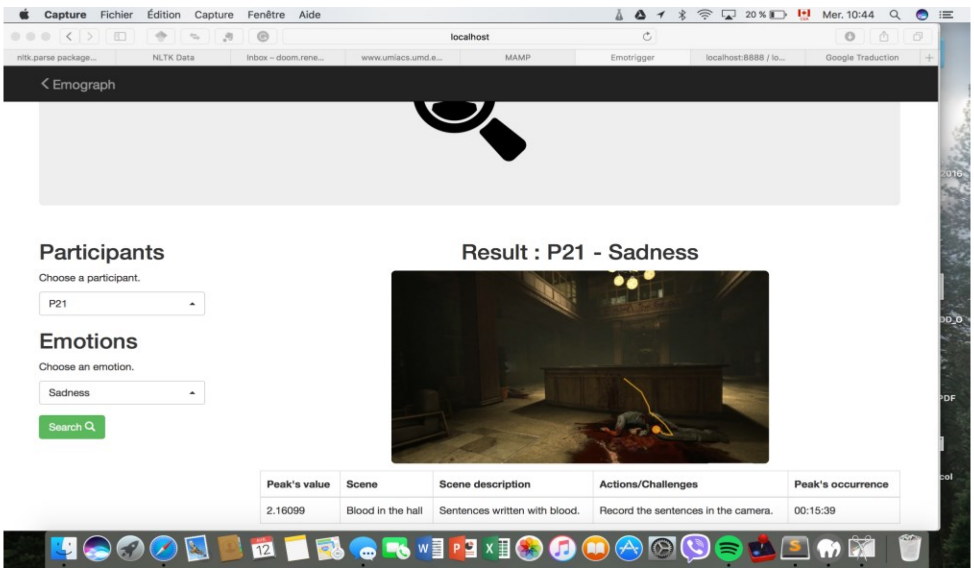


Figure 26. Emotional analysis module.

We present the interfaces and the operating mode of the application's modules. Figure 26 shows the emotional analysis module offering to the game designer the possibility of knowing the dominant emotions of the players and gives additional information about the scene. The interface proposes a search by participants or by AOI. The visualization module that displays the emotional graph is presented in Figure 27. Additionally, by clicking on a node in the visualized emotional graph, it shows a fragment of 30 seconds extracted from the game session video at the time of the dominant emotion.

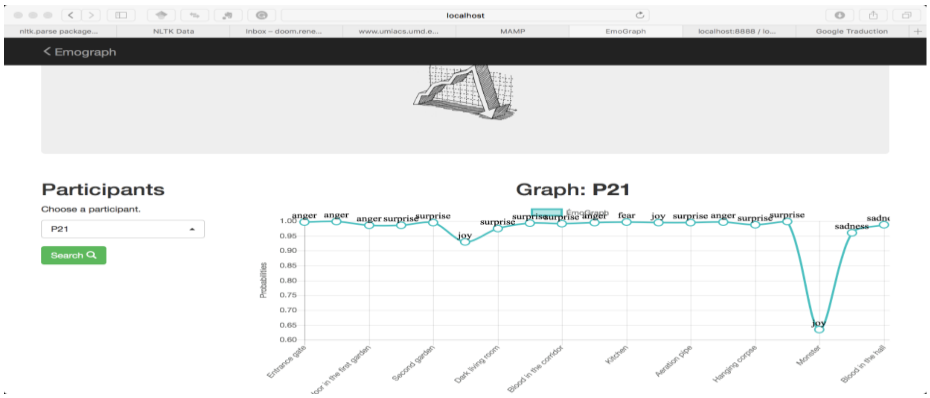


Figure 27. Emotional transition graph module.

The emotional transition graph emphasizes the dominant emotions related to game scenes. The emotions values are presented graphically after being transformed into probabilities (between 0 and 1), which facilitates their interpretation and presentation. This transformation is made according to the following formula: $PP = \frac{1}{1+10e^{-LLR}}$ with LLR being the numerical value (evidence) of the considered emotion. Figure 27 shows the emotion graph for participant P21 displayed by our system. The scenes in the game sequence are on the abscissa axis while the probabilities are on the ordinate axis. Each point gives the dominant emotion of the scene.

7.5.3 Prediction results

7.5.3.1 Individual approach

For the distance-weighted k-NN algorithm, the validation method LOOCV allows each participant to choose the optimal number of neighbors (k) by testing the performance over several values of k. We vary k between 2 and 7. For the random forests algorithm, we found with the same validation method the optimal parameters for the prediction of the model. A leaf with a few examples makes the model more likely to capture noise in training data.

Objective 1: existing person vs new scene. The performance of our model is evaluated individually; the accuracy varies from one participant to another. Accuracy scores range from 42% to 90%. The average accuracy on all participants is 85% with the distance-weighted k-NN algorithm. Accuracy ranges from a minimum of 75% to a maximum of 95%. The average accuracy is 90%, in terms of the random forest algorithm.

Objective 2: existing scene vs new person. For this purpose, precision varies from scene to scene with scores ranging from 27% to 59% accuracy. The average accuracy on all scenes is 30% with the distance-weighted k-NN algorithm. The accuracy ranges from 40% to a maximum of 80%. The average accuracy is 64%, with the random forest algorithm.

7.5.3.2 General approach.

In this approach, we have varied the number of blocks for validation between 2 and 10 for the distance-weighted k-NN and between 5 and 10 for random forests classifier to optimize the parameters (or Hyper-parameters) of the algorithms. For the distance-weighted k-NN algorithm, the number of blocks that gave the highest accuracy rate is K * = 2 with the optimal neighbor number (k) of k * = 4. The accuracy rate is 84 %, also we have varied k between 2 and 7. For the random forest algorithm, we found an optimal number of blocks of K * = 5, an optimal number of trees equivalent to 40, the maximum number of dimensions equivalent to the square root of the total number of blocks. Dimensions, maximum tree depth, the nodes are extended until all the leaves are pure or until all the leaves contain less than 2 samples. We get an accuracy that reaches **96%**.

Table 19. Summary of results from Machine Learning.

	Individual approach		**General approach**
Validation method	LOOCV		K fold
	Objective 1	*Objective 2*	*General*
Distance weighted k-nearest neighbords	85%	30%	83% with K* = 2 and *k*=4*
Random forest classifier	90%	64%	**96% with K*=5**

Table 19 summarizes the results obtained by our algorithms in the approaches. From these results, the random forest classifier algorithm has been integrated into our general approach prediction module. The EMOGRAPH's prediction module interface (Figure 28) proposes to the designers the prediction of the player's emotion, with the reliability indicator which is the f-score recorded during training / validation.

Figure 28. Emotional prediction module.

The EEG data are not necessary in order to use the system. Without EEG data the system will present two outputs one for the Approach case and the second for the Avoidance case. The model can be used for both educational or entertainment games, we only need scenes descriptions with the OCC variables of the planned game in the conceptualization phase, rather than a complete game. We intend that both, designers and teachers, can gain value from this system as a step towards affective recommender system that respond to design and learning objectives.

7.6 Conclusion

In this study, we examined the interactions between video games and player's emotions using game scene's design goals, player characteristics, EEG and eye-tracking data. We presented our method in categorizing the player's behavior as "Approach" or "Avoidance" using the Frontal Alpha Asymmetry (FAA) during time window calculated from eye tracking data. We have also built a machine learning model for predicting player's dominant emotion using game scene's design goal (defined by OCC variables), Approach/Avoidance behavior and the player's personality traits (using the Big Five questionnaire). Based on this experience, we proposed a system for emotional analysis of game session. The proposed tool allows the identification of the dominant emotion expressed by a gamer with the precise time in the scene. The integration of eye-tracking in the analysis process provides another level of accuracy for the game design. The application developed includes many features for game designers. EMOGRAPH not only produces affective analysis for game session and visualizes the player's emotional transitions, but also, the application performs emotions prediction of new person toward an analyzed game scene and for registered player toward a new game scene as well.

Acknowledgments. We thank the Natural Sciences and Engineering Research Council of Canada NSERC and BMU Games for funding this research.

Chapitre 8 : Reconnaissance de la motivation dans un environnement de jeu vidéo

Dans le chapitre précédent, nous avons développé et décrit brièvement le système d'analyse émotionnelle dans les jeux vidéo « EmoGraph ». Nous avons aussi étudié la possibilité de prédire l'émotion dominante des joueurs dans les scènes de jeu en construisant des modèles de classification par des algorithmes d'apprentissage machine. Ce chapitre est une extension des travaux entrepris dans le chapitre précédent. En effet, dans ce chapitre nous présentons une nouvelle méthode d'analyse et de prédiction de la motivation des joueurs dans les scènes de jeu.

Dans cette étude, nous avons alors repris les données collectées lors de notre première expérimentation avec le jeu « Outlast » et nous avons étendu le dataset par l'identification des orientations motivationnelles (Wolters, 2004) en se basant sur le donnée EEG et les questionnaires d'auto-évaluation avant et après la session de jeu. Ensuite, nous avons utilisé des méthodes de prédiction des orientations motivationnelles des joueurs dans les scènes de jeu en utilisant des algorithmes d'apprentissage machine de la bibliothèque Sickit learn.

La théorie d'orientations motivationnelles est une théorie de motivation (Wolters, 2004). Elle vient du domaine de l'éducation et elle regarde comment les objectifs de l'étudiant affectent sa motivation. Par exemple, un étudiant a un but à long terme (*Core goal*) qui est d'être un médecin. Pour achever son but, l'étudiant planifie des buts à court terme (*Proximal goals*), par exemple avoir de bon résultats au niveau du secondaire, faire du bénévolat dans les hôpitaux, suivre des cours libres de médecine pour être accepté dans une université de médecine. Cette théorie distingue quatre orientations de but: (i) **Maîtrise-Approche** lorsque les individus cherchent à atteindre la maîtrise ou l'amélioration de soi, (ii) **Maîtrise-Évitement** lorsque les individus cherchent à éviter d'échouer à la maîtrise d'une tâche, (iii) **Performance-Approche** est que les individus accomplissent et surpassent les autres, et (iv) **Performance-Évitement** lorsque l'on cherche à éviter de faire pire que les autres dans des tâches données.

Dans le but de reconnaitre les orientations motivationnelles du joueur dans une scène de jeu, nous répondons aux questions suivantes : (1) Comment évaluer le but de maîtrise (Maitrise/Performance) du joueur dans un jeu vidéo et caractériser son comportement lié à l'approche (Approche/Évitement) dans une scène de jeu? (2) Quels sont les algorithmes d'apprentissage machine les plus adéquats pour la classification des orientations motivationnelles du joueur dans une scène de jeu?

Pour notre étude, c'est la première fois qu'on applique cette théorie dans les jeux vidéo pour divertissement. Hors, elle colle bien avec ce domaine, en regardant le profil d'un joueur (son type de jeux préféré, combien il joue par semaine, …etc), nous avons une idée par rapport à son but à long terme (*Core goal*), donc on peut estimer son but de maitrise (Maitrise/Performance) à partir de ses réponses au questionnaire de GameFlow (Sweetser et al., 2012) et lors d'une session de jeu, nous avons mesuré son Frontal Alpha Asymmetry (FAA) issu de l'EEG qui détermine son comportement (approche/évitement) dans la scène de jeu et ce qui nous donne ses petits buts à court terme par scène de jeu (proximal goals).

À partir de ces données, nous avons construit un vecteurs de 24 dimensions contenant : 16 variables du modèle OCC, 2 variables sociodémographiques (sexe, âge), 5 valeurs de trait de personnalité en tant que variables d'entrée et les orientations motivationnelles. Ce vecteur de caractéristiques a été utilisé pour entrainer et tester les différents algorithmes d'apprentissage machine. Nous avons construit un modèle de machine Learning qui classifie les quatre orientations de buts du joueur par scène jeu en ayant seulement comme entrée la description de la scène (OCC), les traits de personnalité du joueur (Big5). La classification est étudiée en utilisant trois algorithmes d'apprentissage machine (Random forest, SVM et kNN). Les résultats ont montré que le Random forest a donné la meilleure précision avec 82% pour la classification des orientations motivationnelles des joueurs.

Le reste de ce chapitre est constitué de l'article intitulé « *Game Experience and Brain based Assessment of Motivational Goal Orientations in Video Games* » publié dans la conférence *International Conference on Brain Functions Assessment in Learning, BFAL 2017*. René Doumbouya a participé à l'annotation des joueurs (*Maitrise/Performance*) et la classification en appliquant les algorithmes d'apprentissage machine et a contribué dans certains résultats de cet article. Nous rappelons que notre contribution essentielle consiste à la conception de l'expérimentation, à établir une méthode d'identification des orientations motivationnelles du joueur dans une scène de jeu et à déterminer le vecteur d'entrainement pour la tache de classification et à la rédaction du papier.

Game Experience and Brain based Assessment of Motivational Goal Orientations in Video Games

Benlamine M.S., Doumbouya R., Frasson C., Dufresne A. (2017). Game Experience and Brain based Assessment of Motivational Goal Orientations in Video Games. The *First International Conference on "Brain Function Assessment in Learning", **BFAL-2017***, Patras, Greece, September 24-25, 2017. Springer International publishing.

https://link.springer.com/chapter/10.1007/978-3-319-67615-9_11

Abstract. The current study aims to measure the goal orientations motivation in different scenes of a video-game. The evaluation of player experience was done with both subjective measures through questionnaire and objective measures through brain wave activity (electroencephalography - EEG). We used GameFlow questionnaire to characterize the player's mastery goal in playing video game (Master or Performant). In terms of brain activity, we used the Frontal alpha asymmetry (FAA) to assess the player approach/withdrawal behavior with-in a game scene. Using game scene's design goal (defined by OCC variables) and player personality traits (using Big Five questionnaire), the resulting machine learning model predicts players' motivational goal orientations in order to adapt the game. In this study, we address player's motivation in game scenes by analyzing player's profile, his situation in scene and affective physiological data.

8.1 Introduction

The ultimate question that a game designer is always trying to answer is: "how to make the game more attractive for the player and hold his attention more and more?" This question is all about the **Motivation of the player** in playing the game. Motivation has been extensively studied in psychology and social science. Motivation is defined as: "*The willingness to put effort into achieving goals*" (Driver, 2011). Several researches suggest that goal-directed approach and withdrawal behaviors are regulated by two basic motivational systems: avoidance system and approach system (Elliot & Covington, 2001). As a result of nature selection, Approach-avoidance motivation is deeply embedded in our mind because it is crucial for survival to discriminate between pleasurable and rewarding stimuli that we can approach, and dangerous stimuli that we should avoid (Cosmides & Tooby, 1991). Many studies from neuroscience (Davidson, 2001; Phan et al., 2002) have proven the existence of approach-related motivation in neural circuitry. In fact, behavioral approach is localized in the left anterior cortical regions, and behavioral withdrawal is in the right anterior cortical regions using Functional magnetic resonance imaging (fMRI). Moreover, many studies (Amodio et al., 2008; Davidson, 2004; Horan et al., 2014) associate the Frontal alpha asymmetry (FAA) EEG measure with approach and withdrawal motivational tendencies and individual differences in personality and also other studies (Derbali & Frasson, 2010; Derbali et al., 2013) underscore the role of prefrontal cortex in emotion and motivation process.

During a game, players' motivation is also influenced by achievement need which is a social factor where the player is driven by the motivation to excel and make success in challenging situations. To maintain the player's flow (Nakamura & Csikszentmihalyi, 2002) during the game, the challenges should be in the same level of player's competences. When the challenge surpasses the player's skills, he will feel very frustrated, but the lack of challenge induces boredom (Chanel et al., 2008; Jennett et al., 2008), which leads to make the player unmotivated in both cases and may stop playing the game.

The goal orientations motivational theory takes consideration of both approach withdrawal behavior and mastery goal in assessing the motivation. Coming from educational psychology, this theory investigates the learner's motivation in four achievement goal orientations in educational settings (Elliot, McGregor, & Gable, 1999; Elliot & Murayama, 2008; Wolters, 2004). Up to our knowledge, this is the first approach addressing the study of players' motivational goal orientations in video game settings. In addition, it is important that such environments can detect the player's motivation in game scene and provide help in the right time to keep him playing and evolve his experience during the gameplay.

Modeling of the player's motivation during a game is a difficult process due to the complexity that exists to identify the player achievement goals and behavior. In a different situation, the same cause or stimulus provoke different behaviors depending of important factors like goal, personality and preferences. In this context a cognitive approach proposed by Ortony, Clore and Collins (1988)(Andrew Ortony et al., 1990) was used to represent the game scene goal defined by the designer. The OCC model evaluates a situation with descriptive variables (global, central and local variables), that the designer gives for each scene to represent his scene's goals.

The present paper aims to predicting motivational goal orientations in a game scene by using the scene OCC description and player's personality traits. We ask in this paper the two following research questions: how to assess the player mastery goal in playing a video game and characterize his approach related behavior within a game scene? If so, can we predict players' motivational goal orientations toward a new scene using machine learning model?

The organization of this paper is as follows: in the next section, we present an overview of the motivation theories and video games. In the third section, we explain our empirical approach in assessing players' motivational goal orientations. In the fourth section, we detail our experimental methodology. In the fifth section, we present the obtained results and discuss them. In the last section, we conclude with a discussion of the present research as well as future work.

8.2 Background in motivation and video games

8.2.1 Motivation and motives

Psychologists (Vallerand & Thill, 1993) define motivation as "*hypothetical construct used to describe the internal and / or external forces producing the initiation, direction, intensity and persistence of behavior*"; which gives an idea about the wide broad that covers the concept of motivation. Therefore researchers in their studies approach motivation by working on the narrower more precise concept of motive.

A motive is a need or desire that stimulates and directs behavior towards a goal that is expected to be satisfied. For example, going to the gym can have different motives: interest in workout, want to be in shape and to have healthy body, willingness to impress someone, desire to escape the anxiety associated to overweight, etc. It pushes you to take action to achieve your goal. Two people could produce exactly the same behavior with very different motives. Several theories have been proposed to explain the range of motives that push humans and animals to act. Some focus on physiological needs (Water, food, sleep ...), others on superior needs and social motives such as the need for success. The theory of the hierarchy of needs of Maslow (1943) (Maslow, 1943) covers all these needs. This theory gives an overview of human motives, from the most elementary to the noblest.

Among contemporary motivation theories, the theory of self-determination (Ryan & Deci, 2000), which is based on the existence of two types of motivation (intrinsic vs extrinsic), each leading to different behaviors.

8.2.1.1 Intrinsic motivation

Intrinsic motivation characterizes individuals who practice an activity for self-interest, pleasure and satisfaction (Ryan & Deci, 2000). Intrinsic motivation is characterized by an internal locus of control to meet individual needs for competence and self-determination. The motives for these behaviors have an internal origin. For example, you study because the subject interests you or you eat because you're hungry.

8.2.1.2 Extrinsic motivation

Extrinsic motivations are linked to the interests of an individual to an activity with external tending causal locus, largely directed by external factors (rewards, obligations, pressure, etc.) (Ryan & Deci, 2000). The feeling of self-determination then decreases according to whether the individual loses control over the regulation of his behavior. The motives here have an external origin. For example, you study to have good grade or to avoid having bad one.

8.2.2 Motivation theories

It's generally accepted in contemporary psychology that no instinct really motivates human behavior. In fact, most recognize that biological forces imply human motivation. Moreover, others identify social motives as additional drives that guide and direct human behavior. Unlike the primary drive, social motives are learned: they are acquired through experience and interaction with others, for example: achievement, affiliation, curiosity …etc.

8.2.2.1 Biological approaches

Drive reduction theory

Theory of motivation that proposes: a need generates a drive that the organism is motivated to reduce (Hull, 1935). A *drive* is an internal state of activation or tension generated by an underlying need that the organism is motivated to satisfy. This theory is based on the concept of homeostasis, according to which the organism tends to maintain a state of internal equilibrium essential to its survival (body temperature, glucose level, oxygen level, blood pressure ...). For example, a need to drink or eat disturbs the internal equilibrium, which gives rise to a drive that forces the organism to act to reduce tension by satisfying this need and restoring the state of internal equilibrium.

Arousal theory

This theory claims that an organism is motivated to maintain an *optimal level of arousal*. If the arousal level goes down under an optimal level, we look for increasing it and if it goes over that level, we try to decrease the arousal. If the arousal level is very low, *stimulation motives* push us to increase the arousal level. Stimulation motives refer to processes like: curiosity, exploration desire, playing, and object manipulation. Term arousal level refers to the degree of activity of the organism in a continuum that ranges from sleep to stress through various degrees of awakening, alertness and alertness. Stimulus with high intensity (like high noise, flashing lights…), stimulants (like caffeine, nicotine, cocaine …), emotions (like anger, joy, surprise ...) or biological needs increase the arousal level. According to law Yerkes Dodson (Diamond et al., 2007; Teigen, 1994; Yerkes & Dodson, 1908), there is a relationship between the arousal level and performance by attention and concentration, but only up to a point. A task is accomplished more efficiently when the arousal level suites the degree of difficulty: simple and routine tasks requiring a relatively high arousal level (to increase motivation); the moderately difficult tasks ask an average level of arousal and difficult and complex task, ask a lower arousal level (to facilitate concentration).

8.2.2.2 Personality and social motives approaches

Achievement Motivation theory

The achievement need is an important dimension of human motivation; it is a desire to accomplish something difficult, and to excel in it. This need is influenced by internal drive for action (intrinsic motivation), and the pressure exerted by the expectations of others (extrinsic motivation). According to Henry Murray (1938) (Murray, 1938), this need is particular since the more success is made, the more the person is motivated to make more achievements. According to researchers (Conroy, 2001; David C McClelland et al., 1976; David Clarence McClelland, 1984), achievement-motivated people set ambitious but nevertheless realistic and achievable goals.

For these people, easy-to-reach goals are irrelevant because the easily attained success is not a genuine achievement. On the other hand, they avoid setting unrealistic goals and taking risks too high, which would be a waste of time. According to other research (Elliott & Dweck, 1988), low achiever people take no risk since they are motivated by the fear of failure rather than by the success possibilities. Their goals are either very easy to make or very difficult so they are not embarrassed by the failure.

Goal orientation theory

According to the goal orientation theory, the motivation to succeed varies depending on whether the goal is one of mastery (goal defined according to self) or performance (goal defined by comparison with others), and according to whether it aims at Approach (get something nice) or avoidance (avoid something unpleasant). This theory distinguishes **four goal orientations**: (i) *mastery-approach* where individuals seek to achieve mastery or self-improvement, (ii) *mastery-avoidance* where individuals seek to avoid failing achievement of a task mastery, (iii) *performance-approach* where the main focus is for individuals to accomplish and outperform others, and (iv) *performance-avoidance* where one seeks to avoid doing worse than others in given tasks (Elliot, 1999; Elliot & Covington, 2001; Wolters, 2004).

8.2.3 Motivation and video games

Looking at how users are engaged in playing with intrinsically motivating games, Malone (1981) (Malone, 1981) has been interested in studying the theory behind intrinsically motivating learning, or learning to which the individual engages without Motivation (rewards or punishments). He describes the characteristics of environments that make them intrinsically motivating, with individual motivations such as challenge, fantasy, curiosity and control (Malone & Lepper, 1987). These characteristics can be considered as theories on how to make learning fun (Lepper & Malone, 1987). In more recent research (Przybylski et al., 2010; Ryan, Rigby, & Przybylski, 2006) Przybylski, Rigby and Ryan identified two other motivational factors associated with games, autonomy and competence, originating from self-determination theory (Ryan & Deci, 2000). Games with motivating experience can be explained by the concept of flow (Mihaly, 1990), a term invented by Csikszentmihalyi to describe a condition in which a person experiences a challenge that extends its competences without being too difficult nor too easy (engaging in an appropriate level of challenge depending on the player's skill) and have clear objectives and immediate feedback on progress (Nakamura & Csikszentmihalyi, 2002). In his book "A Theory of Fun for Game Design" (Koster, 2013), the game developer Ralph Koster stressed out the importance of integrating psychological theories in game design. Additional study (Yee, 2006) on MMORPG games, finds that factors addressed by extrinsic motivation theories (rewards) and intrinsic motivation (exploration, social needs, competence and mastery) contribute to players game enjoy. These findings actually provide some evidence supporting that intrinsic and extrinsic motivations are a false dichotomy.

8.3 Experimental settings

8.3.1 Participants

This study involves 21 participants (12 males; 9 female), aged between 18 and 35 years from a North American university. We have discarded 2 participants due to technical problem while collecting data. The players were categorized according to hours of play per week (5 extreme players, 6 intermediates and 8 novices).

8.3.2 The game - Outlast

In this study, participants were asked to play the first level of a horror commercial game named *Outlast* (developed by Red Barrels Games). This first-person game takes the player in a horrifying hospital full of dead bodies and monsters. The player takes the role of an investigative reporter that gathers evidences against the *Murkoff Corporation* who made horrible experiments on mental patients in the hospital. The only means of survival are to flee the enemies or to hide from them, the player cannot attack them. This genre of games stimulates the player's emotional reactions and his approach/avoidance motivation.

8.3.3 Experiment and equipment

The Experiment begins by receiving the participant in our laboratory; we introduce him to the testing room. To avoid interferences with the equipment, the participant was invited to turn off his phone. The participant signs a consent form to register his experiment agreement. After installing the EEG and EDA (Electro-dermal activity) sensors on the participant, the researcher checks the webcam recording the user's face and the eye-tracker calibration. The experiment Design was conducted in the platform iMotions that allows the multimodal synchronization of different sensors. By clicking the start button, iMotions platform launches the data recording and the game where the participant plays the first level.

8.3.4 Measures

Because of the complexity of the motivation concept and its components, it is difficult to measure motivation. Different studies have used several methods, subjective or objective, in order to evaluate the motivation.

8.3.4.1 Subjective measures

We have used questionnaires before and after the play session to get the users information and their game evaluations. We first gave pre-test questionnaires to collect socio-demographic data and the player's profile (school level, their preferred games and hours of play per week) and also the "Big Five" questionnaire (Goldberg, 1992) for the assessment of the participant's personality traits (openness, neuroticism, extraversion, agreeableness and conscientiousness). In the post-test, we used the immersion and flow questionnaire. The questionnaire was adapted from the

GameFlow questionnaire (Sweetser et al., 2012; Sweetser & Wyeth, 2005). These questionnaires are composed of Likert-type items where the answer is expressed on a scale between 1-"disagree at all" and 7-"completely agree". In addition, participants were asked to answer a final questionnaire about their emotions felt during stages of the game. In fact, post-test questionnaires can serve as an indicator of player's motivation. Since, improved motivation may bring improved performance of the player.

8.3.4.2 Objective measures

We believe that monitoring and analyzing objective measures like ocular and physiological signals is the most appropriate methodology for emotional and cognitive states recognition in the gaming context. We have collected multimodal data from the player's body and face to analyze his affective and mental state. In this study, we are using among these measures: EEG data to compute the *Frontal Alpha Asymmetry* (FAA) and eye-tracking data to detect the scenes visualization time and duration.

Frontal Alpha Asymmetry

The frontal asymmetry (FAA) is an unfiltered and unbiased phenomenon associated with emotion and motivation. Brain scientists have consistently found that higher engagement of the left compared to the right frontal brain is related to positive feelings and higher engagement (Coan & Allen, 2003). Due to the inverse relationship between alpha power (8-12 Hz) and cortical activity, decreased alpha power reflects increased engagement. The special effect of the asymmetry in frontal alpha power was initially detected in studies investigating biomarkers of personality (Hagemann et al., 2002). While this "emotional" effect was found to be indicative of a personality trait (supposed to be very stable across the life span), recent evidence suggests that it also varies depending on emotional stimulation, reflecting whether or not someone is drawn towards or away from something or someone. In short, this "approach/avoidance effect" reflects someone's motivation (Harmon-Jones, Gable, & Peterson, 2010).

Eye-tracking analysis

For each participant, we annotated his play-session by defining Areas of Interest (AOI). The AOIs refer to the game stages/scenes. Through the game, a participant may take different duration for the different game scene. We used the software iMotions to replay and annotate chronologically the participants' game session. The annotation allows the software to compute the gaze statistics by AOI. Using hit time and the time spent metrics, we identified the time and duration that the player spent for each scene.

8.4 Method

Thus, we performed a multimodal analysis using several sources of information to determine the level of affective reactions of the players. We used several sources of information such as: questionnaires, eye tracking and EEG physiological data. This require the use of statistical analyzes and AI techniques for the selection and extraction of characteristics and the construction of the player's motivational model.

The EEG headset consists of 14 data-collecting electrodes and 2 reference electrodes, located and labeled according to the international 10-20 system. The participant's EEG data was calibrated using his neutral state of mind when looking at gray screen for 6 seconds which is considered as our baseline period.

8.4.1 The FAA computation

The frontal asymmetry index was computed from raw frontal EEG data using electrodes F3/F7 and F4/F8 (see Figure 29). We calculated FAA by following the steps below:

- Preprocess the data to attenuate artifacts (eye blink and muscle artifacts) by applying an order 5 Butterworth filter between 0.5 and 50Hz.
- Epoch the baseline-corrected data. In this step, the continuous data was broken into smaller parts of 2 seconds epoch (Tomarken, Davidson, & Henriques, 1990) (with 75% overlap). For each epoch, we compute the Alpha Power using the library PyEEG[36](a python module to extract EEG features). Using Fourier frequency analysis, the original signal is split up in frequencies in order to remove specific frequencies, before transforming back the signal with only the frequencies of interest.

[36] http://pyeeg.sourceforge.net/

- Average the Alpha power across all of the artifact-free epochs of the baseline and game scenes signal.
- Compute the normalized FAA as the log of the difference between alpha power density of the right hemisphere and the left hemisphere divided by their sum.

$$FAA = \log\left(\frac{Alpha\ Power_{Right} - Alpha\ Power_{Left}}{Alpha\ Power_{Right} + Alpha\ Power_{Left}}\right)$$

Higher scores on this asymmetry index indicate greater relative left hemisphere activation which means that the scenes' motivation is APPROACH oriented otherwise it is AVOIDANCE oriented.

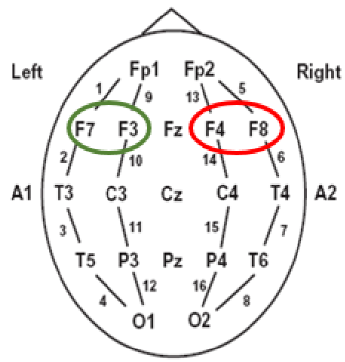


Figure 29. Used EEG sensors in the FAA calculation.

8.4.2 Player mastery goal assessment

To characterize the player category (Master or Performant), we identified the performance question in the 'Immersion-Game experience' survey (see Table 20). We categorized the player's achievement goal as Master if:

*Average(Mastery) >= 1.5*Average(Performance).*

We fixed the "1.5" threshold by checking information related to the preferred games, the play time per week and the GPA in the demographic self-report which is matching with the participants' report of 'immersion-Game experience' survey.

Table 20. Achievement goal related questions from the 'Immersion-Game experience' survey.

Achievement goal	**Questions**
Mastery	I would have liked to succeed in the game.
Mastery	I wanted to replay the game despite the failures.
Mastery	I feel satisfied when I see that the game is progressing.
Performance	Winning or losing the game did not interest me. I was not interested in the outcome of the game (win or lose).
Mastery	At the end of the session, I regretted not being able to continue the game.
Performance	The level of the challenge of the game corresponded to my skills as a player.
Mastery	I try to play as best as I can to win the game.
Performance	Sometimes I wanted to give up because the level of difficulty was too high.
Performance	The actions to be performed were becoming more and more difficult as I progressed through the game.
Mastery	The game was easy to win.
Performance	I think I made some progress at the end of the session compared to the beginning.
Mastery	The more I progressed in the game, the more I got interesting rewards (number of fans, bonuses, new weapons, new abilities, medals ...).
Performance	Sometimes I would have wanted to change the keys used to better control the game.
Mastery	I would like to replay this game.

8.4.3 OCC game scene representation

To predict the player motivation for new game scenes, we need first to characterize the game scenes according to the designer perception and goals using variables from OCC model (Andrew Ortony et al., 1990). The OCC model evaluates a situation with descriptive variables (global, central and local variables). These cognitive variables characterize a person's interpretation of a situation as desirable or undesirable, expected or unexpected, etc. Global variables are included in all situations, whereas the central and local variables are specific to certain situations characterizing their informational content.

Table 21. Global, central and local variables and their associated values.

	Global variables		Central variables			Local variables										
Evaluation Variable	Surprise	Sense of reality	Desirability	Approval	Attraction	Desirability	Esteem for other	Merit for other	Likelihood	Realization	Effort	Agent	Power of the	Deviation	Disposition	Familiarity
Values	0 (False) or 1 (True)	0 (False) or 1 (True)	-1, -0.5, 0, 0.5, 1	-1, -0.5, 0, 0.5, 1	-1, -0.5, 0, 0.5, 1	-1, -0.5, 0, 0.5, 1	-1, -0.5, 0, 0.5, 1	-1, -0.5, 0, 0.5, 1	-1, -0.5, 0, 0.5, 1	-1, -0.5, 0, 0.5, 1	-1, -0.5, 0, 0.5, 1	0 (other) or 1 (self)	0, 0.5, 1	-1, -0.5, 0, 0.5, 1	0, 0.5, 1	-1, -0.5, 0, 0.5, 1

In the table above (see Table 21), we summarized the OCC variables and their possible values. Thus, we can formally represent a game scene with vector of 16 numerical values. Thanks to its generality, The OCC model representation of the scene remains applicable to game scenes for learning or entertainment purposes.

8.4.4 The Motivation Prediction

In this section we describe our approach to training and evaluating classifiers for the task of detecting the motivational state of mind of a person given the person's cognitive situation in game, personal and demographical data. We approach this problem as a 4-class classification problem.

8.4.4.1 Features extraction

We perform a feature selection over the feature vector by extracting features using Principal Component Analysis (PCA) and Univariate feature selection (Univariate feature selection examines each feature individually to determine the strength of the relationship of the feature with the response variable). Then we combine the results through a pipeline into a single transformer. The 10-fold Cross validation method allowed to train and validate our model with better generalization.

8.4.4.2 Training set construction

In order to train a scene's motivation predictive model, we used a training set containing scene descriptions, participant's socio-demographic information and personality traits and also the participant's mastery goal and his approach related behavior in the game scene (Performant/Master-Approach/Avoidance). That we determined through the combination of the FAA computation section above and the categorization of the players we present above. The model use 23-dimensions vector: 16 variables from the OCC model, 2 socio-demographic variables (gender, age), 5 personality trait values as **input** variables and the motivational goal orientations as the **output result**. The method has been developed with Python language and scikit-learn[37] library. We have trained and validated our model using the 10-fold Cross validation method. We are interested in inducing a classifier of the following form:

- MotivationClassifier(*Playerdata*)→[*Performant-Approach, Performant-Avoidance, Master-Approach, Master-Avoidance*]

Where "*Playerdata*" is the 23-dimension vector presented above [*Performant-Approach, Performant-Avoidance, Master-Approach, Master-Avoidance*] is the sets of motivational states to be discriminated. The dataset contains 245 examples distributed over the 4 classes as follow Performant-Avoidance: 66/245 – 26%, Performant-Approach: 52/245 – 21%, Master-Avoidance: 71/245 – 28%, Master-Approach: 55/245 – 22%. Thus if a classifier always predict the most present class which is Master-Avoidance it will get 28% of precision that will be considered as our baseline.

8.5 Result

In this approach, we evaluated classifier by performing the standard 10-fold cross validation in which 10% of the training set is held out in turn as test data while the remaining 90% is used as training data. The optimum parameters for the classifier were found with a grid search.

[37] http://scikit-learn.org/stable/

Table 22. the classifiers F-score and parameters.

CLASSIFIER	F1-Score (parameters)
RFC	**81%** (PCA = 3, Univ=10, n_estim = 10)
SVM	**75%** (PCA=15, Univ=10, Kernel ='linear', C=1.0, gamma= 0.04)
KNN	**73%** (PCA=3, Univ=2, k=70)

Compared to KNN and SVM methods, Random Forest achieves the best performance with the average accuracy of 81% for all goal orientations classes, as illustrated in Table 22.

The confusion matrix of Random Forest is shown in Figure 30, which gives details of the strength and weakness of the generated model. Each row of the confusion matrix represents the target class and each column represents the predicted class. The element (i, j) is the percentage of samples in class i that is classified as class j. We can see that the goal orientations are generally recognized with very high accuracy of near ninety percent. The model differentiates between Mastery and Performance classes. But it makes some errors in recognizing whether it is Performance-Avoidance or Performance-Approach and also fewer errors between Mastery-Avoidance and Mastery-Approach.

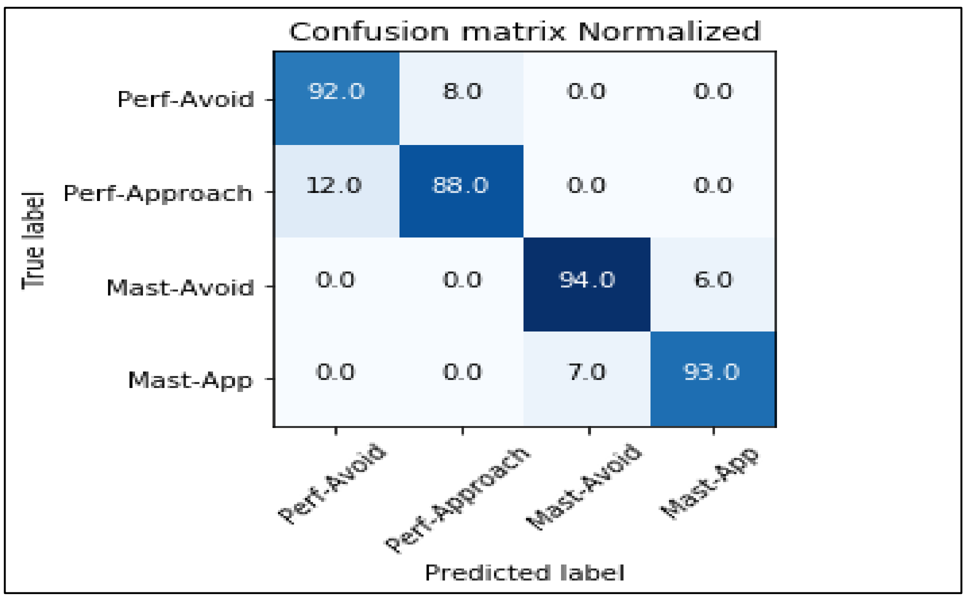


Figure 30. Confusion matrix of the motivational goal orientations prediction with RFC.

With these results, our motivation prediction approach provides a simple and reliable way to predict the motivational goal orientations of the learner/player that can be implemented in a distant game environment. The resulting model predicts the learner /player’s goal orientations from only the game scene’s OCC representation and the Big Five result without using invasive technique (physiological sensors, EEG, etc.).

Additionally, based on this assessment, we can make adaptations in the environment according to learner’s motivation. In fact, Kaplan and Maehr (Kaplan & Maehr, 2007) made comparison between aspects of educational environments (Task, Authority, Recognition, Grouping, Evaluation and Time) in mastery goals Vs performance goals emphasizing environments. For example, these environment aspects can be modified based on this assessment to foster learner’s motivation. Moreover, based on the model of motivated action theory (DeShon & Gillespie, 2005), Goal orientations describe a profile of structured hierarchies of goals that lead to the learner/player's specific action plan goals (Seek feedback, Manage impression, Allocate resources and Explore problem). Using this predicted player's actions, the adaptation strategies can modify the environment by providing hints, messages, feedbacks, rewards or changing the problem difficulty.

8.6 Conclusion

In this paper, we assessed players’ motivational goal orientations in their interaction with the commercial game “Outlast” using game scene's design goals, player characteristics, EEG and eye-tracking data. We presented our method in categorizing the player as "Master" or "Performant" using the GameFlow questionnaire. We also assessed the Approach withdrawal behavior toward a visualized game scene using the Frontal Alpha Asymmetry (FAA) during time window calculated from eye tracking data. We have also built a machine learning model for predicting player’s motivational goal orientation using game scene's design goal (defined by OCC variables) and the player’s personality traits (using the Big Five questionnaire).

The obtained results are very promising for their future integration in a motivationally intelligent serious game. This integration would clearly contribute to learning since it combines the game scene's design objectives to the learner/player's motivation. Furthermore, a practical real-time non-invasive assessment of learners’ motivation is now feasible, since we can rely on this assessment as a substitute for self-reports that can disturb a learning/gaming session. Moreover, the system can become more adaptive in terms of its response to learner’s motivation within the game scenes. In further work, we will target the learner's reported emotions in the prediction which would contribute to more comprehensive models of learner/player affect.

Acknowledgments. We thank the Natural Sciences and Engineering Research Council of Canada, NSERC, and BMU Games for funding this research.

Chapitre 9 : Évaluation affective et adaptation dans un environnement de réalité virtuelle

Nous décrivons dans ce chapitre notre dernière contribution qui s'intéresse à l'adaptation du contenu des jeux vidéo selon l'état émotionnel du joueur. Cette troisième et dernière contribution consiste à concevoir et réaliser un outil qui permet de définir et d'appliquer des règles d'adaptation aux éléments du jeu selon l'état émotionnel du joueur afin d'optimiser l'expérience utilisateur dans les environnements de jeu vidéo.

Ce chapitre vise donc l'adaptation des environnements de réalité virtuelle en fonction des données affectives de l'utilisateur. Dans cette étude, nous avons présenté un nouvel outil que nous avons développé, intitulé « BARGAIN » qui permet de créer des règles d'adaptation des éléments du jeu selon l'état émotionnel du joueur. Cet outil « BARGAIN » peut être facilement intégré à n'importe quel jeu Unity existant et le transformer en un jeu émotionnellement adaptatif. Nous avons appliqué cet outil dans un jeu de réalité virtuelle pour le développement socio-moral, nommé « Dilemmas_VR ». Composé de neuf dilemmes, l'environnement Dilemmas_VR met le joueur dans différentes situations. À chaque dilemme, le joueur est demandé à répondre à une question par rapport son opinion initiale par « oui » ou « non » et se justifier oralement pendant qu'il est enregistré. Ensuite, il rencontre cinq autres personnages non-joueurs qui donnent chacun son avis que le joueur doit évaluer en cliquant sur un bouton « like » ou « dislike ». Ce jeu a été développé par Sameh Chaieb dans le cadre de sa maitrise en art et communication en version écran. J'ai aussi participé avec lui pour transformer le jeu en version réalité virtuelle.

Dans ce travail, nous cherchons à répondre aux questions de recherche suivantes: (1) Pouvons-nous développer un outil général permettant de définir des règles d'adaptation du comportement des éléments du jeu en fonction de l'état émotionnel du joueur? (2) Pouvons-nous intégrer le système résultant dans un environnement de réalité virtuelle existant (dans Unity) afin de le rendre émotionnellement intelligent? Et (3) Pouvons-nous évaluer l'effet des règles d'adaptation dans cet environnement de réalité virtuelle émotionnellement intelligent sur l'expérience subjective du joueur ainsi que sur ses réactions émotionnelles?

Dans l'étude expérimentale, nous avons recruté 30 participants de l'Université de Montréal pour une période de 1 mois et une semaine. Chaque session d'un participant dure en moyenne 30 minutes. Chaque participant a été équipé d'un casque Emotiv (EEG), le casque VR (Oculus), une manette de jeu sans fil et un bracelet GSR pour détecter la conductance de peau. Pour la détection des émotions du joueur dans un environnement VR (portant un casque Oculus qui cache son visage), nous avons utilisé le système « NeuroExpress » pour la reconnaissance des expressions faciales à partir des signaux cérébraux (EEG), développé dans l'étude antérieure (Chapitre 6). Le système de reconnaissance des émotions est basé sur un modèle d'apprentissage machine des expressions faciales de l'utilisateur, comportant en entrée ses signaux physiologiques EEG. Cette méthode est applicable à tous les types de jeux mais plus adaptée aux jeux de réalité virtuelle, où les expressions faciales ne sont pas accessibles en raison de l'utilisation du casque de réalité virtuelle, qui masque le visage de l'utilisateur. De plus, l'utilisation de l'EEG donne plus d'informations sur les mesures cognitives au cours de l'interaction.

Dans cette étude nous avons présenté les analyses statistiques sur les données d'expérimentation afin de détecter les éventuelles corrélations entre les mesures affectives et cognitives, les événements de jeu et l'expérience de l'utilisateur (rapportée par le questionnaire GEQ), ainsi que tout effet possible des règles émotionnelles sur les mesures affectives et cognitives du joueur. Dans les résultats, cette étude montre que les mesures affectives et cognitives peuvent être un indicateur de l'expérience de jeu, comme le montre la corrélation avec les rapports subjectifs et les événements du jeu. De plus, l'effet des règles d'adaptation ne peut être statistiquement significatif que pour les mesures affectives et pas significatif pour les mesures cognitives. Ceci peux être dût au fait que les règles d'adaptation ont seulement été appliquées qu'à la musique de fond du jeu et aux animations faciales et gestuelles des avatars (suscitant plus le plan affectif et ne nécessitant pas beaucoup d'effort mental) et non à la difficulté du jeu et des taches du joueur. Ceci suggère que les mesures cognitives sont meilleures pour analyser les événements du jeu qui nécessitent un effort mental (comme le montrent les corrélations avec les dimensions du GEQ).

Le reste de ce chapitre est constitué de l'article intitulé « BARGAIN: *Behavioral Affective Rule-based Games Adaptation Interface. Towards Emotional Intelligent Games:*

application on a virtual reality environment for socio-moral development » soumis (le 21 novembre 2018) au journal *User Modeling and User-Adapted Interaction*, UMUAI. Nous rappelons que ma contribution essentielle consiste au développement de Framework BARGAIN et l'integrer dans le jeu Dilema_VR, à la conception de l'expérimentation, à la collecte des données, à l'analyse des données et l'interprétation des résultats et à la rédaction du papier.

BARGAIN: <u>B</u>ehavioral <u>A</u>ffective <u>R</u>ule-based <u>G</u>ames <u>A</u>daptation <u>In</u>terface

Towards Emotional Intelligent Games: application on a virtual reality environment for socio-moral development

Benlamine, M.S., Dufresne, A., Beauchamp, M., & Frasson, C. (under review 2018). BARGAIN: Behavioral Affective Rule-based Games Adaptation Interface Towards Emotional Intelligent Games: application on a virtual reality environment for socio-moral development. Submitted to the journal *User Modeling and User-Adapted Interaction* ***UMUAI*** at the 21st November 2018, Manuscript number: UMUI-D-18-00119.

Abstract. This paper presents a framework for adapting game elements to the player's affective state and the integration of the framework in a virtual reality environment for moral development. These game elements include gestural and facial expressions of avatars during dialogues with the player, background music, the score, game mechanics, aesthetics and learning. The framework BARGAIN (Behavioral Affective Rule-based Games Adaptation Interface) is an authoring tool for affective game design providing a visual interface based on finite state ma-chine (FSM) technique to represent the affective rules as state transitions graph dependent on the player emotional state assessed using facial expression recognition system based on electroencephalography (EEG) data. We conducted a user study (n = 30) examining the effects of the resulting affective virtual reality game on players' experience using the Game experience Questionnaire (GEQ) (IJsselsteijn, De Kort, & Poels, 2013). The results show significant correlation between the GEQ dimensions and the player's facial expressions during his interaction with the Non Player Characters (NPCs) within the VR game. These findings highlight that adapting games to user's emotions enhances the players' experience.

9.1 Introduction

There is a long-standing interest in adapting virtual environment to the user's affective state, such as affective computing (Picard, 1997), physiological computing (Fairclough, 2010), human-computer interaction and video game fields, with regard to both theoretical and practical issues. To make video games responsive of the player emotional reactions, it is necessary to adapt the design and the structures of video games (Böckle et al., 2018), and also the game elements involved in interactions with the players. New tendencies, especially in game industry, target the user experience by becoming more aware of the player's emotions (Bontchev, 2016). Therefore, it is necessary to develop tools that takes in consideration the player's emotions within the design of video games.

Existent studies (Arellano, Tokarchuk, & Gunes, 2016; Christy & Kuncheva, 2018) focused on specific physiological signals and measures; or presented general principles (de Byl, 2015; i Badia et al., 2018) and rules of affective feedback frameworks (L. E. Nacke et al., 2011; Nogueira et al., 2013); or realized specific game prototype and its mechanics (Abdessalem & Frasson, 2017; Kosunen et al., 2016) without providing a reusable authoring tool that can be easily used to change an existent normal game to emotional intelligent game. Such tool will give more easiness to game designers and game developers in making such games with efficient development process in term of time and resources.

In this work, we aim to answer the following research questions: **(1)** Can we develop a general framework that allows the definition of adaptive rules to modify the behavior of game elements depending on the player's emotional state? **(2)** Can we integrate the resulting system in an existing virtual reality environment (in Unity) to transform it into an emotionally intelligent virtual reality environment that automatically adapt to the user's emotion? And **(3)** Can we assess the effect of this emotionally intelligent virtual reality environment on the player's subjective experience and also on his emotional reactions? To answer these questions, we developed "BARGAIN" (Behavioral Affective Rule-based Games Adaptation Interface) an authoring tool that can be integrated in a unity game, that offer the possibility to receive the affective information and to create emotional rules that adapt the game elements depending on the user's emotional state.

This paper describes the resulting system which is an authoring tool for creating adaptive rules embeddable in any Unity game or interactive system as a package that allows:

- The reception of the emotional values from network asynchronously and the identification of the current emotional state of the player;
- The creation of the emotional rules to adapt the game objects' behavior to the user's emotional state.

The BARGAIN system was used to integrate emotionally intelligent adaptation in the game 'Dilemmas_VR', a virtual reality game where the player is exposed to moral situations where he makes a choice and gives a justification orally. Then, the player meets five avatars that give their opinions about the dilemma and chooses whether to agree or not with them. The player gets a social score according to his decisions in the game. During the game session, the participant was equipped with virtual reality headset (Oculus[38]) on his eyes and EEG (encephalography) headset (Emotive EPOC[39]) to record his brain activity. This prototype is based on new technology that capture the user's facial expression via the EEG (M. Benlamine et al., 2016) headset (Brain physiological signal).

The evaluation of player experiences in game design is very important and especially for VR games. Player experience is linked to concepts like flow, challenge and immersion (Bernhaupt, Eckschlager, & Tscheligi, 2007). The player's interactions with game elements (for example: NPC's gestures and facial expressions, Background music, lights ...) influence his affect and cognition. It is therefore important in game design to be able to measure the effect of the game elements on player's experiences. This paper presents facial expressions analysis for evaluating player experiences by investigating correlations between game events, affective and cognitive measures and self-reported Game Experience Questionnaire (GEQ) dimensions (IJsselsteijn et al., 2013). The paper also presents the analyses of the impact of the different adaptive rules on the emotional and cognitive reactions of users in the different contexts of the game.

[38] www.oculus.com

[39] www.emotive.com

9.2 Experimental settings

9.2.1 The game – Dilemmas_VR

In this study, participants were asked to play "Dilemmas_VR" a virtual reality serious game for socio-moral reasoning that was developed in our laboratory (See Figure 31).

Figure 31. Some scenes from the Dilemmas_VR environment

The game presents nine everyday life dilemmas where the player was asked to take a decision and provide a justification orally to be recorded. Then the game asks players to meet the other characters to know what their decisions would be and why. For each situation, five characters represent consecutive developmental levels of social and moral maturity (Chiasson et al., 2017; Gibbs, 2013). After meeting a character, players are asked to evaluate the character's opinion and to express their agreement or disagreement by selecting a Like or a Dislike button (see Figure 32). The NPC then shows a non-verbal reaction to the player's evaluation with facial and gestural animations. A first version of this game (Tato et al., 2018; Tato et al., 2017), was developed for normal screen monitors and was tested with school children (aged under 18).

Figure 32. Examples of NPC's non-verbal reactions (facial and gestural)

9.2.2 Participants

The experiment involved 30 participants (16 males; 14 female), aged between 20 and 35 years from a North American university. We have discarded 1 participant due to technical problem while collecting data.

9.2.3 Experiment and equipment

The Experiment begins by receiving the participant in our laboratory; we introduce him to the testing room. After signing a consent form, Participants answered an initial questionnaire about their education, type of played games and hour of play per week and also the big-five questionnaire about their personality traits. After presenting the game and how to use the joystick buttons in the game, the experimenter installs the Emotiv EEG headset on the participant's head and launches the NeuroExpress application (developed previously (M. Benlamine et al., 2016; M. S. Benlamine et al., 2016)) that detects the player's facial expressions only from EEG. NeuroExpress sends the user's affective data to the VR game that adapts its game elements to the user's emotional state. The experimenter installs also the Oculus VR headset on the player (see Figure 33) and let him play the game (that takes about 30 minutes).

After playing the VR game, participants filled the Game Experience Questionnaire (GEQ) (IJsselsteijn et al., 2013).

Figure 33. A participant playing the Dilemmas_VR game with experiment settings.

9.2.4 Measures

In order to evaluate the player's affective and cognitive state, subjective and objective measures have been used.

9.2.4.1 Subjective measures

In order to get the users' information and their feedback about their interaction with the game, we used questionnaires before and after the play session. We first gave pre-test questionnaires to collect socio-demographic data and the player's profile (school level, their preferred games and hours of play per week) and also the "Big Five" questionnaire (Goldberg, 1992) for the assessment of the participant's personality traits. In the post-test, we used the Game Experience questionnaire (GEQ) (IJsselsteijn et al., 2013). These questionnaire is composed of three modules: the core questionnaire, the social presence module and the post-game module. The GEQ assesses player's game experience dimensions with good reliability (Ahmad et al., 2017; L. Nacke & Lindley, 2008) : Immersion, tension, competence, flow, negative affect, positive affect and challenge. In addition, participants were asked to answer an extra question related to the opinion change felt after meeting the characters and knowing their opinions within the game.

9.2.4.2 Objective measures

In a virtual reality environment, monitoring and analyzing objective measures like the player's physiological data is very important to analyze the user's behavior. As the user's face is hidden because he's wearing the Oculus VR-headset, we have not the access to his facial expression data using a camera. For that reason, we have developed, in a previous study (M. Benlamine et al., 2016), NeuroExpress (Figure 34), a real-time application for the facial expression recognition from EEG data. In this study, besides the affective measures (facial expression) that we get from NeuroExpress, we have extended the NeuroExpress algorithm to provide additional cognitive measures computed from the EEG input data. The additional cognitive measures are: The *short-term* and *long-term Engagement*, the *Frontal Alpha Asymmetry* (FAA), and the *Attention* index.

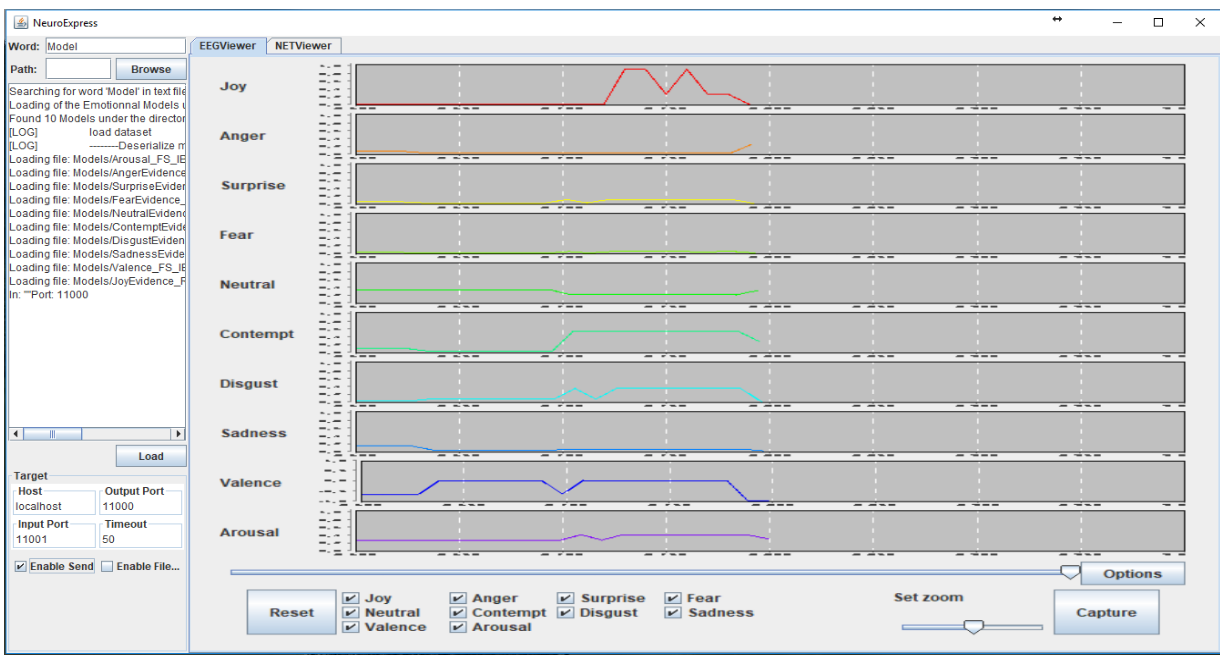


Figure 34. The interface of NeuroExpress application.

Facial Expressions via EEG.

Using the AI models, NeuroExpress recognize facial expressions data from the EEG cerebral signals with a precision reaching 92% compared to the camera based Facial expression recognition software iMotions Facet[40] (M. Benlamine et al., 2016). The tool NeuroExpress visualizes in real time the intensity of each facial expression between [0..1] as curves over time (6 outputs/sec). We have 10 facial expressions outputs classified in 6 basic emotions (Joy, Anger, Surprise, Fear, Disgust, and Sadness), 1 secondary emotion (Contempt), Valence, Arousal and Neutral emotion.

Frontal Alpha Asymmetry.

The special effect of frontal alpha asymmetry (FAA) was initially detected in studies investigating biomarkers of personality (Hagemann et al., 2002). The frontal asymmetry index was computed from raw frontal EEG data using electrodes F3/F7 and F4/F8.We calculated FAA using the formula below:

$$FAA = \log\left(\frac{Alpha\ Power_{Right} - Alpha\ Power_{Left}}{Alpha\ Power_{Right} + Alpha\ Power_{Left}}\right)$$

Recent studies (Harmon-Jones et al., 2010; Harmon-Jones & Gable, 2018) suggests that FAA varies depending on emotional stimulation, reflecting whether or not someone is drawn towards or away from something or someone, this “approach/avoidance effect” reflects someone’s motivation. Higher scores on this asymmetry index indicate greater relative left hemisphere activation which means that the player’s behavior in the scene is APPROACH otherwise it is AVOIDANCE.

Engagement measure.

The engagement index (Chaouachi et al., 2010a; Chaouachi & Frasson, 2012b; Chaouachi, Jraidi, & Frasson, 2015b; Pope, Bogart, & Bartolome, 1995b) is computed from three EEG frequency bands: Alpha (8-12hz), Beta (12-22Hz) and Theta (4-8hz):

Engagement = Beta / (Alpha + Theta).

Long-Term Engagement measure.

The long-term engagement (LT-Eng) is calculated as the 3rd quarter on a 10 sec of the engagement data.

40 https://imotions.com/emotient/

Distraction measure.

The EEG index of Distraction = Theta / Beta (Putman et al., 2010), is negatively correlated with the level of attention of learners. So, a low value reflects a normal attention state while a high value of this index reflects an excessive Theta rhythms, characterizing a state of inattention (Distraction).

9.3 BARGAIN System Architecture

We developed the BARGAIN authoring tool to design affective game using the technique of finite state machine (FSM) to represent the affective rules as a state transitions graph sensitive to the player emotional state. As shown in Figure 35, the system integrates NeuroExpress measures, calculates the User's emotional state every 30 sec (Bekinschtein et al., 2004) and offer the possibility to define rules to adapt the game elements depending on user emotional state. By simply double-clicking on it, this tool can be imported into the current open Unity project, a folder with name BARGAIN is created in the project hierarchy containing all the necessary scripts and prefab. This framework allows the reception of the NeuroExpress measurements, identifying the user's emotional state and also the creation and the integration of the emotional rules in the game elements. This tool consists of two main components Biometrics and BARGAIN rule editor.

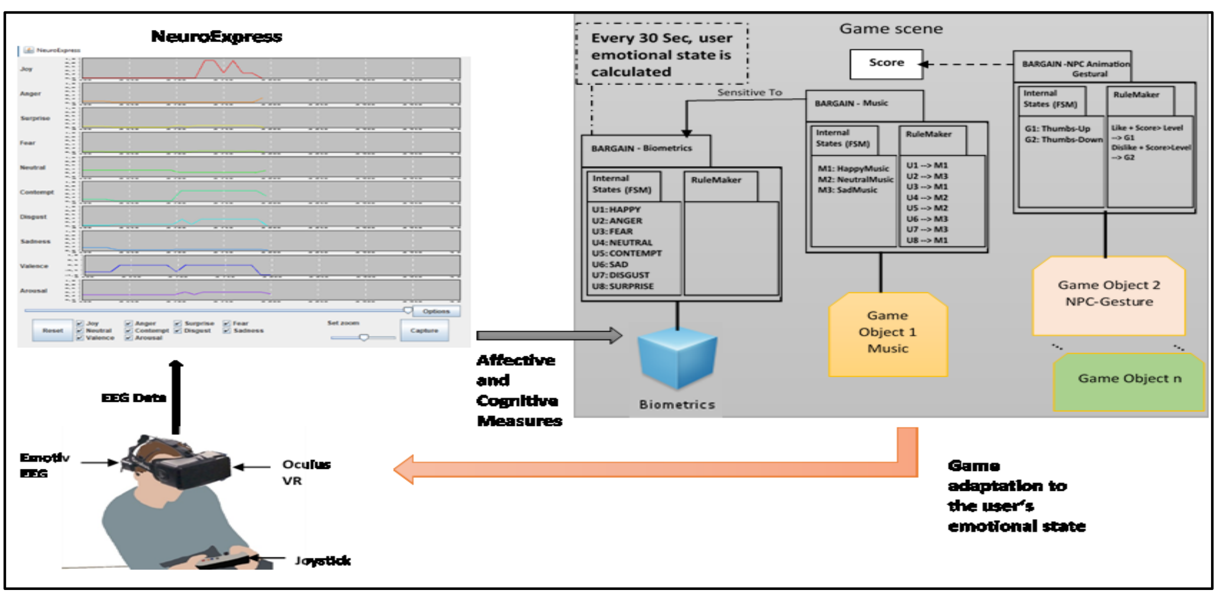


Figure 35. The system architecture of the BARGAIN framework.

9.3.1 Biometrics

Biometrics (see Figure 35) is an important component of the framework, it is responsible for establishing the UDP connection with NeuroExpress to get the affective measures. This component is also in charge of identifying the user's emotional state by making the average values of the affective measures every 30 sec and select the state with the maximum average value. Biometrics has another important function which is the log management where it records the game event's name along with the time stamp, the 30 seconds average measures and the average measures of the current second. For easiness sack, we created a Biometrics prefab included in the package, so that the developer has only to drag and drop it in the hierarchy and make some configurations (ip-address, port number, initial user's state …).

9.3.2 BARGAIN Rule editor

Now with the Biometric component, the User's emotional state is available within the game. But the game elements reaction to the user's emotional state has to be specified by the game designer. So we implemented a graphical Interface (see Figure 35) to allow the game designer to create the different states that can have a game element, the transitions between the states and the emotional rules that according to the current user state, activate a game object state. For example, a change in the user's emotional state (Anger) can trigger a change in the type of music (Sad music), as in Figure 35.

The BARGAIN interface is composed of two views: the work view (Left) and the property view (Right). The work view is where the designer can create the states and the transitions between the states as a finite state machine. The property view contains the "resource directory" field where the generated scripts of the states will be saved, the "target" field that contains the associated game object and the rules interface (framed in red). The rules interface allows the creation of the emotional rules. By dragging the Biometrics object into the "sensitive to" field, the game element becomes sensitive to the Biometrics object. In The panel below, we can add an emotional rule by clicking the plus button. Therefore, we can specify for each Biometrics state which internal transition in the game element will be activated, and the game element enters to the corresponding state. We can also control another game object by dragging it in the "transmitting to" field. So we can add new rules by clicking the plus button in the panel below and specify for each state of the game element which transition to activate in the distant game element. For example, as seen in Figure 36, the Music object changes the game's background music depending on the current user's emotional state calculated in the biometrics object according to the defined rules in the rule interface of the music object.

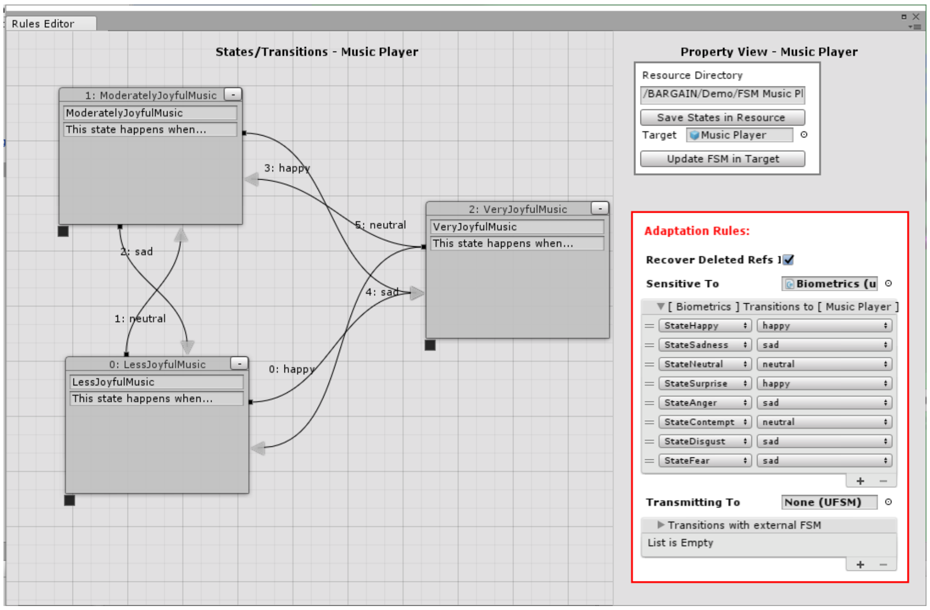


Figure 36. The BARGAIN interface containing the game object's states/transitions (Left) and the adaptive emotional rules (Bottom-Right).

By clicking on a state, the state's public variables are shown on the property view. The variables can be tweaked and their effect can be seen when running the scene. With BARGAIN interface, the active state can be tracked when running the scene, which is a useful feature to verify the execution of rules within the running game (as shown in Figure 37).

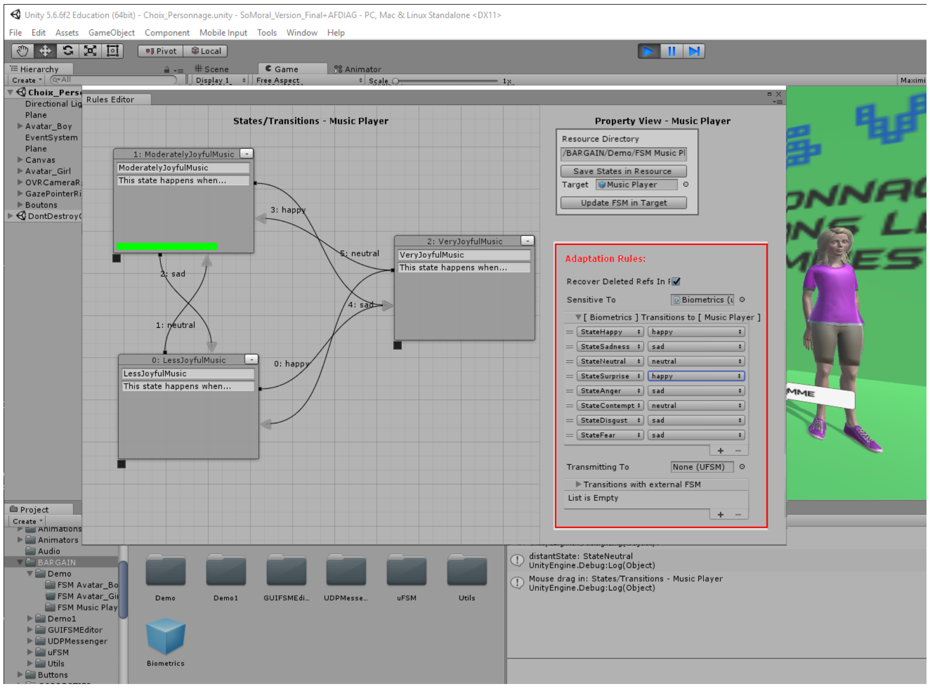


Figure 37. The “BARGAIN” interface when running the scene in Unity, the active state is with a running green bar.

By double clicking on a state in the work view, we can access to the associated script in visual studio. A state is derived from MonoBehavior standard Class, so we have access to the target objects’ variables (Ex. Score variable in Figure 35) and to all standard functions (Ex. Awake(), Update() and Start()) with two additional functions:

- OnEnter() : The called function when state is entered.
- OnExit() : This function is called when the state is exited.

The BARGAIN interface can be accessed in the unity software, but it is not compiled when generating the game. The BARGAIN interface updates two components attached to the game element (uFSM and RuleMaker) that will be integrated during the game compilation.

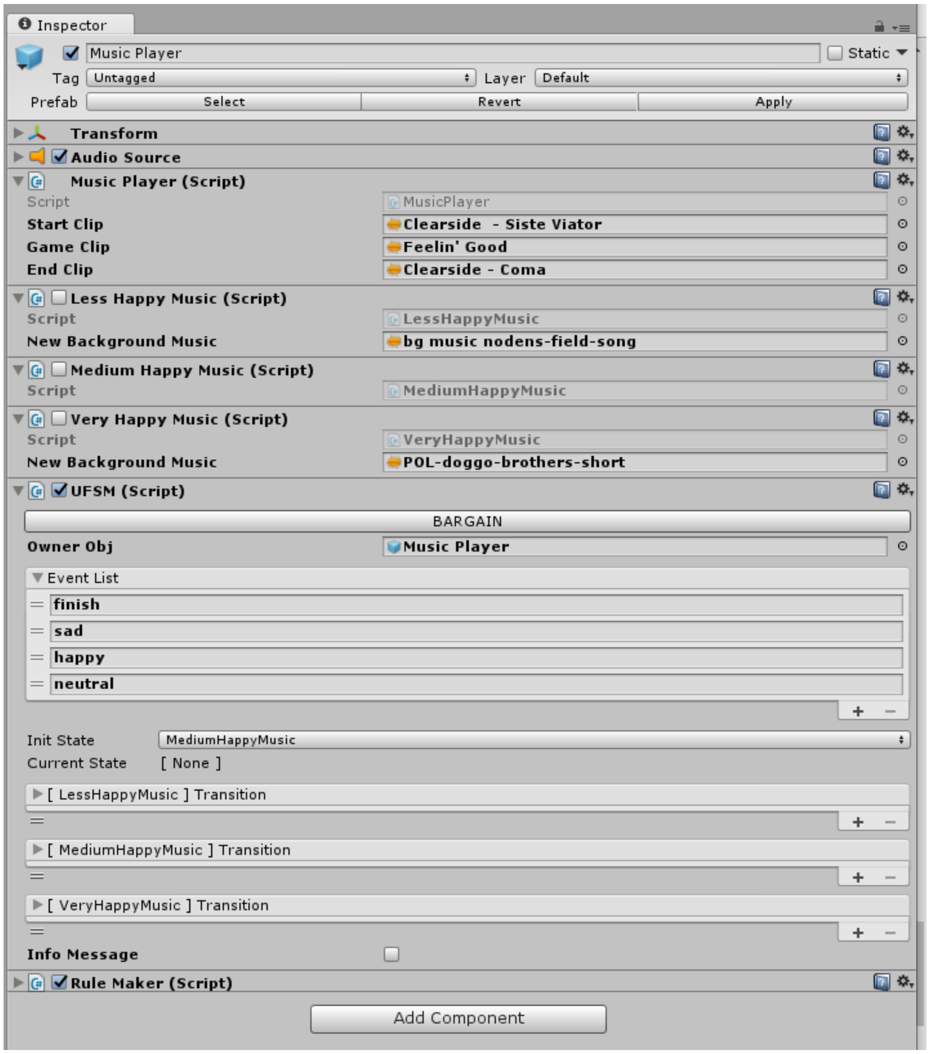


Figure 38. The music-player properties in the inspector within the Unity software with the uFSM and the RuleMaker attached components

This two components can be seen in the inspector attached to the game object (See Figure 38):

- uFSM: The component that manages the state machine. This component is updated in the BARGAIN interface in the work view by clicking "update target" button.
- RuleMaker: The component that manage the rules. This component is updated in the BARGAIN interface when modifying the rules interface.

Even if the game designer does not have the ability to modify the state code, he can define the states with a description (in the field within the state) and create the emotional rules and test the system even if the new states are not implemented yet. So, later the programmer could implement the code of the states.

9.3.3 Emotional adaptation rules

Using BARGAIN framework, we have created emotional adaptation rules to make the Dilemmas_VR game emotionally intelligent. In fact, we have mainly 4 emotional adaptation rules as follow:

9.3.3.1 Mimetic rules.

The Non-Player Character (NPC) mimes the detected player's emotional state to maintain the player's emotion and mirror his feeling as an empathic action trying to build a social relation with him (see Figure 39). We set the mimetic emotional rules as follow:

1) If the detected player's emotional state is Happy or Surprise (during the last 30 seconds) then the NPC will display a Happy Face.
2) If the detected player's emotional state is Neutral, Contempt or Fear (during the last 30 seconds) then the NPC will display a Neutral Face.
3) If the detected player's emotional state is Disgust or Anger (during the last 30 seconds) then the NPC will display a Disgust Face.

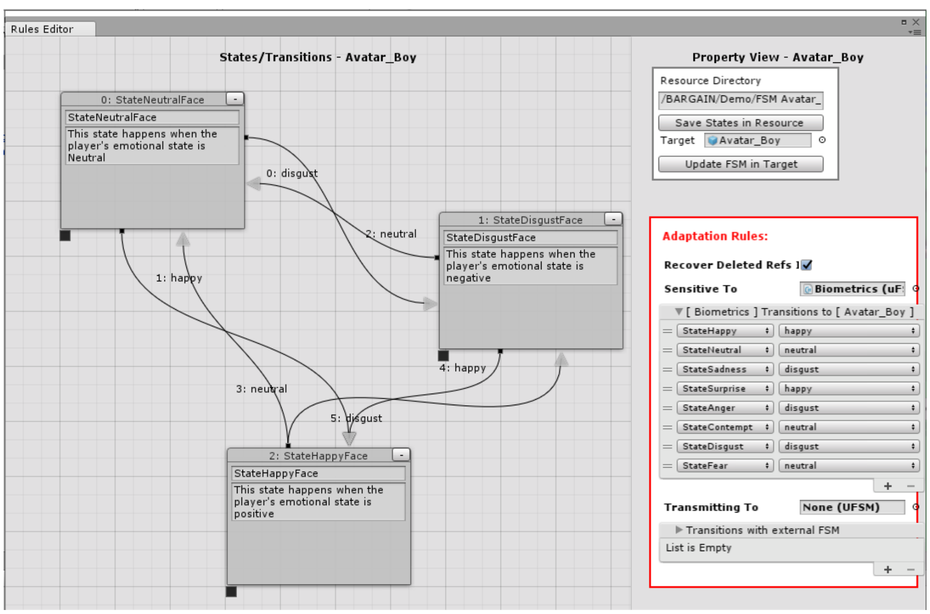


Figure 39. NPC's facial states and the emotional rules interface (framed in red) in the BARGAIN interface.

9.3.3.2 NPC`s Facial non-verbal reaction rules.

The Non-Playable Character (NPC) reacts to the player's evaluation with a facial animation, as follow:

1) If the player selected the Like button, the NPC will show a Happy face animation.
2) Otherwise, if the player selected the Dislike button, NPC will show a Disgust face animation.

9.3.3.3 NPC's Gestural non-verbal reaction rules.

The Non-Playable Character (NPC) reacts to the player's evaluation with a gestural animations with a consideration to the player's social score, as follow:

1) If the player selected the Like button and his social score is less to the NPC social level, the NPC will make a thumbs-up animation.
2) Elsewise, if the player selected the Dislike button and his social score is less to the NPC social level, the NPC will make a thumbs-down animation.

This gives a hint to the player in the evaluation and push him to develop his socio-moral score.

9.3.3.4 Music-player rules.

Using this rules, the game's background music is aware of the player's emotional state (see Figure 37). In fact, depending on the detected player's emotional state the background music changes to maintain the player's emotion and reflect his feelings within the game. We set the rules of the emotional aware music-player as follow:

1) If the detected player's emotional state is Happy, Surprise or Fear (during the last 30 seconds) then a Very_Joyful music[41] is played.
2) If the detected player's emotional state is Neutral, Contempt (during the last 30 seconds) then a Moderately_Joyful music[42] is played.
3) If the detected player's emotional state is Disgust, Anger or Sad (during the last 30 seconds) then a Less_Joyful music[43] is played.

---

[41] Very_Joyful Music: https://www.playonloop.com/2017-music-loops/doggo-brothers/

[42] Moderately_Joyful Music: www.audiomicro.com/feelin-good-wav-royalty-free-stock-music-1438027

[43] Less_Joyful Music: https://freesound.org/people/axtoncrolley/sounds/172707/

9.4 Evaluation and results

We implemented a multimodal analysis using log file information to determine the affective reactions of the players to game events. We used several sources: games events, player's reaction measures and subjective assessment in the questionnaires. We first analyzed the relation of the affective and cognitive measures during the game and the assessment of the player in the assessment of the player in the subjective measures after the game. We then assessed in the context the impact of the application of the emotional rules on the following player's affective and cognitive measures.

9.4.1 Correlation Analysis between emotional and cognitive measures and the subjective evaluation for different game events (context)

With the framework BARGAIN, we have created the emotional adaptation rules having as condition the player emotional state calculated every 30 seconds as the maximum of the average of the facial expressions' values received from NeuroExpress (Figure 33). Using SPSS, we calculated Pearson's correlation between game experience dimension in the GEQ, game events and average facial expression intensities, as shown in Table 23. We recorded affective (the six basic emotions plus neutral) and cognitive (engagement, long-term-engagement, distraction and frontal Alpha Asymmetry (FAA) indexes) measures. These measures was calculated as the average of 1 second after each of these five game events: Player Gives Initial Opinion, NPC Starts Talk, NPC Ends Talk, Player Evaluates NPC, NPC Makes Non Verbal Reaction.

Table 23. Pearson correlation between GEQ dimensions and Player's affective and cognitive measures in the game events (context).

GEQ\Events	Initial Opinion	NPC Starts Talk	NPC Ends Talk	NPC Evaluated	NPC Reaction
Competence	**FAA**=-.403*	**FAA**=-.443*	**FAA**=-.443* Valence=.379*	**FAA**= -.440* Happy=-.371* **Sad**=.496**	**FAA**=-.417* **Sad**=.394*
Immersion	Sad=-.384*	Neutral=.386*	Valence=.391*	Disgust =-.479** Valence=.407*	**Neutral**= .573**
Flow		**Distraction**= -.416*	**Distraction** = -.415*	**Distraction** = -.406*	**Distraction** = -.402*
Tension	**Happy**= -.463*		Contempt=.382*		Surprise= .383*
Challenge	**Contempt**= .584**		**Contempt**=.414* **Fear**= .491**		
Negative affect	**Happy**= -.444* **Neutral**= .436*				
Positive affect	**Happy**=.619**			**Disgust**=-.425*	
Negative Experience	**Distraction** = -.422*				
Tiredness		**Contempt**= -.466* **Valence**=-.539**	**Contempt**= -.455*	**Contempt**= -.498** Arousal=-.436*	**Contempt**= -.521**
Behavioural Involvement	**Sad**= -.528**				
Persuasion (opinion change)	LT-Eng =.373* Engagement= .386*	LT-Eng=.375* **Disgust**=-.531** Fear=.376* Sad=-.427*	LT-Eng =.372* **Disgust** = -.453* Anger= -.416*	LT-Eng=.371* **Disgust**=-.527**	LT-Eng= .370* **Disgust** =- .595**
Social Score				Surprise=.368*	

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).
Bold. Confirmed with Bayesian factor analysis with at least moderate level of evidence.

The table 23 presents the correlations between GEQ dimensions and Player's affective measures in the five game events that occur in each level (Player Gives Initial Opinion, NPC Starts Talk, NPC Ends Talk, Player Evaluates NPC, and NPC Makes Non-Verbal Reaction). For each situation, we look at the facial expressions of the player (for a second) and their correlations with the values of the questionnaire at the end (we look at the significant correlations).

This table (Table 23) shows us the correlations between the GEQ scores and the affective measures in each game events, which give us an idea about the relation between the GEQ dimension and the game context more precisely. In other word, we can identify the game event's triggered emotions that correlates with GEQ dimension's score.

Nevertheless, inferences from these correlations can only be drawn with caution, given they would not have survived multiple comparison corrections and were only carried out as post-hoc exploratory analyses. So we carried out a Bayesian correlation pairs analysis with a conservative beta prior of 0.5 as confirmatory analysis, to determine the level of evidence for the null hypothesis (H0) or the alternative hypothesis (H1) (correlation between GEQ score and the emotional metric in a game event).

Competence: The GEQ dimension of Competence is correlated (negatively) with the Frontal Alpha Asymmetry (FAA) in all game events with moderate to large effect sizes ($-0.65 < r < -0.30$, $p < 0.05$) (Table 23). This is confirmed by Bayes correlations of GEQ Competence score with the FAA metric in all the game events with anecdotal to moderate levels of evidence. For example, for the NPC ends talking event the Bayes Factor for the correlation hypothesis (H1) was favoured 4.448 times over the null hypothesis (H0) between Competence score and FAA, with moderate level of evidence (Figure 40).

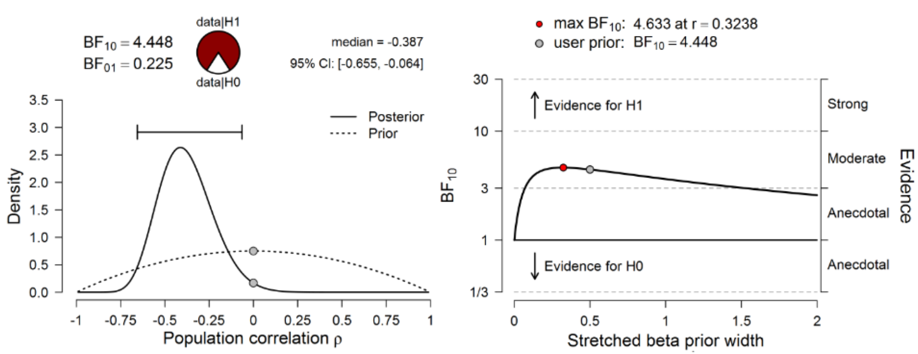


Figure 40. Results from Bayesian correlation pairs analyses for GEQ Competence score and FAA (Frontal Alpha Asymmetry) in the NPC Ends Talk event. CI, Credibility Interval; BF, Bayes Factor.

This can be explained by the fact that the more competent you feel the more you are in avoidance mode, as the FAA is a measure related to approach/avoidance motivation. We note also, from Table 23, that the competence score is correlated (positively) with Valence measure when NPC finished talking. But this suggests small trends as the Bayes Factor analysis shows only anecdotal evidence for the correlation hypothesis (H1) over the null hypothesis (H0) between Competence score and Valence metric (Figure 41).

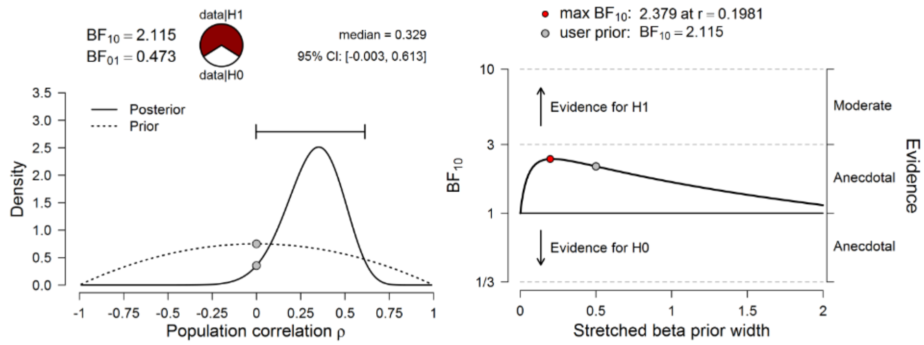


Figure 41. Results from Bayesian correlation pairs analyses for GEQ Competence score and Valence metric in the NPC start Talk event. CI, Credibility Interval; BF, Bayes Factor.

Furthermore, when the player evaluates the NPC, Competence GEQ dimension was significantly positively correlated with Sadness ($r=.496$, $p < 0.01$). This is confirmed by Bayes Factor for the correlation hypothesis (H1) with 9.548 times over the null hypothesis (H0) between GEQ Competence score and the Sadness metric in the NPC ends talking event, with moderate (close to strong) level of evidence (Figure 42).

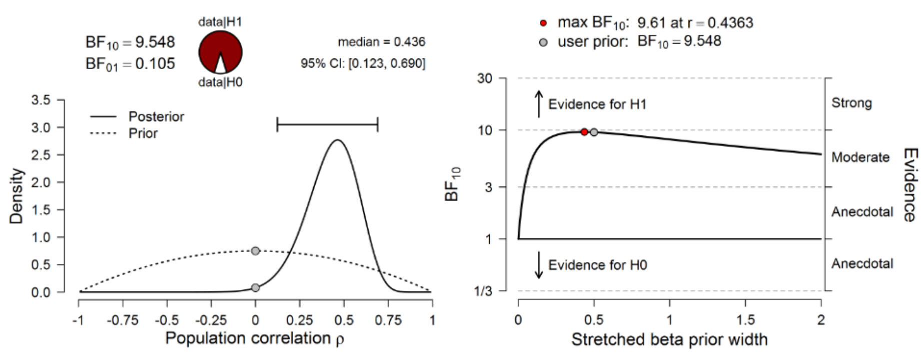


Figure 42. Results from Bayesian correlation pairs analyses for GEQ Competence score and Sadness metric in the Player evalutes NPC event. CI, Credibility Interval; BF, Bayes Factor.

There is also, in the same game event, a significant negatively correlation between Competence GEQ score and Happy metric ($r=-.371$, $p <0.05$). But again the Bayes Factor Analysis find an anecdotal level of evidence for the correlation hypothesis (H1) over the null hypothesis (H0) showing small trends to correlation between Competence score and the Happy metric in the NPC ends talking. This suggest that, at the moment of the evaluation using Like/Dislike button, the participant feels more competent if he is more sad and tending to be less happy about what have been said (the participant didn’t like the NPC's opinion).

Immersion: For each game event, the Immersion dimension has different affective measures correlations. In fact, in the Initial Opinion phase the Immersion is more correlated with Sadness ($r=.414$, $p<0.05$). Also, when the NPC ends talking, the Immersion score is significantly correlated (negatively) with Disgust ($r=.479$, $p<0.01$) and also with Valence ($r=.391$, $p<0.05$). But with Bayesian Factor analysis, we get only anecdotal evidences to correlation hypotheses for all of these measures with Immersion score. Thus, we cannot fully confirm these correlations result.

Moreover, there is significant correlation between Immersion score and Neutral ($r=.573$, $p<0.01$) in the NPC non-verbal reaction event. This is confirmed by Bayes correlations of Immersion score with Neutral metric in the NPC non-verbal reaction event with very strong level of evidence for hypothesis (H1) 36.819 times over null hypothesis (H0) (Figure 43).

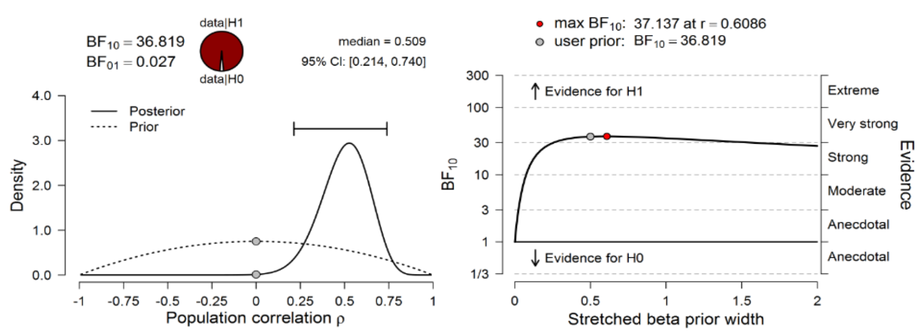


Figure 43. Results from Bayesian correlation pairs analyses for GEQ Immersion score and Neutral metric in NPC non-verbal reaction event. CI, Credibility Interval; BF, Bayes Factor.

Flow: According to Table 23, the GEQ dimension of Flow is negatively correlated with the Distraction index in all game events except the initial opinion event with moderate to large effect sizes ($-0.65 < r < -0.30$, $p < 0.05$). The Bayes factor analysis reaches moderate levels of evidence for the correlation hypothesis (H1) of Flow score with Distraction metric in all game events with except initial opinion event. For example, for the NPC starts talking event, Bayes Factor analysis favoured by 3.169 folds the alternative hypothesis (H1) over the null hypothesis (H0), with moderate level of evidence (Figure 44). This suggests that the more in flow the participants are, the less they are distracted.

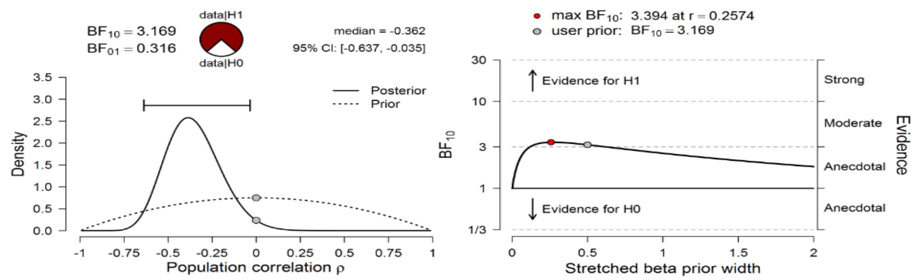


Figure 44. Results from Bayesian correlation pairs analyses for GEQ Flow score and Distraction metric in the NPC starts talking event. CI, Credibility Interval; BF, Bayes Factor.

Tension: The tension GEQ dimension is correlated (negatively) with Happy measure in the Initial Opinion phase ($r=-.463$, $p<0.05$). This is shown by Bayes Factor for the correlation hypothesis (H1) with 5.855 times over the null hypothesis (H0) between GEQ Tension score and the Happy metric in the Initial Opinion event, with moderate level of evidence (Figure 45).

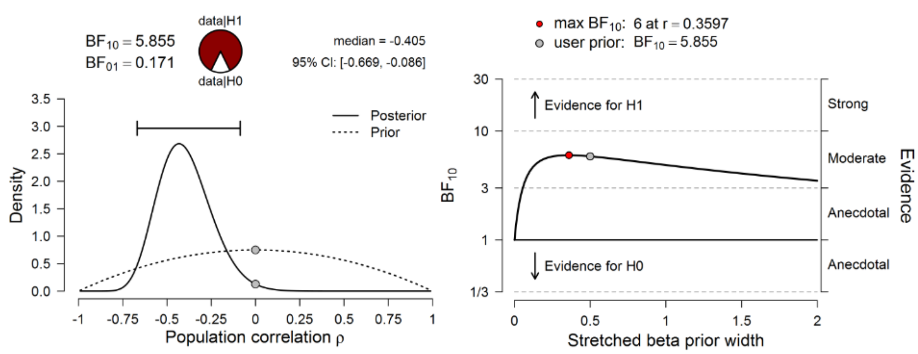


Figure 45. Results from Bayesian correlation pairs analyses for GEQ Tension score and Happy metric in the Initial opinion event. CI, Credibility Interval; BF, Bayes Factor.

This shows the player's perceived tension in taking decision about the dilemma and also providing justification about his decision (orally). Moreover, there is significant correlation between Tension and Contempt measure ($r=.382$, $p<0.05$) when the NPC finished his talking and also between Tension and surprise measure ($r=.383$, $p<0.05$) when the NPC showed non-verbal reaction (Table 23). But the Bayes Factor analysis shows small trends with only anecdotal evidences for the correlation hypotheses (H1) over the null hypothesis (H0). This suggests some trends that the participants feel some tension to the evaluation of the NPC's opinion that correlates with Contempt measure and also with Surprise measure when the NPC shows non-verbal reaction (that can be unexpected to the participant).

Challenge: The Challenge GEQ dimension is significantly correlated (positively) with Contempt measure in the Initial Opinion phase ($r=.584$, $p< 0.01$) and also when the NPC finished talking ($r=.414$, $p<0.05$) (Table 23). This is confirmed by Bayes correlations of GEQ Challenge score with the Contempt metric in the Initial Opinion event with very strong level of evidence for the hypothesis (H1) 46.416 times over the null hypothesis (H0) (Figure 46).

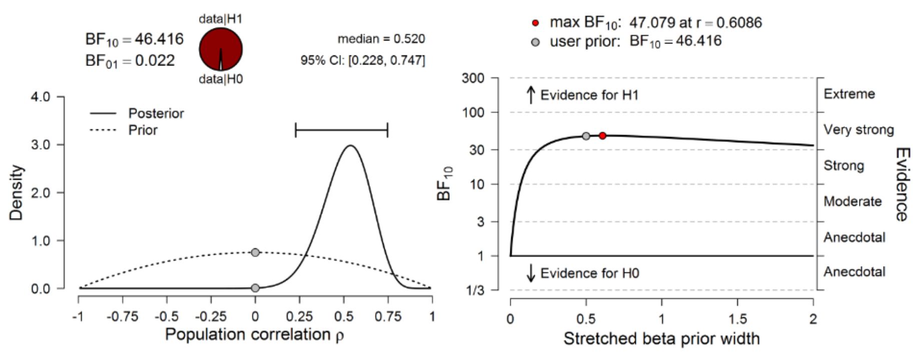


Figure 46. Results from Bayesian correlation pairs analyses for GEQ Challenge score and Contempt metric in the Initial opinion event. CI, Credibility Interval; BF, Bayes Factor.

This shows the player's perceived challenge in taking decision about the dilemma and also providing justification about his decision (orally). Moreover, there is significant correlation between Challenge and Fear ($r=.414$, $p< 0.05$) when the NPC finished his talking. This is shown by Bayes Factor for the correlation hypothesis (H1) with 8.783 times over the null hypothesis (H0) between GEQ Challenge score and the Fear metric in the NPC ends talking event, with moderate (close to strong) level of evidence (Figure 47).

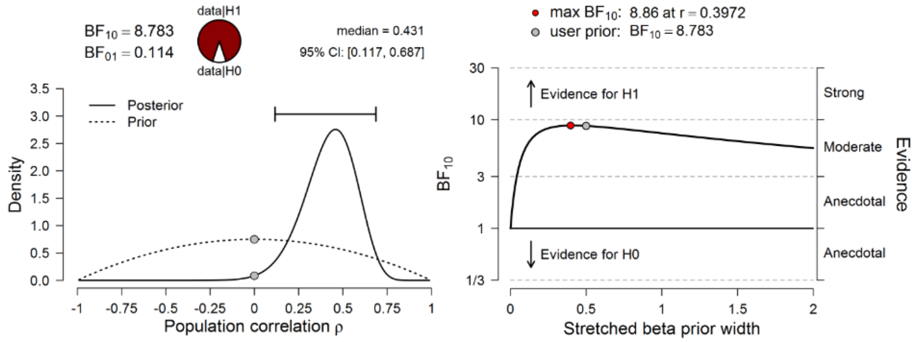


Figure 47. Results from Bayesian correlation pairs analyses for GEQ Challenge score and Fear metric in the NPC ends talk event. CI, Credibility Interval; BF, Bayes Factor.

This suggests that the participants felt the challenge of the evaluation of the NPC's opinion with Contempt and Fear emotions.

Negative affect: The Negative affect GEQ dimension is more correlated (negatively) with Happy measure in the Initial Opinion phase (r=-.444, p<0.05) and also with Neutral (r=.436, p<0.05) (Table 23). This is confirmed by Bayes correlations of GEQ Negative affect score with the Neutral metric in the Initial Opinion event with moderate level of evidence for the hypothesis (H1) 4.058 times over the null hypothesis (H0) (Figure 48).

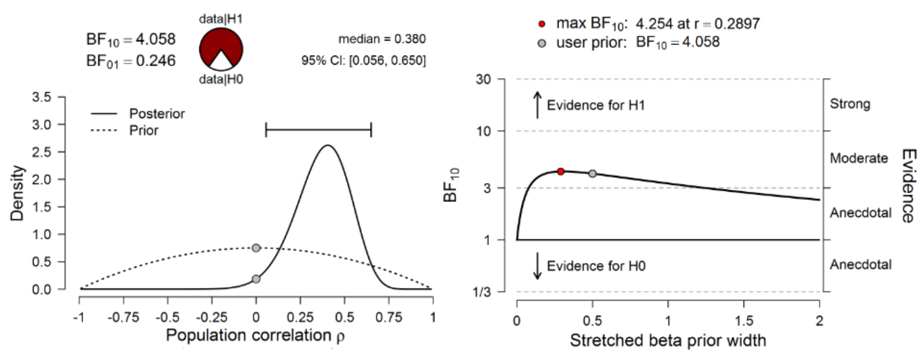


Figure 48. Results from Bayesian correlation pairs analyses for GEQ Negative Affect score and Neutral metric in the Initial opinion event. CI, Credibility Interval; BF, Bayes Factor.

Always in the same event (Initial opinion), the Bayes Factor for the correlation hypothesis (H1) was favoured 4.522 times over the null hypothesis (H0) between Negative affect score and Happy measure, with moderate level of evidence (Figure 49).

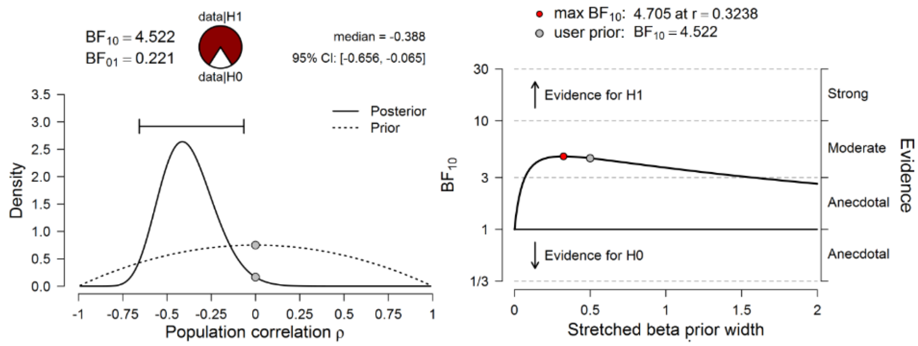


Figure 49. Results from Bayesian correlation pairs analyses for GEQ Negative Affect score and Happy metric in the Initial opinion event. CI, Credibility Interval; BF, Bayes Factor.

This shows the player's felt negative affect in taking decision about the dilemma and also providing justification about his decision (especially when "No" has to be said).

Positive affect: The Positive affect GEQ dimension is significantly correlated (positively) with Happy measure in the Initial Opinion phase ($r=.619$, $p<0.01$) (Table 23). This is confirmed by Bayes correlations of GEQ Positive affect score with the Happy metric in the Initial Opinion event with extreme level of evidence for the hypothesis (H1) 101.987 times over the null hypothesis (H0) (Figure 50).

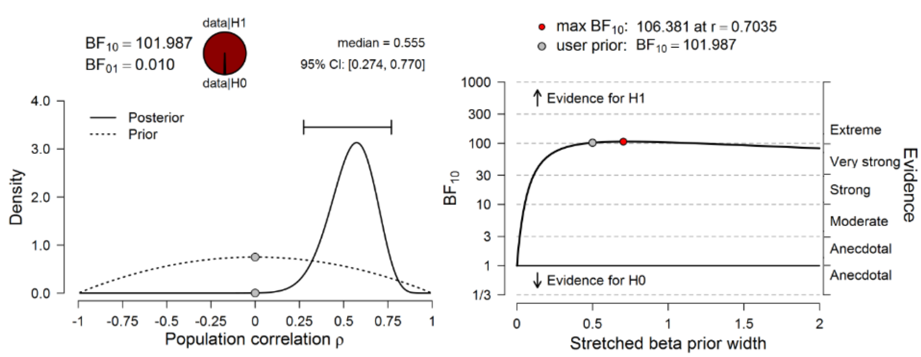


Figure 50. Results from Bayesian correlation pairs analyses for GEQ Positive Affect score and Happy metric in the Initial opinion event. CI, Credibility Interval; BF, Bayes Factor.

This shows the player's felt positive affect in taking decision about a simple dilemma (according to them) and providing justification (especially when the justification is obvious from his point of view). Moreover, there is correlation (negative) between GEQ Positive affect score and disgust measure ($r=-.425$, $p<0.05$) when the Player evaluates the NPC (Table 23).

This is shown by Bayes correlations of GEQ Positive affect score with the Disgust metric in the Player Evaluates NPC event with moderate level of evidence for the hypothesis (H1) 3.564 times over the null hypothesis (H0) (Figure 51).

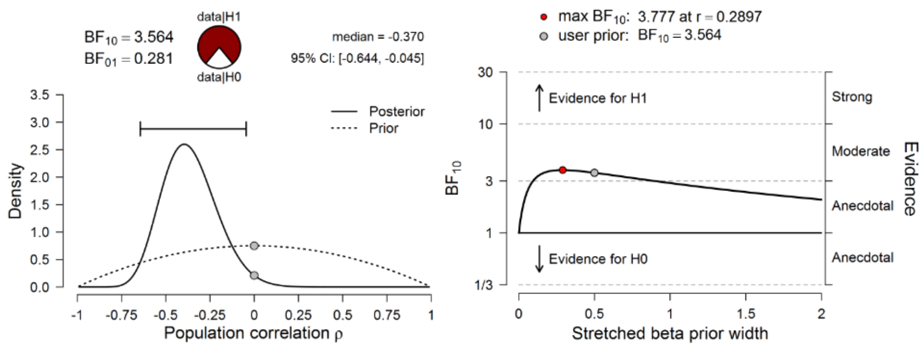


Figure 51. Results from Bayesian correlation pairs analyses for GEQ Positive Affect score and Disgust metric in the Player evaluates NPC event. CI, Credibility Interval; BF, Bayes Factor.

This suggests that the more the participant are not disgusted with the NPCs' opinions the more they have positive affect to the game.

Negative Experience: The Negative Experience GEQ dimension is more correlated (negatively) with Distraction measure in the Initial Opinion phase ($r=-.422$, $p<0.05$) (Table 23). This is shown by Bayes correlations of Negative experience score with the Distraction metric in the Initial opinion event with moderate level of evidence for the hypothesis (H1) 3.405 times over the null hypothesis (H0) (Figure 52).

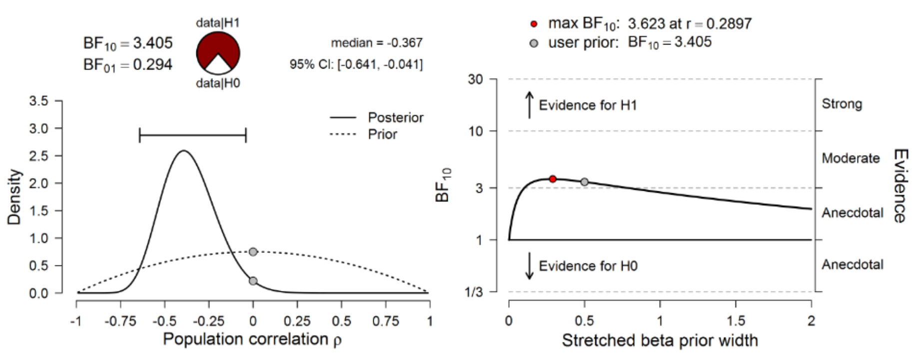


Figure 52. Results from Bayesian correlation pairs analyses for GEQ Negative experience score and Distraction metric in the Initial opinion event. CI, Credibility Interval; BF, Bayes Factor.

This shows the player's felt negative Experience in taking decision about the dilemma where he has to pay attention in providing justification about his decision (Low values of Distraction index show high attention state).

Tiredness: According to Table 23, the GEQ dimension of Tiredness is correlated (negatively) with the Contempt measure in all game events except the initial opinion event with moderate to large effect sizes. For NPC talk start and talk end events correlations are respectively r=-.466 and r=-.455 with p <0.05. The correlations are more significant for the events NPC Evaluation event r=-.498 and NPC non-verbal reaction event r=-.521 with p <0.01. This is confirmed by Bayes correlations of GEQ Competence score with the Contempt metric in all the game events except the initial opinion event with moderate to strong levels of evidence. For example, for the NPC non-verbal reaction event the Bayes Factor for the correlation hypothesis (H1) was favoured 14.154 times over the null hypothesis (H0) between Tiredness score and Contempt, with strong level of evidence (Figure 53).

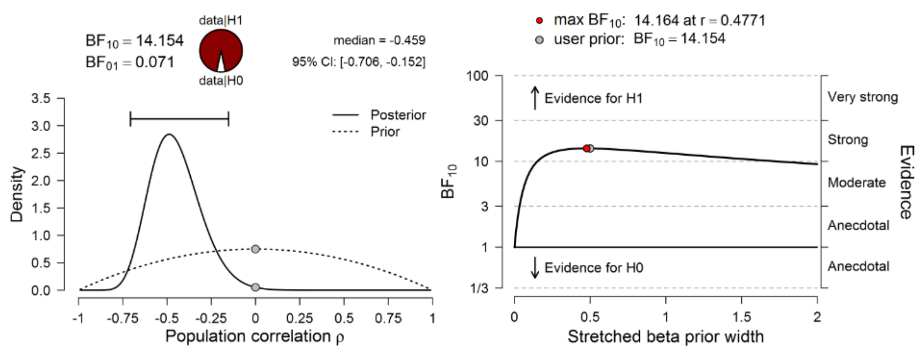


Figure 53. Results from Bayesian correlation pairs analyses for GEQ Tiredness score and Contempt metric in the NPC non-verbal reaction event. CI, Credibility Interval; BF, Bayes Factor.

This means that the more they are feeling tired the less they show a contempt emotion. Moreover, there is significant correlation between tiredness and valence measure (r=-.539, p<0.01) when the NPC started his talking. This is shown by Bayes Factor for the correlation hypothesis (H1) with 19.460 times over the null hypothesis (H0) between GEQ Tiredness score and Valence metric in the NPC starts talking event, with strong level of evidence (Figure 54).

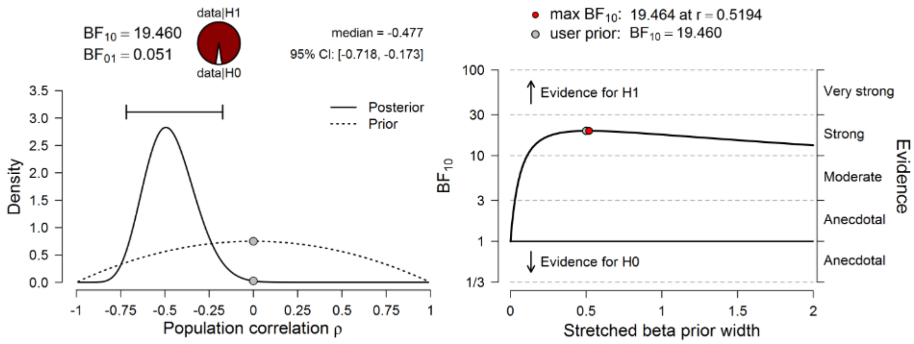


Figure 54. Results from Bayesian correlation pairs analyses for GEQ Tiredness score and Valence metric in the NPC starts talking event. CI, Credibility Interval; BF, Bayes Factor.

There is also a correlation between Tiredness and Arousal (r=-.436, $p< 0.05$) when the player evaluated the NPC. But we cannot confirm this correlation because the Bayesian Factor analysis shows only anecdotal evidence to correlation hypothesis (H1). We can conclude that the player felt tired of the confrontation with the NPC at the beginning of the interaction with NPC and also when the NPC shows his non-verbal reaction after the evaluation.

Behavioural Involvement: The behavioural involvement is about the social interaction with the between the player and the NPC. This dimension is significantly correlated negatively with the sad measure (r=-.528, $p<.01$) at the initial opinion phase (Table 23). This is confirmed by Bayes correlations of GEQ Behavioural involvement score with Sadness metric in the Initial opinion event with strong level of evidence for the hypothesis (H1) 16.079 times over the null hypothesis (H0) (Figure 55).

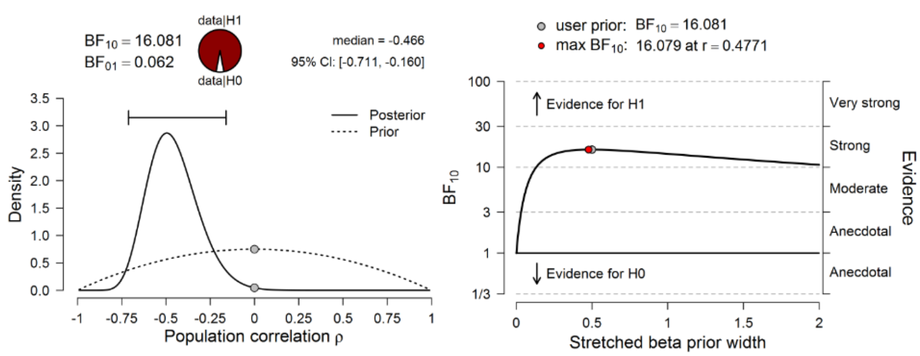


Figure 55. Results from Bayesian correlation pairs analyses for GEQ Behavioural involvement score and Sadness metric in Initial opinion event. CI, Credibility Interval; BF, Bayes Factor.

This means the less the player is sad about the dilemma, even before meeting the NPC, the more he feels involved in the situation.

Persuasion (Opinion Change): The persuasion score is based on supplementary question about how much the participant believe he changed his opinion after hearing the other NPCs' point of view (the answer between 1-(not-at-all) to 7-(definitely)) is correlated (positively) with the Long-term Engagement (LT-Eng) in all game events with moderate to large effect sizes ($0.30 < r < 0.65$, $p < 0.05$). The long term engagement (LT-Eng) is calculated as the 3rd quarter on a 10 sec the engagement index as it was found significant measure in studies on persuasion in online debates (M. S. Benlamine et al., 2017; Villata et al., 2017). This can be explained by the fact that a persuading argument is more engaging than a rejected argument but this Engagement must be measured on a longer period (10 sec and not 1 sec). We note that the persuasion score is also correlated (positively) with the Engagement index only in the Initial opinion event when the participant has to take decision about the dilemma. But with Bayesian Factor analysis we get only anecdotal evidences to correlation hypothesis between Persuasion and Long term engagement or persuasion and engagement.

Moreover, the Persuasion score is correlated (negatively) with the Disgust measure in all game events except the initial opinion event with moderate to large effect sizes. The correlations are more significant for the events NPC starts talk event $r= -.531$, NPC Evaluation event $r=-.527$ and NPC non-verbal reaction event $r=-.595$ with $p <0.01$. This is confirmed by Bayes correlations of Persuasion score with the Disgust metric in all the game events except the initial opinion event with moderate to very strong levels of evidence. For example, for the NPC non-verbal reaction event the Bayes Factor for the correlation hypothesis (H1) was favoured 58.406 times over the null hypothesis (H0) between Persuasion score and Disgust, with very strong level of evidence (Figure 56).

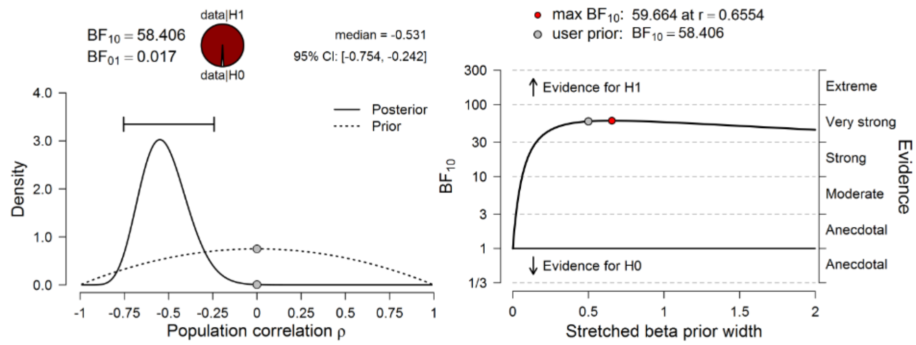


Figure 56. Results from Bayesian correlation pairs analyses for Persuasion score and Disgust metric in the NPC non-verbal reaction event. CI, Credibility Interval; BF, Bayes Factor.

This means that the more the participant are feeling disgusted the less they are persuaded especially when the NPC makes non-verbal reaction.

Social Score: the social score is calculated according to the reference answer and the comparison between the player's score and the NPC's social level. This score is correlated (Positively) with the surprise measure ($r=.368$, $p<.05$) at the NPC evaluation phase (Table 23). But with Bayesian Factor analysis, we get only anecdotal evidence to correlation hypothesis (H1). Thus, we cannot confirm this correlation result.

9.4.2 Assessment of the effect of the adaptive rules

The second analysis aims to assess the impact of the different adaptive rules in context. For example, is there an impact of the different facial animations of the avatars on the participant's average measure of emotions (as measured one second after the interaction)? In other words, can we verify that a facial animation of the NPC during the interaction influences the participant's emotions? Or has the music change any impact on the player's affective and cognitive measures? To test this hypothesis, we have performed repeated measures analysis of variance (ANOVA) for each of the affective measures when the rule is triggered. We only present the measures where we found significant results. The sample size is variously reported because there was some missing observations for some participants (for example, some participants did not have Disgust face with thumbs down because their responses are correct with regards to the reference answers). In SPSS, subjects with missing data are excludes in the analysis (i.e. listwise exclusion).

9.4.2.1 Dialogue with avatar: (Mimetic at the end of NPC talk).

The dependent variables are the participant's average value of the affective measure during one second when the NPC finished talking for each of the facial animations of the NPC (NPCHappyFace, NPCNeutralFace and NPCDisgustFace).

Happy.

Let's take a look at the Descriptive Statistics table shown below (Table 24). Happy face got maximum average value of participants' Happy measure (m = 32.21). Disgust face got the minimum average value of Happy (m = 22.14). This shows a large difference (≈10) between the two NPC facial animations in term of the participant's Happy measure.

Table 24. Mimetic rules' descriptive statistics (Happy measure).

NPC Facial Animations	Mean	Std. Deviation	N
DisgustFace	22.14	16.68	27
HappyFace	32.21	14.09	27
NeutralFace	25.53	8.75	27

To verify the sphericity assumption on our data, we use Mauchly's test. According to this test, $\chi^2(2)=.904$, $p=.636>.05$ did not indicate any violation of sphericity. So, for the within-subject effect test, we have a significant difference between the means of Happy measure for each NPC facial animations: $F(2, 52) = 4.271$ and $p=.019$. In this case, we have enough evidence to reject the null hypothesis and say that it is unlikely that the average measure of Happy for each group of NPC Facial Animations will be the same.

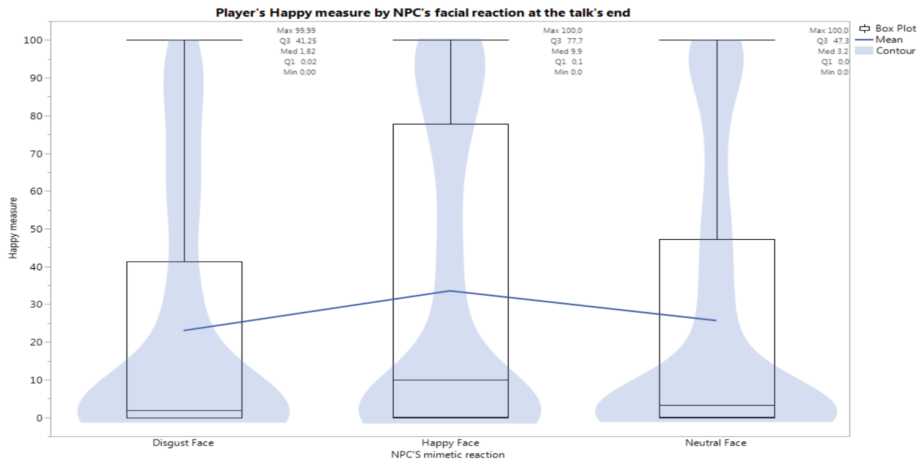


Figure 57. The mean effect of NPC mimetic (facial animations) on the participant's Happy measure in the NPC_end_talk event.

As shown in Figure 57, we present the corresponding measure of Happy after NPC finished his talk to compare the effect of NPC mimetic (facial animation) on the participant's Happy measure. The mimetic rules are based on the emotional state of the participant during the last 30 seconds where the NPC mimes the participant's emotional state in order to maintain his emotion. Using Bonferroni post-hoc test, we analysed the difference of the effect between the NPC Facial animations on the measure of Happy. On the one hand, the average participants' Happy measure when the NPC displays a Happy_Face differs significantly from those when the NPC displays a Disgust_face ($d=10.072$, $p_{bonf}=.032$). On the other hand, we observe a non-significant difference in mean between the Neutral_Face and Disgust_face ($d=3.388$, $p_{bonf}=1$) and between Happy_Face and Neutral_Face ($d=6.684$, $p_{bonf}=.132$).

Anger.

Let's take a look at the Descriptive Statistics table shown below (Table 25). Disgust face got maximum average value of participants' Anger measure ($m = 18.53$). Happy face got the minimum average value of Anger ($m = 8.88$). This shows a large difference (≈10) between the two NPC facial animations in term of the participant's Anger measure.

Table 25. Mimetic rules' descriptive statistics (Anger measure)

NPC Facial Animations	Mean	Std. Deviation	N
DisgustFace	18.53	15.52	27
HappyFace	8.88	4.64	27
NeutralFace	12.04	3.74	27

To verify the sphericity assumption on our data, we use Mauchly's test. According to this test, $\chi^2(2)=25.6$, $p=.000<.05$ indicating a violation of sphericity assumption. According to this test, we assess the significance of the corresponding F with Greenhouse-Geisser correction ($p=.609<0.75$). So, for the within-subject effect test by applying Greenhouse-Geisser correction, we have a significant difference between the means of Anger measure for each NPC facial animations: $F(1.219, 31.691) = 7.093$ and $p=.009$. In this case, we have enough evidence to reject the null hypothesis and say that it is unlikely that the average measure of Anger for each group of NPC Facial Animations will be the same.

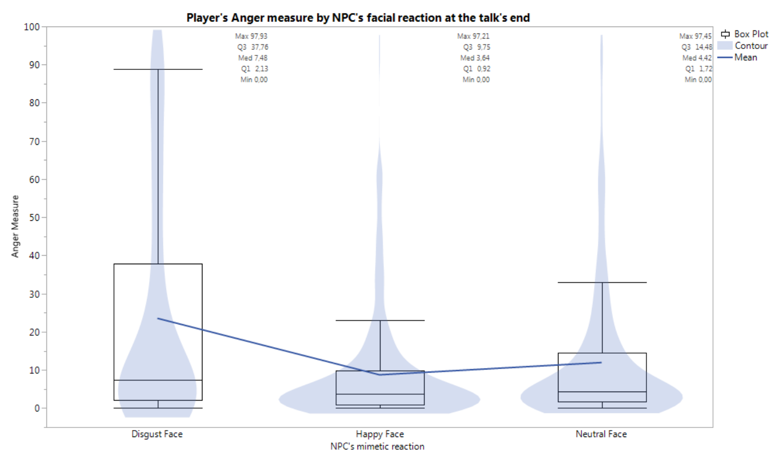


Figure 58. The mean effect of NPC mimetic (facial animations) on the participant's Anger measure in the NPC_end_talk event

As shown in Figure 58, we present the corresponding measure of Anger after NPC finished his talk to compare the effect of NPC mimetic (facial animation) on the participant's Anger measure. Using Bonferroni post-hoc test, we analysed the difference of the effect between the NPC Facial animations on the measure of Anger. On the one hand, there are significant differences on average participants' Anger measure between the NPC Disgust_face and Happy_Face ($d=9.655$, $p_{bonf}=.011$) and also between Neutral_Face and Happy face ($d=3.524$,

p_{bonf} =.016). On the other hand, we observe non-significant difference (but close) -in mean between the Disgust_face and Neutral_Face (d=6.131, p_{bonf} =.176).

9.4.2.2 Evaluating the avatar (NPC non-verbal reaction to player's evaluation of the NPC).

The dependent variables are the participant's average value of the affective measure during one second when the NPC made a non-verbal reaction (NPCDisgustFace, NPCDisgustFace + Thumbs_Down, NPCHappyFace and NPCHappyFace + Thumbs_Up).

Happy.

Let's take a look at the Descriptive Statistics table shown below (Table 26). DisgustFace got maximum average value of participants' Happy measure (m = 30.76). DisgustFace+ThumbsDown got the minimum average value of Happy (m = 19.35). This shows a large difference (≈10) between the two NPC non-verbal reactions in term of the participant's Happy measure.

Table 26. Non-verbal reaction rules' descriptive statistics (Happy measure).

NPC Non-verbal Reaction	Mean	Std. Deviation	N
DisgustFace	30.76	10.32	27
DisgustFace + ThumbsDown	19.35	19.43	27
HappyFace	28.24	9.80	27
HappyFace + ThumbsUp	29.49	13.05	27

To verify the sphericity assumption on our data, we use Mauchly's test. According to this test, $\chi^2(5)=13.663$, p=.018<.05 indicating a violation of sphericity assumption. According to this test, we assess the significance of the corresponding F with Greenhouse-Geisser correction (p=.748<0.75). So, for the within-subject effect test by applying Greenhouse-Geisser correction, we have a significant difference between the means of Happy measure for each NPC non-verbal reactions: $F(2.243, 58.307) = 4.912$ and p=.008. In this case, we have enough evidence to reject the null hypothesis and say that it is unlikely that the average measure of Happy for each group of NPC non-verbal reactions will be the same.

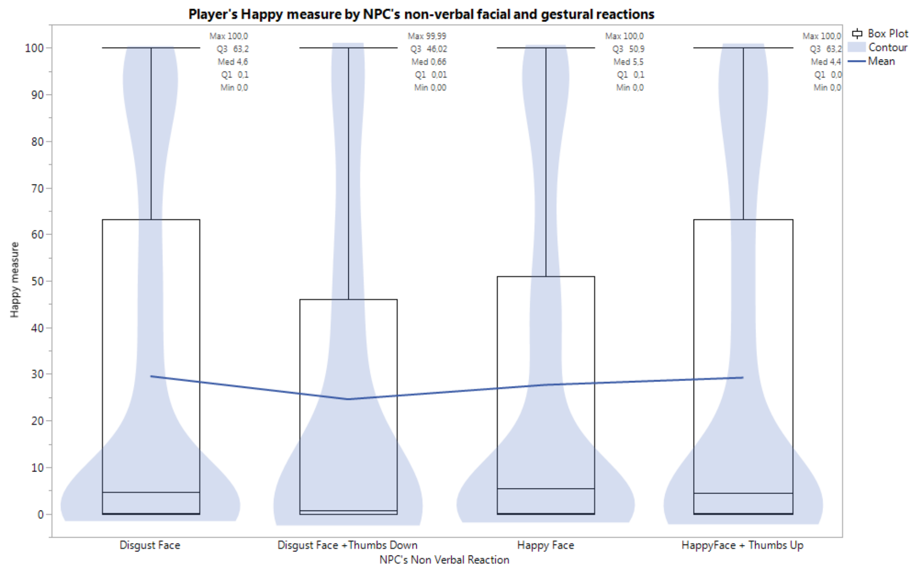


Figure 59. The mean effect of NPC non-verbal reaction (facial and gestural animations) on the participant's Happy measure in the NPC_non-verbal reaction event

As shown in Figure 59, we present the corresponding measure of Happy after interacting with the NPC to compare the effect of NPC non-verbal reaction (facial and gestural feedbacks) on the participant's emotion of Happy. The non-verbal reaction of the NPC to the player evaluation is based on two emotional rules (facial expression feedback rules and gestural feedback rules) where the NPC make facial animation depending on being liked or dislike by the participant and also gestural animation (Thumbs up or Thumbs down) that depends also to the participant's social score compared to the NPC social level. Using Bonferroni post-hoc test, we analysed the difference of the effect between the NPC's non-verbal reactions on the measure of Happy. On the one hand, the average participants' Happy measure has significant difference between Disgust_Face and Disgust_face+Thumbs_down ($d=11.409$, $p_{bonf}=.011$). On the other hand, we did not find any other significant difference.

9.4.2.3 Music Rules.

For the music change effect on player affective and cognitive measures, we have done a repeated measures ANOVA analyzes for each of the measures to compare the impact of the three types of music (Very_Joyful, Moderately_Joyful and Less_Joyful).

Happy.

Let's take a look at the Descriptive Statistics table shown below (Table 27). Very_Joyful Music got maximum average value of participants' Happy measure (m = 52.81). Moderately_Joyful Music got the minimum average value of Happy (m = 11.17). This shows a large difference (≈40) between the two types of background music in term of the participant's Happy measure.

Table 27. Music rules' descriptive statistics (Happy measure).

Music types	Mean	Std. Deviation	N
Very_Joyful	52.81	15.22	28
Moderately_Joyful	11.17	6.57	28
Less_Joyful	14.95	12.12	28

To verify the sphericity assumption on our data, we use Mauchly's test. According to this test, $\chi^2(2)=11.156$, $p=.004<.05$ indicating a violation of sphericity assumption. According to this test, we assess the significance of the corresponding F with Greenhouse-Geisser correction ($p=.741<0.75$). So, for the within-subject effect test by applying Greenhouse-Geisser correction, we have a significant difference between the means of Happy measure for each of background music types: $F(1.483, 40.033) = 106.184$ and $p<.001$. In this case, we have enough evidence to reject the null hypothesis and say that it is unlikely that the average measure of Happy for each group of music types will be the same.

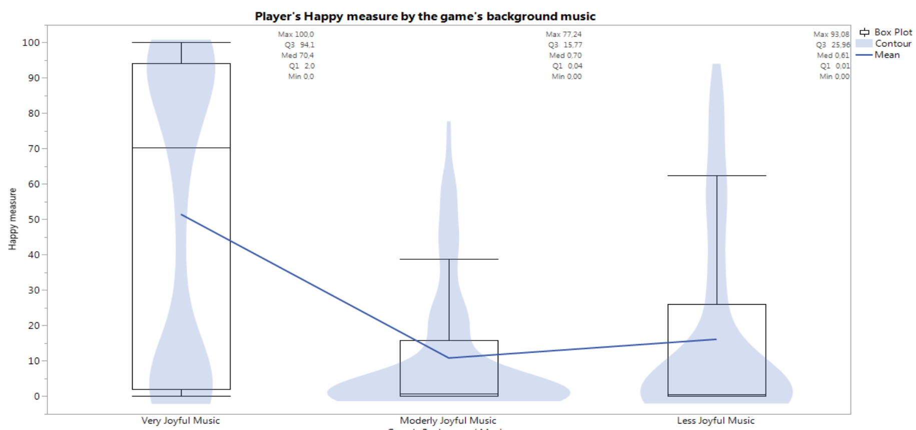


Figure 60. The mean effect of Music types on user's Happy measure when the music change.

As shown in Figure 60, we present the corresponding measure of Happy after background music change (according to the music emotional rules depending on the player emotional state) to compare the effect of Music Types on the participant's emotion of Happy. The music player rules are based on the emotional state of the participant during the last 30 seconds where the game background music meets the participant's emotional state in order to maintain his emotion. Using Bonferroni post-hoc test, we analysed the difference of the effect between the music types on the measure of Happy. On the one hand, the average participants' Happy measure has significant difference between Very_Joyful and Moderately_Joyful music (d=41.641, p_{bonf} <.001) and also Very_Joyful and Less_Joyful music (d=37.864, p_{bonf} <.001). On the other hand, we observe a non-significant difference in mean between the Less_Joyful and Moderately_Joyful music (d=3.777, p_{bonf} =.398).

Anger.

Let's take a look at the Descriptive Statistics table shown below (Table 28). Less_Joyful music got maximum average value of participants' Anger measure (m = 32.93). Very_Joyful music got the minimum average value of Anger (m = 8.56). This shows a large difference (≈24) between the two types of background music in term of the participant's Anger measure.

Table 28. Music rules' descriptive statistics (Anger measure).

Music types	Mean	Std. Deviation	N
Very_Joyful	8.56	4.97	28
Moderately_Joyful	12.38	4.55	28
Less_Joyful	32.93	20.78	28

To verify the sphericity assumption on our data, we use Mauchly's test. According to this test, $\chi^2(2)$=33.931, p=.000<.05 indicating a violation of sphericity assumption. According to this test, we assess the significance of the corresponding F with Greenhouse-Geisser correction (p=.578<0.75). So, for the within-subject effect test by applying Greenhouse-Geisser correction, we have a significant difference between the means of Anger measure for each of background music types: F(1. 157, 31.235) = 28.319 and p<.001. In this case, we have enough evidence to reject the null hypothesis and say that it is unlikely that the average measure of Anger for each group of music types will be the same.

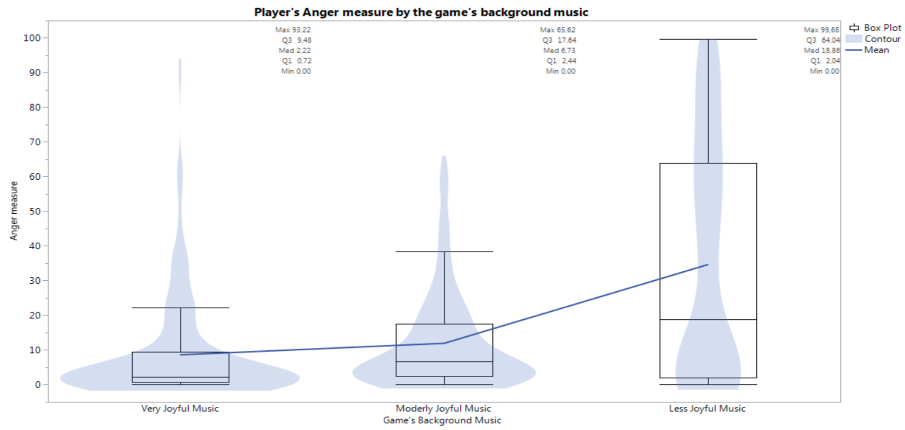


Figure 61. The mean effect of Music types on user's Anger measure when the music change.

As shown in Figure 61, we present the corresponding measure of Anger after background music change to compare the effect of Music Types on the participant's emotion of Anger. Using Bonferroni post-hoc test, we analysed the difference of the effect between the music types on the measure of Anger. The average participants' Anger measure has significant differences between the three music types: Moderately_Joyful and Very_Joyful music ($d=3.825$, $p_{bonf}=.025$), Less_Joyful and Very_Joyful music ($d=24.369$, $p_{bonf}<.001$) and also Less_Joyful and Moderately_Joyful music ($d=20.544$, $p_{bonf}<.001$).

Surprise.

Let's take a look at the Descriptive Statistics table shown below (Table 29). Very_Joyful music got maximum average value of participants' Surprise measure ($m = 34.65$). Less_Joyful music got the minimum average value of Surprise ($m = 12.92$). This shows a large difference (≈22) between the two types of background music in term of the participant's Surprise measure.

Table 29. Music rules' descriptive statistics (Surprise measure).

Music types	Mean	Std. Deviation	N
Very_Joyful	34.65	8.94	28
Moderately_Joyful	19.38	7.94	28
Less_Joyful	12.92	13.08	28

To verify the sphericity assumption on our data, we use Mauchly's test. According to this test, $\chi^2(2)=3.813$, $p=.149>.05$ did not indicate any violation of sphericity. So, for the within-subject effect test, we have a significant difference between the means of Surprise measure for each of background music types: $F(2, 54) = 37.501$ and $p<.001$. In this case, we have enough evidence to reject the null hypothesis and say that it is unlikely that the average measure of Surprise for each group of music types will be the same.

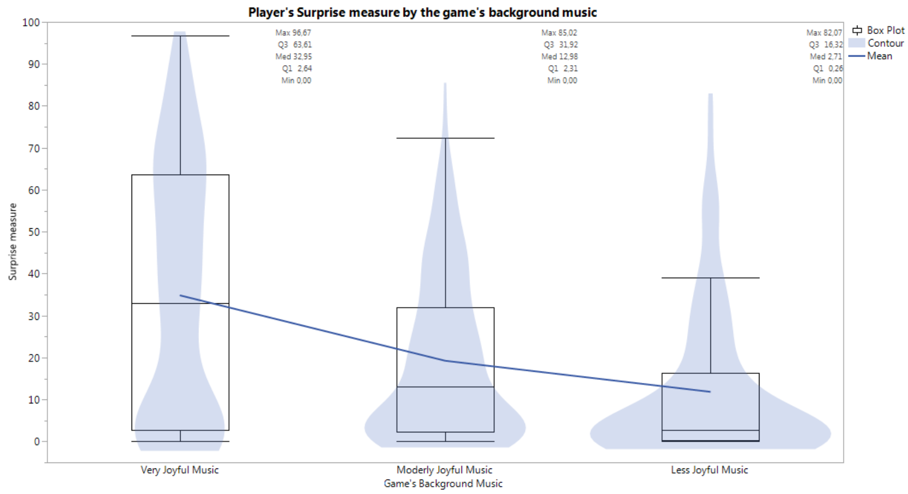


Figure 62. The mean effect of Music types on user's Surprise measure when the music change.

As shown in Figure 62, we present the corresponding measure of Surprise after background music change to compare the effect of Music Types on the participant's emotion of Surprise. Using Bonferroni post-hoc test, we analysed the difference of the effect between the music types on the measure of Surprise. The average participants' Surprise measure has significant differences between the three music types: Very_Joyful and Moderately_Joyful music ($d=15.269$, $p_{bonf} <.001$), Very_Joyful and Less_Joyful music ($d=21.729$, $p_{bonf} <.000$) and also Moderately_Joyful and Less_Joyful music ($d=6.460$, $p_{bonf} =.045$).

Contempt.

Let's take a look at the Descriptive Statistics table shown below (Table 30). Moderately_Joyful music got maximum average value of participants' Contempt measure ($m = 51.95$). Less_joyful music got the minimum average value of contempt ($m = 25.47$). This shows a large difference (≈26) between the two types of background music in term of the participant's Contempt measure.

Table 30. Music rules' descriptive statistics (Contempt measure).

Music types	Mean	Std. Deviation	N
Very_Joyful	36.79	8.56	28
Moderately_Joyful	51.95	9.98	28
Less_Joyful	25.47	12.45	28

To verify the sphericity assumption on our data, we use Mauchly's test. According to this test, $\chi^2(2)=2.263$, $p=.322>.05$ did not indicate any violation of sphericity. So, for the within-subject effect test, we have a significant difference between the means of Contempt measure for each of background music types: $F(2, 54) = 53.123$ and $p<.001$. In this case, we have enough evidence to reject the null hypothesis and say that it is unlikely that the average measure of Contempt for each group of music types will be the same.

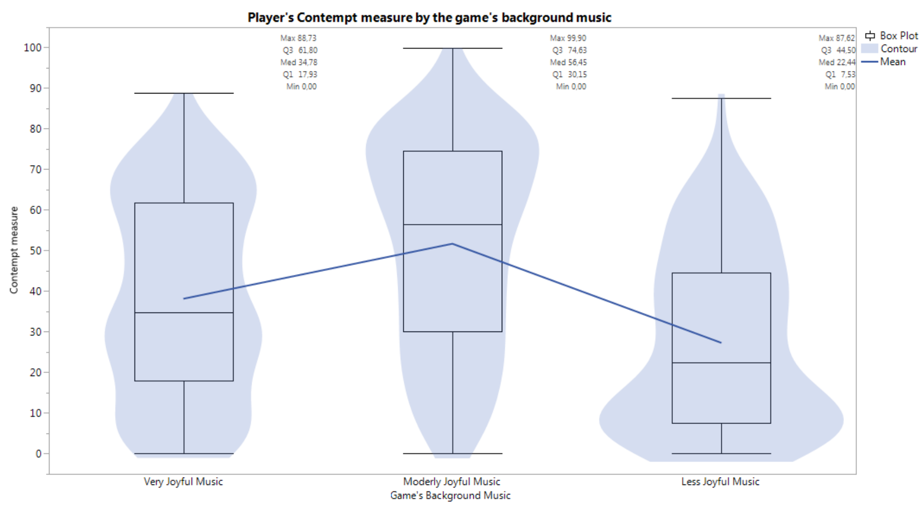


Figure 63. The mean effect of Music types on user's Contempt measure when the music change.

As shown in Figure 63, we present the corresponding measure of Contempt after background music change to compare the effect of Music Types on the participant's emotion of Contempt. Using Bonferroni post-hoc test, we analysed the difference of the effect between the music types on the measure of Contempt. The average participants' Contempt measure has significant differences between the three music types: Moderately_Joyful and Very_Joyful music ($d=15.155$, $p_{bonf} <.000$), Moderately_Joyful and Less_Joyful music ($d=26.478$, $p_{bonf} <.000$) and also Very_Joyful and Less_Joyful music ($d=11.323$, $p_{bonf} =.001$).

Neutral.

Let's take a look at the Descriptive Statistics table shown below (Table 31). Moderately_Joyful music got maximum average value of participants' Neutral measure (m = 87.98). Less_Joyful music got the minimum average value of Neutral (m = 56.37). This shows a large difference (≈10) between the two types of background music in term of the participant's Neutral measure.

Table 31. Music rules' descriptive statistics (Neutral measure).

Music types	Mean	Std. Deviation	N
Very_Joyful	64.29	15.108	28
Moderately_Joyful	87.98	6.712	28
Less_Joyful	56.37	21.245	28

To verify the sphericity assumption on our data, we use Mauchly's test. According to this test, $\chi^2(2)=6.694$, $p=.035<.05$ indicating a violation of sphericity assumption. According to this test, we assess the significance of the corresponding F with Huynh-Feldt correction (because Greenhouse-Geisser $p=.815>0.75$). So, for the within-subject effect test by applying Huynh-Feldt correction, we have a significant difference between the means of Neutral measure for each of background music types: $F(1.\ 720, 46.444) = 28.295$ and $p<.000$. In this case, we have enough evidence to reject the null hypothesis and say that it is unlikely that the average measure of Neutral for each group of music types will be the same.

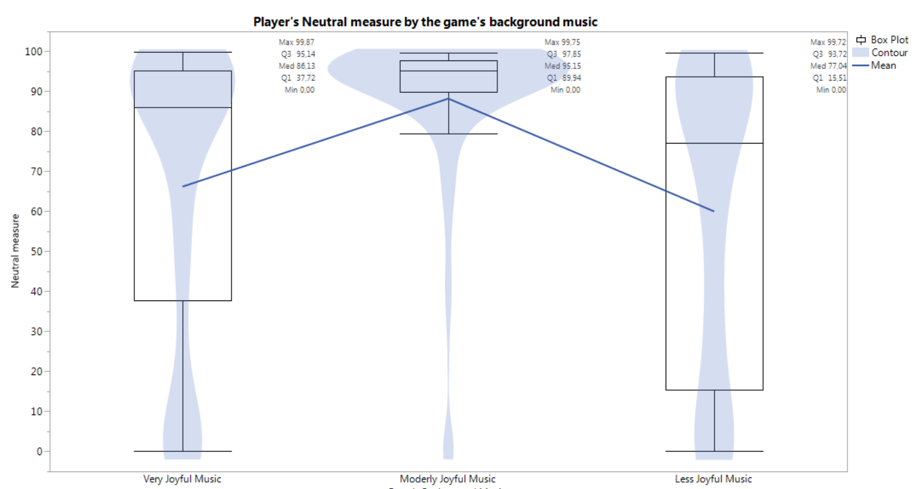


Figure 64. The mean effect of Music types on user's Neutral measure when the music change.

As shown in Figure 64, we present the corresponding measure of Neutral after background music change to compare the effect of Music Types on the participant's emotion of Neutral. Using Bonferroni post-hoc test, we analysed the difference of the effect between the music types on the measure of Neutral. On the one hand, the average participants' Neutral measure has significant difference between Moderately_Joyful and Very_Joyful music (d=23.692, p_{bonf}<.001) and also Moderately_Joyful and Less_Joyful music (d=31.615, p_{bonf} <.001). On the other hand, we observe a non-significant difference in mean between the Very_Joyful and Less_Joyful music (d=7.922, p_{bonf} =.412).

Disgust.

Let's take a look at the Descriptive Statistics table shown below (Table 32). Less_Joyful music got maximum average value of participants' Disgust measure (m = 39.88). Moderately_Joyful music got the minimum average value of Disgust (m = 10.24). This shows a large difference (≈29) between the two types of background music in term of the participant's Disgust measure.

Table 32. Music rules' descriptive statistics (Disgust measure).

Music types	Mean	Std. Deviation	N
Very_Joyful	21.20	10.24	28
Moderately_Joyful	10.24	6.16	28
Less_Joyful	39.88	21.18	28

To verify the sphericity assumption on our data, we use Mauchly's test. According to this test, $\chi^2(2)$=15.502, p=.000<.05 indicating a violation of sphericity assumption. According to this test, we assess the significance of the corresponding F with Greenhouse-Geisser correction (p=.690<0.75). So, for the within-subject effect test by applying Greenhouse-Geisser correction, we have a significant difference between the means of Disgust measure for each of background music types: F(1.380, 37.264) = 35.924 and p<.000. In this case, we have enough evidence to reject the null hypothesis and say that it is unlikely that the average measure of Disgust for each group of music types will be the same.

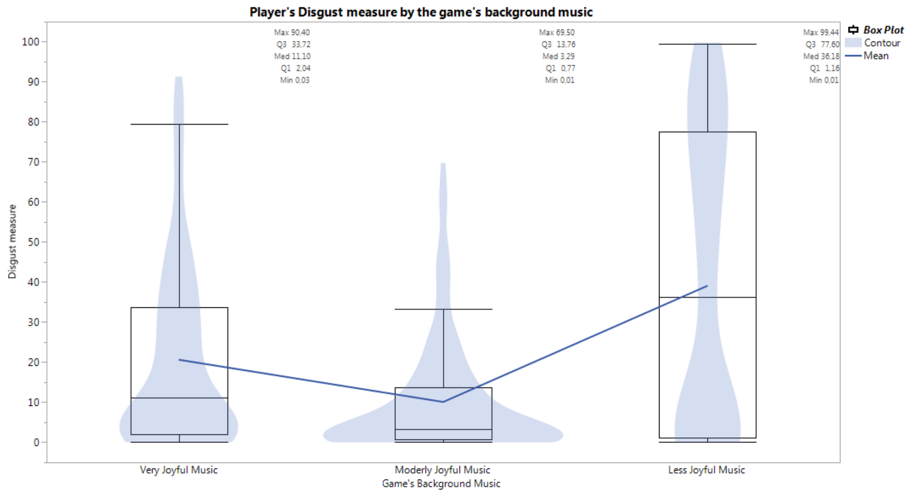


Figure 65. The mean effect of Music types on the user's Disgust measure when the music change.

As shown in Figure 65, we present the corresponding measure of Disgust after background music change to compare the effect of Music Types on the participant's emotion of Disgust. Using Bonferroni post-hoc test, we analysed the difference of the effect between the music types on the measure of Disgust. The average participants' Disgust measure has significant differences between the three music types: Very_Joyful and Moderately_Joyful music ($d=10.968$, $p_{bonf}<.001$), Less_Joyful and Very_Joyful music ($d=18.671$, $p_{bonf} <.000$) and also Less_Joyful and Moderately_Joyful music ($d=29.639$, $p_{bonf} <.000$).

Valence.

Let's take a look at the Descriptive Statistics table shown below (Table 33). Very_Joyful music got maximum average value of participants' Valence measure ($m = -11.46$). Moderately_Joyful music got the minimum average value of Valence ($m = -31.17$). This shows a large difference (≈20) between the two types of background music in term of the participant's Valence measure.

Table 33. Music rules' descriptive statistics (Valence measure).

Music types	Mean	Std. Deviation	N
Very_Joyful	-11.46	13.69	28
Moderately_Joyful	-31.17	15.80	28
Less_Joyful	-23.45	28.65	28

To verify the sphericity assumption on our data, we use Mauchly's test. According to this test, $\chi^2(2)=4.484$, $p=.106>.05$ did not indicate any violation of sphericity. So, for the within-subject effect test, we have a significant difference between the means of Valence measure for each of background music types: $F(2, 54) = 6.620$ and $p=.003$. In this case, we have enough evidence to reject the null hypothesis and say that it is unlikely that the average measure of Valence for each group of music types will be the same.

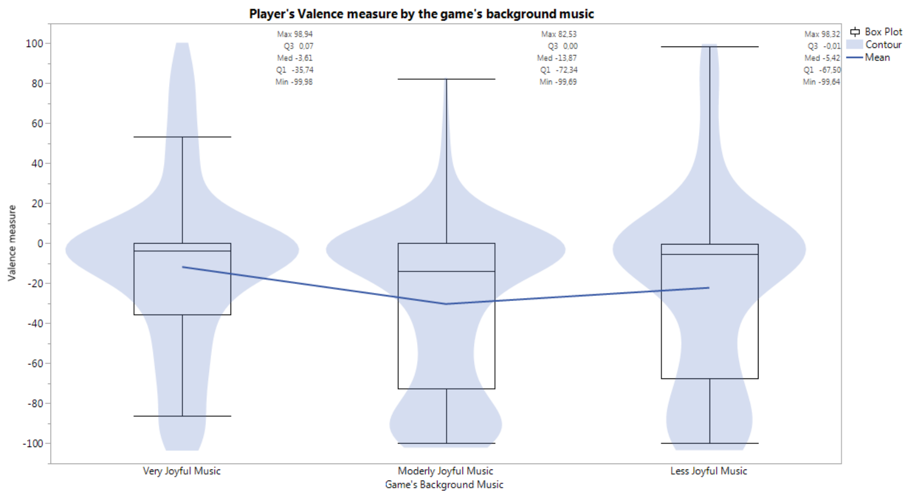


Figure 66. The mean effect of Music types on the user's Valence measure when the music change.

As shown in Figure 66, we present the corresponding measure of Valence after background music change to compare the effect of Music Types on the participant's measure of Valence. Using Bonferroni post-hoc test, we analysed the difference of the effect between the music types on the measure of Valence. On the one hand, the average participants' Valence measure has significant difference between Very_Joyful and Moderately_Joyful music ($d=19.708$, $p_{bonf}<.001$). On the other hand, we observe a non-significant difference in mean between the Less_Joyful and Moderately_Joyful music ($d=7.722$, $p_{bonf}=.706$) and also Very_Joyful and Less_Joyful music ($d=11.986$, $p_{bonf}=.103$).

Arousal.

Let's take a look at the Descriptive Statistics table shown below (Table 34). Less_Joyful music got maximum average value of participants' Arousal measure (m = 35.65). Moderately_Joyful music got the minimum average value of Arousal (m = 32.42).

Table 34. Music rules' descriptive statistics (Arousal measure).

Music types	Mean	Std. Deviation	N
Very_Joyful	34.10	3.71	28
Moderately_Joyful	32.42	2.39	28
Less_Joyful	35.65	4.46	28

To verify the sphericity assumption on our data, we use Mauchly's test. According to this test, $\chi^2(2)=3.330$, $p=.189>.05$ did not indicate any violation of sphericity. So, for the within-subject effect test, we have a significant difference between the means of Arousal measure for each of background music types: $F(2, 54) = 5.389$ and $p=.007$. In this case, we have enough evidence to reject the null hypothesis and say that it is unlikely that the average measure of Arousal for each group of music types will be the same.

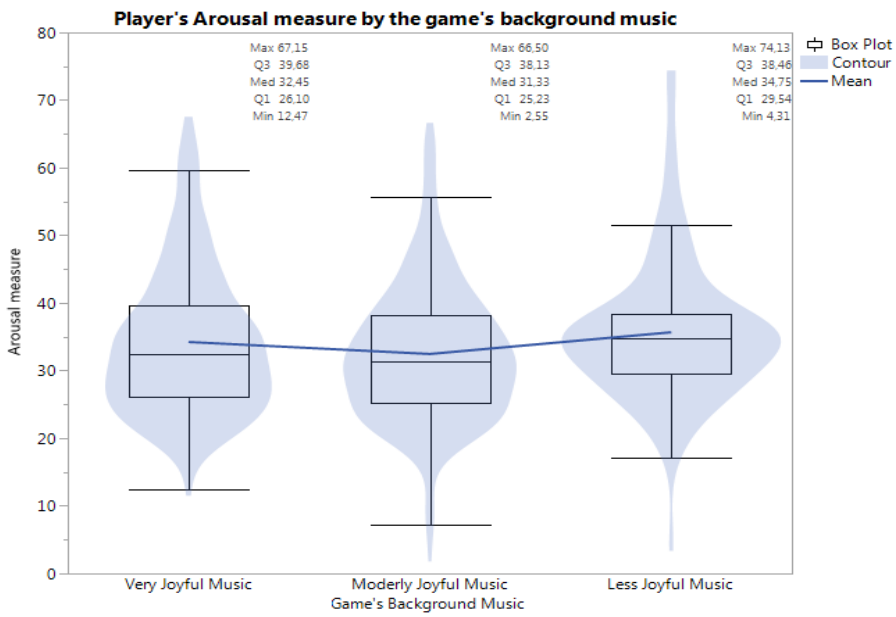


Figure 67. The mean effect of Music types on the user's Arousal when the music change.

As shown in Figure 67, we present the corresponding measure of Arousal after background music change to compare the effect of Music Types on the participant's measure of Arousal. Using Bonferroni post-hoc test, we analysed the difference of the effect between the music types on the measure of Arousal. On the one hand, the average participants' Arousal measure has significant difference between Less_Joyful and Moderately_Joyful music ($d=3.262$, $p_{bonf}=.003$). On the other hand, we observe a non-significant difference in mean between the Very_Joyful and Moderately_Joyful music (d=1.676, p_{bonf}=.220) and also Less_Joyful and Very_Joyful music (d=1.550, p_{bonf}=.556).

9.5 Discussion and Conclusion

This paper presented the BARGAIN framework and its application in adapting a virtual reality game “Dillemmas_VR” according to the player’s emotional state by creating emotional rules and the analysis of the experimentation data to detect any possible correlations between measures, game events and self-reported subjective experience, and also any possible effect of the emotional rules on the player’s affective and cognitive measures.

To begin with, the correlation analysis showed that cognitive measures has relationship with game experience questionnaire components, but that only flow and competence showed consistent correlation within the game events, respectively the flow is negatively correlated with Distraction index and competence is negatively correlated with FAA. There is also the persuasion (which is not from the GEQ questionnaire) is positively correlated with long-term engagement. This show that flow and competence can be automatically inferred from the cognitive measure. For the affective measure there is consistent negative correlation between tiredness and contempt measure. We have also consistent negative correlation between persuasion and disgust. This shows that the other correlations with GEQ components are game context sensitive and might change from a game to another.

Moreover, the repeated measure ANOVA showed significant effects of the mimetic rules with significant effects of the NPC‘s facial animations only on the Happy and Anger affective measures. The measurement of Happy showed statistically significant differences between Disgust Face + Thumbs down animation and the other NPC non-verbal reaction. For the music-player rules, we found that it has significant effect between the music types (Very_Joyful, Moderately_Joyful and Less_Joyful music) on all the affective measures except for fear and sad measures. Music types have different effects on affective measures that can be grouped in three groups: positively-valenced (Happy, Surprise and Valence), negatively-valenced (Anger, Disgust and Arousal) and non-valenced measures (Neutral and Contempt) as in (Jason M Harley et al., 2015). In fact the positively valenced measures were characterized with high levels for the Very_Joyful music, low levels for the Moderately_Joyful Music and medium levels for the Less_Joyful Music, the negatively-valenced measures have medium levels for the Ver_Joyful music, low levels for the Moderately_Joyful Music and high levels for the Less_Joyful Music and the non-valenced measures medium levels for the Very_Joyful music, High levels for the Moderately_Joyful Music and low levels for the Less_Joyful Music. We note also that the observed groups of affective reactions are coherent with the emotional states in the conditions of the music emotional rules. So, to measure the impact of the adaptive rule we should use the relative values, taking into account the difference between the initial affective measure before the application of the rule and the measures after.

Finally, the cognitive measure does not have significant differences between the facial animations in the mimetic rules, between the nonverbal reaction in the NPC reaction rules and also between the music types in the music-player rules. This may be due to the fact that the rules have only targeted the game's background music and the avatars' facial and gestural animations (do not require much mental effort) and not the difficulty of the game and the player's tasks. This suggest that the cognitive measures are better in analysing the game events that require mental effort (as shown by the correlations with GEQ dimensions).

In conclusion, the study reported here shows that affective and cognitive measures can be an indicator of the gameplay experience, as shown by correlation with subjective reports and game events. And also, the effect of the adaptation rules can be found statistically significant only on the affective measures. Ongoing work will examine the relationship between players' personality traits (Big-five questionnaire), affective and cognitive responses and game context in greater detail. We will also investigate the participants' attitude in the NPC opinion evaluations and what influences the user's persuasion in the NPC speech and facial animation.

Acknowledgments. The authors acknowledge support of the NSERC (Natural Sciences and Engineering Research Council of Canada), the FRQSC (Fonds de Recherche du Québec Société et Culture) and BMU Games for funding this research.

Chapitre 10 : Conclusion

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

10.1 Contributions

La **première contribution** de cette thèse a été de réaliser des études empiriques sur l'analyse affective dans les environnements virtuels (débat en ligne et jeux vidéo) et de développer un système de reconnaissance des expressions faciales à partir des signaux EEG. À cet effet, nous avons fait deux études empiriques sur l'analyse affective dans les environnements de débat en ligne. La première étude avait pour but d'analyser les relations entre des arguments et leurs rapports (attaque/support) avec les états émotionnels et l'engagement des participants. La deuxième étude avait pour but d'analyser l'effet des stratégies de persuasion (Ethos, Logos et Pathos) sur les mesures d'émotions et d'engagement des participants et aussi sur leurs changements d'opinion à la fin du débat (persuasion).

Ensuite, nous avons fait une expérience sur un jeu de tir « Danger Island » où nous avons intégré un menu d'auto-évaluation au cours du jeu. Avec ce menu, les joueurs peuvent rapporter leurs états émotionnels au cours du jeu. Nous avons analysé les états émotionnels rapportés et les niveaux d'engagements dans le jeu selon le genre des participants et leurs niveaux d'experience de jeux (*Gamers* vs *non-Gamers*). Enfin, dans la même expérience sur le jeu « danger Island », avant de commencer la session de jeu, nous avons demandé aux participants de regarder des images venant de la base d'images standardisées IAPS et rapporter leur émotion ressentie. Nous avons ainsi proposé une approche physiologique pour la détection des expressions faciales reliées aux émotions à partir des signaux d'EEG. Nous avons recueilli une base de données des caractéristiques extraites des signaux EEG et des expressions faciales correspondantes des participants via camera. Nous avons ensuite entrainé des algorithmes d'apprentissage machine pour construit les modèles de régression pour la prédiction des intensités des émotions à partir des signaux EEG. Nous avons trouvé que le Random-Forest donnait les meilleurs scores de prédiction. Cette contribution a fait l'objet de quatre publications présentées dans le chapitre 3, le chapitre 4, le chapitre 5, et le chapitre 6.

Notre **deuxième contribution** a consisté à développer « Emograph » une solution d'analyse émotionnelle dans les scènes des jeux et de prédire l'émotion dominante du joueur et sa motivation (orientations de but) dans une autre scène de jeu. Les expérimentations ont été faites sur un jeu commercial d'horreur « Outlast ». Pour ce faire, nous avons opté pour une approche multimodale qui regroupe plusieurs mesures de différents senseurs tel que le traceur de regard, les signaux EEG et les expressions faciales. Nous avons formalisé la description des scènes de jeu à l'aide de variables numériques venant du modèle OCC. Ensuite nous avons élaboré une méthode d'extraction de l'émotion dominante dans une scène de jeu en se basant sur les mesures d'expression faciale et les réponses d'auto-évaluation sur l'émotion ressentie par le joueur pour chacune des scènes de jeu. Nous avons construit des modèles de classification pour prédire l'émotion dominante dans une scènes de jeu en ayant comme entrée la description de la scènes sous forme de variable OCC, les traits de personnalité du joueur et les informations sociodémographiques (genre, âge et *Gamer*/*non-Gamer*). De même, nous avons opté pour prédire la motivation des joueurs en termes de leurs orientations de but. Pour ce faire, nous avons élaboré une méthode d'identification du but de maitrise (Maitrise/Performance) à partir du questionnaire du GameFlow en regardant les préférences de jeux du participant, sa moyenne scolaire, le nombre d'heures de jeu par semaine. Ensuite, nous avons identifié son comportement d'approche (Approche/Évitement) pour chacune des scènes de jeu à partir des moyennes de mesures FAA (*Frontal Alpha Asymmetry*). Nous avons de même construit des modèles de classification pour prédire l'orientation du but du joueur dans une scènes de jeu en ayant comme entrée la description de la scène sous forme de variables OCC, les traits de personnalité du joueur et les informations sociodémographiques. Pour ce faire, une étude expérimentale a été faite où nous avons collecté les données de 20 participants de l'Université de Montréal. Cette étude a duré 3 semaines. Durant cette étude, nous avons collecté différentes données physiologiques de l'apprenant. Ces données ont servi à construire notre approche, entrainer et tester différents algorithmes d'apprentissage machine. Les résultats préliminaires ont montré que le Random-Forest donne de meilleurs résultats pour la classification de l'émotion dominante et l'orientation de but de joueur dans une scène de jeu.

Notre **troisième contribution** a consisté à réaliser une interface de conception de règles d'adaptations selon l'état émotionnel du joueur intégrable dans n'importe quel jeu en Unity. Cette interface, intitule « BARGAIN » permet la représentation des états internes de l'élément de jeu (par exemple les états internes de l'élément musique de fond sont « HappyMusic », « SadMusic » et « NeutralMusic ») et de changer l'état actuel d'un élément de jeu selon l'état émotionnel courant du joueur. Nous avons intégré notre Framework d'adaptation émotionnelle dans un jeu de réalité virtuelle pour le développement socio-morale, nommé « Dilemmas_VR ». Avec BARGAIN, nous avons créé plusieurs règles d'adaptation selon l'état émotionnel du joueur comme adapter la musique de fond dans le jeu, les expressions faciales des personnages non-joueur et leurs expression non verbales gestuelles. Pour étudier l'impact de ces règles d'adaptation, nous avons recueilli, lors de l'activation de chaque règle, les mesures d'émotions provenant de notre système de reconnaissance des expressions faciales basées sur les signaux cérébraux NeuroExpress. Pour cette étude, nous avons modifié l'algorithme de NeuroExpress pour fournir outre les mesures affectives (expressions faciales), des mesures cognitives (index d'engagement, la FAA et la distraction). Pour ce faire, une étude expérimentale a été réalisée sur 30 participants de l'Université de Montréal. Cette étude a duré 1 mois et une semaine. Les résultats préliminaires avec l'expérimentation de cet environnement ont montré que les mesures affectives et cognitives peuvent être un indicateur de l'expérience de jeu, comme le montre la corrélation avec les réponses subjectives au questionnaire et les événements du jeu. De plus, l'effet des règles d'adaptation n'apparait statistiquement significatif que sur les mesures affectives du joueur mais ne le sont pas sur les mesures cognitives.

10.2 Travaux futurs

L'objectif était de construire un environnement virtuel émotionnellement intelligent. Tout d'abord, nous avons analysé les états émotionnels et mentaux dans deux types d'environnements : les environnements de débats en ligne et les environnements de jeux vidéo. Ensuite, nous avons proposé une méthode de reconnaissance des réactions émotionnelles à partir des signaux cérébraux (EEG) utilisable dans les environnements de réalité virtuelle. Enfin, nous avons mis au point une interface de définition et d'application de règles d'adaptation émotionnelle dans les jeux vidéo. Nous avons expérimenté et évalué l'impact de règles d’adaptation dans un jeu de réalité virtuelle. Les résultats de nos travaux ont des implications sur le design des jeux vidéo, 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 optimiser son expérience d'interaction. Comme conseil général pour les concepteurs d’interfaces et d’environnement virtuels, il faut toujours trouver de nouveaux moyens (des mécaniques d’interaction) pour varier les effets sonores et visuels afin d’avoir une expérience engageante et plus riche.

Nos travaux présentent cependant quelques limitations qui laissent remarquer de nombreuses perspectives de recherche. En voici quelques exemples :

- **Étendre la base de données émotionnelle du système « Emograph »** : Notre base de données émotionnelle contient un nombre restreint de participants pour le jeu Outlast. Pour améliorer les résultats des algorithmes d'apprentissage machine et généraliser les résultats trouvés, nous comptons ajouter d'autres participants dans les expérimentations et recueillir d'avantage de mesures (suivi oculaire, expressions faciales et EEG). Ces mesures permettront de tester et de valider les résultats trouvés par les algorithmes d'apprentissage machine.
- **Tester la fiabilité des mesures et des algorithmes sur d'autres environnements de jeu vidéo**: Pour que les résultats d'apprentissage machine soient généraux et valides pour des environnements de jeux vidéo, il serait utile d'expérimenter d'autres environnements de différents genres de jeux et recueillir de nouvelles données. Par la suite, nous mènerons des expérimentations sur les environnements développés et validerons nos hypothèses et résultats par rapport à l’émotion dominante et la motivation dans les scènes de jeu.
- **Estimer la durée minimale d’un état émotionnel par apprentissage machine**: dans la composante Biometrics du Framework BARGAIN, nous avons défini qu’un état émotionnel minimal ayant un seuil de 30sec pour recalculer la moyenne des mesures et identifier un nouvel état émotionnel. Cette méthode est acceptable mais n'est pas assez intelligente puisque ça peut changer d’une personne à un autre. Nous proposons plutôt de tester un jeu émotionnellement adaptatif avec différents seuils de calcul de l’état émotionnel et d’entrainer des algorithmes d'apprentissage machine qui permettront de

prédire le seuil de calcul de l'état émotionnel selon les caractéristiques du joueur. Cette méthode serait plus flexible et intelligente que la méthode actuelle.

- **Adapter les jeux selon les états cognitifs d'engagement et de motivation** : le Framework BARGAIN utilise seulement les mesures affectives dans l'estimation de l'état émotionnel. Nous pouvons étendre les règles d'adaptation selon l'état cognitif en considérant les mesures cognitives telles que l'engagement, la motivation et l'attention.
- **Intégrer des techniques de planification automatique** : pour rendre les éléments émotionnellement adaptatifs du jeu capables de planifier tout seuls leurs actions et de prendre des décisions selon l'état du joueur et l'état de l'environnement, nous pouvons intégrer des méthodes IA de planification automatique. Un tel outil utilisera des planificateurs hiérarchiques (HTN hierarchical task network) tels que SHOP2 (Nau et al., 2001) et TLPlan (Bacchus & Kabanza, 1995) avec une définition du problème en langage PDDL (Planning Domain Definition Language (McDermott et al., 1998)).

Bibliographie

Abdessalem, Hamdi Ben, & Frasson, Claude. (2017). *Real-time Brain Assessment for Adaptive Virtual Reality Game: A Neurofeedback Approach.* Paper presented at the International Conference on Brain Function Assessment in Learning.

Aguirre, Héctor Rafael Orozco, Corchado, Félix Francisco Ramos, Ruvalcaba, Luis Alfonso Razo, Rios, Jaime Alberto Zaragoza, & Thalmann, Daniel. (2008). *Use of Intelligent Emotional Agents in the Animation of Autonomous Virtual Creatures.* Paper presented at the Artificial Intelligence, 2008. MICAI'08. Seventh Mexican International Conference on.

Ahmad, Ibrahim, Hamid, Erman, Abdullasim, Nazreen, & Jaafar, Azizah. (2017). *Game Interface Design: Measuring the player's gameplay experience.* Paper presented at the International Visual Informatics Conference.

AlZoubi, Omar, Calvo, Rafael A, & Stevens, Ronald H. (2009). Classification of EEG for affect recognition: an adaptive approach *AI 2009: Advances in Artificial Intelligence* (pp. 52-61): Springer.

Amodio, David M, Master, Sarah L, Yee, Cindy M, & Taylor, Shelley E. (2008). Neurocognitive components of the behavioral inhibition and activation systems: Implications for theories of self-regulation. *Psychophysiology, 45*(1), 11-19.

Andrade, Gustavo, Ramalho, Geber, Santana, Hugo, & Corruble, Vincent. (2005). *Extending reinforcement learning to provide dynamic game balancing.* Paper presented at the Proceedings of the Workshop on Reasoning, Representation, and Learning in Computer Games, 19th International Joint Conference on Artificial Intelligence (IJCAI).

Arellano, Daniel Gábana, Tokarchuk, Laurissa, & Gunes, Hatice. (2016). *Measuring affective, physiological and behavioural differences in solo, competitive and collaborative games.* Paper presented at the International Conference on Intelligent Technologies for Interactive Entertainment.

Avram, Julia, Balteş, Felicia Rodica, Miclea, Mircea, & Miu, Andrei C. (2010). Frontal EEG activation asymmetry reflects cognitive biases in anxiety: evidence from an emotional face stroop task. *Applied psychophysiology and biofeedback, 35*(4), 285-292.

Bacchus, Fahiem, & Kabanza, Froduald. (1995). *Using temporal logic to control search in a forward chaining planner.* Paper presented at the Proceedings of the 3rd European Workshop on Planning.

Barrett, Lisa Feldman. (2017). *How emotions are made: The secret life of the brain*: Houghton Mifflin Harcourt.

Barrett, Lisa Feldman, Gendron, Maria, & Huang, Yang-Ming. (2009). Do discrete emotions exist? *Philosophical Psychology, 22*(4), 427-437.

Bekinschtein, T, Niklison, J, Sigman, L, Manes, FRLJA, Leiguarda, R, Armony, J, . . . Olmos, L. (2004). Emotion processing in the minimally conscious state. *Journal of Neurology, Neurosurgery & Psychiatry, 75*(5), 788-788.

Benlamine, M. , Chaouachi, M. , Frasson, C. , & Dufresne, A. . (2016). Physiology-based Recognition of Facial Micro-expressions using EEG and Identification of the Relevant Sensors by Emotion. *Proceedings of the 3rd International Conference on Physiological Computing Systems - Volume 1: PhyCS.*

Benlamine, Mohamed S, Villata, Serena, Ghali, Ramla, Frasson, Claude, Gandon, Fabien, & Cabrio, Elena. (2017). *Persuasive Argumentation and Emotions: An Empirical Evaluation with Users.* Paper presented at the International Conference on Human-Computer Interaction.

Benlamine, Mohamed Sahbi , Chaouachi, Maher , Frasson, Claude , & Dufresne, Aude. (2016). *Predicting Spontaneous Facial Expressions from EEG.* Paper presented at the Intelligent Tutoring Systems.

Benlamine, Sahbi, Chaouachi, Maher, Villata, Serena, Cabrio, Elena, Frasson, Claude, & Gandon, Fabien. (2015). *Emotions in Argumentation: an Empirical Evaluation.* Paper presented at the IJCAI, IJCAI.

Berka, Chris, Levendowski, Daniel J, Cvetinovic, Milenko M, Petrovic, Miroslav M, Davis, Gene, Lumicao, Michelle N, . . . Olmstead, Richard. (2004). Real-time analysis of EEG indexes of alertness, cognition, and memory acquired with a wireless EEG headset. *International Journal of Human-Computer Interaction, 17*(2), 151-170.

Bernhaupt, Regina, Eckschlager, Manfred, & Tscheligi, Manfred. (2007). *Methods for evaluating games: how to measure usability and user experience in games?* Paper presented at the Proceedings of the international conference on Advances in computer entertainment technology.

Böckle, Martin, Micheel, Isabel, Bick, Markus, & Novak, Jasminko. (2018). *A Design Framework for Adaptive Gamification Applications.* Paper presented at the Proceedings of the 51st Hawaii International Conference on System Sciences.

Bontchev, Boyan. (2016). Adaptation in affective video games: A literature review. *Cybernetics and Information Technologies, 16*(3), 3-34.

Bradley, Margaret M, & Lang, Peter J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry, 25*(1), 49-59.

Breton, Philippe. (2006). *L'argumentation dans la communication*: La découverte.

Broekens, Joost, & Brinkman, Willem-Paul. (2013). AffectButton: A method for reliable and valid affective self-report. *International Journal of Human-Computer Studies, 71*(6), 641-667.

Cabrio, Elena, & Villata, Serena. (2013). A natural language bipolar argumentation approach to support users in online debate interactions. *Argument & Computation, 4*(3), 209-230.

Cacioppo, John T, Cacioppo, Stephanie, & Petty, Richard E. (2018). The neuroscience of persuasion: A review with an emphasis on issues and opportunities. *Social neuroscience, 13*(2), 129-172.

Calvo, Rafael A, & D'Mello, Sidney. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on affective computing, 1*(1), 18-37.

Carletta, Jean. (1996). Assessing agreement on classification tasks: the kappa statistic. *Computational linguistics, 22*(2), 249-254.

Carofiglio, Valeria, & de Rosis, F d. (2003). *Combining logical with emotional reasoning in natural argumentation.* Paper presented at the 3rd Workshop on Affective and Attitude User Modeling.

Cayrol, Claudette, & Lagasquie-Schiex, Marie-Christine. (2013). Bipolarity in argumentation graphs: Towards a better understanding. *International Journal of Approximate Reasoning, 54*(7), 876-899.

Cerutti, Federico, Tintarev, Nava, & Oren, Nir. (2014). *Formal Arguments, Preferences, and Natural Language Interfaces to Humans: an Empirical Evaluation.* Paper presented at the ECAI.

Chanel, Guillaume, Rebetez, Cyril, Bétrancourt, Mireille, & Pun, Thierry. (2008). *Boredom, engagement and anxiety as indicators for adaptation to difficulty in games.* Paper presented at the Proceedings of the 12th international conference on Entertainment and media in the ubiquitous era.

Chaouachi, Maher, Chalfoun, Pierre, Jraidi, Imène, & Frasson, Claude. (2010a). *Affect and mental engagement: towards adaptability for intelligent systems.* Paper presented at the Proceedings of the 23rd International FLAIRS Conference.

Chaouachi, Maher, Chalfoun, Pierre, Jraidi, Imène, & Frasson, Claude. (2010b). *Affect and mental engagement: towards adaptability for intelligent systems.* Paper presented at the Proceedings of the 23rd International FLAIRS Conference, Daytona Beach, FL. http://citeseerx. ist. psu. edu/viewdoc/download.

Chaouachi, Maher, & Frasson, Claude. (2012a). *Mental workload, engagement and emotions: an exploratory study for intelligent tutoring systems.* Paper presented at the Intelligent Tutoring Systems.

Chaouachi, Maher, & Frasson, Claude. (2012b). *Mental workload, engagement and emotions: an exploratory study for intelligent tutoring systems.* Paper presented at the International Conference on Intelligent Tutoring Systems.

Chaouachi, Maher, Jraidi, Imène, & Frasson, Claude. (2015a). MENTOR: A Physiologically Controlled Tutoring System *User Modeling, Adaptation and Personalization* (pp. 56-67): Springer.

Chaouachi, Maher, Jraidi, Imène, & Frasson, Claude. (2015b). *MENTOR: a physiologically controlled tutoring system.* Paper presented at the International Conference on User Modeling, Adaptation, and Personalization.

Chi, Yu Mike, Wang, Yu-Te, Wang, Yijun, Maier, Christoph, Jung, Tzyy-Ping, & Cauwenberghs, Gert. (2012). Dry and noncontact EEG sensors for mobile brain–computer interfaces. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on, 20*(2), 228-235.

Chiasson, V, Vera-Estay, E, Lalonde, G, Dooley, JJ, & Beauchamp, MH. (2017). Assessing social cognition: age-related changes in moral reasoning in childhood and adolescence. *The Clinical Neuropsychologist, 31*(3), 515-530.

Christy, Thomas, & Kuncheva, Ludmila I. (2018). Technological advancements in affective gaming: A historical survey. *GSTF Journal on Computing (JoC), 3*(4).

Coan, James A, & Allen, John JB. (2003). Frontal EEG asymmetry and the behavioral activation and inhibition systems. *Psychophysiology, 40*(1), 106-114.

Conati, Cristina, Jaques, Natasha, & Muir, Mary. (2013). Understanding attention to adaptive hints in educational games: an eye-tracking study. *International Journal of Artificial Intelligence in Education, 23*(1-4), 136-161.

Conroy, David E. (2001). Progress in the development of a multidimensional measure of fear of failure: The Performance Failure Appraisal Inventory (PFAI). *Anxiety, Stress and Coping, 14*(4), 431-452.

Cosmides, Leda, & Tooby, John. (1991). Reasoning and natural selection. *Encyclopedia of human biology, 6*, 493-503.

D'Mello, Sidney, Lehman, Blair, Pekrun, Reinhard, & Graesser, Art. (2014). Confusion can be beneficial for learning. *Learning and Instruction, 29*, 153-170.

Damasio, Antonio R. (2002). Descartes' Error: Emotion, Reason and the Human Brain. *Bulletin of the American Meteorological Society, 83*(5), 742.

Davidson, Richard J. (2001). The neural circuitry of emotion and affective style: Prefrontal cortex and amygdala contributions. *Social Science Information, 40*(1), 11-37.

Davidson, Richard J. (2004). What does the prefrontal cortex "do" in affect: perspectives on frontal EEG asymmetry research. *Biological psychology, 67*(1), 219-234.

de Byl, Penny. (2015). A conceptual affective design framework for the use of emotions in computer game design. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace, 9*(3).

Dekker, Andrew, & Champion, Erik. (2007). *Please Biofeed the Zombies: Enhancing the Gameplay and Display of a Horror Game Using Biofeedback.* Paper presented at the DiGRA Conference.

Denis, Guillaume. (2006). *Jeux vidéo éducatifs et motivation: application à l'enseignement du jazz.* Paris, ENMP.

Derbali, Lotfi, & Frasson, Claude. (2010). *Prediction of players motivational states using electrophysiological measures during serious game play.* Paper presented at the Advanced Learning Technologies (ICALT), 2010 IEEE 10th International Conference on.

Derbali, Lotfi, Ghali, Ramla, & Frasson, Claude. (2013). *Assessing Motivational Strategies in Serious Games Using Hidden Markov Models.* Paper presented at the The Twenty-Sixth International FLAIRS Conference.

DeShon, Richard P, & Gillespie, Jennifer Z. (2005). A motivated action theory account of goal orientation. *Journal of Applied Psychology, 90*(6), 1096.

DeSteno, David, Petty, Richard E, Rucker, Derek D, Wegener, Duane T, & Braverman, Julia. (2004). Discrete emotions and persuasion: the role of emotion-induced expectancies. *Journal of personality and social psychology, 86*(1), 43.

Diamond, David M, Campbell, Adam M, Park, Collin R, Halonen, Joshua, & Zoladz, Phillip R. (2007). The temporal dynamics model of emotional memory processing: a synthesis on the neurobiological basis of stress-induced amnesia, flashbulb and traumatic memories, and the Yerkes-Dodson law. *Neural plasticity, 2007*.

Driver, Matt. (2011). *Coaching Positively: Lessons For Coaches From Positive Psychology: Lessons for Coaches from Positive Psychology*: McGraw-Hill Education (UK).

Dung, Phan Minh. (1995). On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial intelligence, 77*(2), 321-357.

Ekman, Paul. (2005). Basic Emotions, pages 45--60: John Wiley & Sons, Ltd.

Ekman, Paul. (2007). *Emotions revealed: Recognizing faces and feelings to improve communication and emotional life*: Macmillan.

Ekman, Paul, & Rosenberg, Erika L. (1997). *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*: Oxford University Press, USA.

Elliot, Andrew J. (1999). Approach and avoidance motivation and achievement goals. *Educational psychologist, 34*(3), 169-189.

Elliot, Andrew J, & Covington, Martin V. (2001). Approach and avoidance motivation. *Educational Psychology Review, 13*(2), 73-92.

Elliot, Andrew J, McGregor, Holly A, & Gable, Shelly. (1999). Achievement goals, study strategies, and exam performance: A mediational analysis. *Journal of educational psychology, 91*(3), 549.

Elliot, Andrew J, & Murayama, Kou. (2008). On the measurement of achievement goals: Critique, illustration, and application. *Journal of Educational Psychology, 100*(3), 613.

Elliott, Elaine S, & Dweck, Carol S. (1988). Goals: An approach to motivation and achievement. *Journal of personality and social psychology, 54*(1), 5.

Emotient. (2015). Emotient Accuracy Measures and Methodology. Retrieved from Emotient.com website: http://www.emotient.com/wp-content/uploads/Emotient-Accuracy-Methods-May-2015.pdf

Facet, iMotions. (2013). Attention Tool FACET Module Guide Retrieved from imotions.com website: https://imotions.com/wp-content/uploads/2013/08/130806_FACET_Guide.pdf

Fairclough, Stephen H. (2010). Physiological computing: interfacing with the human nervous system *Sensing emotions* (pp. 1-20): Springer.

Freeman, Frederick G, Mikulka, Peter J, Prinzel, Lawrence J, & Scerbo, Mark W. (1999). Evaluation of an adaptive automation system using three EEG indices with a visual tracking task. *Biological psychology, 50*(1), 61-76.

Freeman, Frederick G, Mikulka, Peter J, Scerbo, Mark W, Prinzel, Lawrence J, & Clouatre, Keith. (2000). Evaluation of a psychophysiologically controlled adaptive automation system, using performance on a tracking task. *Applied Psychophysiology and Biofeedback, 25*(2), 103-115.

Frijda, Nico H. (1986). *The emotions*: Cambridge University Press.

Fulkerson, Richard. (1993). The Place of Emotion in Argument: JSTOR.

Garner, Tom Alexander. (2013). Game Sound from Behind the Sofa: An Exploration into the Fear Potential of Sound & Psychophysiological Approaches to Audio-centric, Adaptive Gameplay.

Gass, Robert H, & Seiter, John S. (2015). *Persuasion: Social influence and compliance gaining*: Routledge.

Ghali, Ramla, Frasson, Claude, & Ouellet, Sébastien. (2016). *Using Electroencephalogram to Track Learner's Reasoning in Serious Games.* Paper presented at the International Conference on Intelligent Tutoring Systems.

Gibbs, John C. (2013). *Moral development and reality: Beyond the theories of Kohlberg, Hoffman, and Haidt*: Oxford University Press.

Gilbert, Michael A. (1995). *Emotional argumentation, or, why do argumentation theorists argue with their mates.* Paper presented at the Analysis and evaluation: Proceedings of the third ISSA conference on argumentation.

Goldberg, Lewis R. (1992). The development of markers for the Big-Five factor structure. *Psychological assessment, 4*(1), 26.

Goleman, Daniel. (1995). Emotional intelligence: Why it can matter more than IQ: New York: Bantam Books.

Goleman, Daniel. (2006). *Emotional intelligence*: Bantam.

Gratch, Jonathan, Marsella, Stacy, & Petta, Paolo. (2009). Modeling the cognitive antecedents and consequences of emotion. *Cognitive Systems Research, 10*(1), 1-5.

Greenhouse, Samuel W, & Geisser, Seymour. (1959). On methods in the analysis of profile data. *Psychometrika, 24*(2), 95-112.

Hagemann, Dirk, Naumann, Ewald, Thayer, Julian F, & Bartussek, Dieter. (2002). Does resting electroencephalograph asymmetry reflect a trait? an application of latent state-trait theory. *Journal of personality and social psychology, 82*(4), 619.

Hall, Mark, Frank, Eibe, Holmes, Geoffrey, Pfahringer, Bernhard, Reutemann, Peter, & Witten, Ian H. (2009). The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter, 11*(1), 10-18.

Hamari, Juho, Shernoff, David J, Rowe, Elizabeth, Coller, Brianno, Asbell-Clarke, Jodi, & Edwards, Teon. (2016). Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. *Computers in Human Behavior, 54*, 170-179.

Harley, Jason M, & Azevedo, Roger. (2014). Toward a Feature-Driven Understanding of Students' Emotions during Interactions with Agent-Based Learning Environments: A Selective Review. *International Journal of Gaming and Computer-Mediated Simulations (IJGCMS), 6*(3), 17-34.

Harley, Jason M, Bouchet, François, & Azevedo, Roger. (2013). *Aligning and comparing data on emotions experienced during learning with MetaTutor.* Paper presented at the International Conference on Artificial Intelligence in Education.

Harley, Jason M, Bouchet, François, Hussain, M Sazzad, Azevedo, Roger, & Calvo, Rafael. (2015). A multi-componential analysis of emotions during complex learning with an intelligent multi-agent system. *Computers in Human Behavior, 48*, 615-625.

Harley, Jason Matthew. (2016). Measuring emotions: a survey of cutting edge methodologies used in computer-based learning environment research *Emotions, technology, design, and learning* (pp. 89-114): Elsevier.

Harmon-Jones, Eddie, Gable, Philip A, & Peterson, Carly K. (2010). The role of asymmetric frontal cortical activity in emotion-related phenomena: A review and update. *Biological psychology, 84*(3), 451-462.

Harmon-Jones, Eddie, & Gable, Philip A. (2018). On the role of asymmetric frontal cortical activity in approach and withdrawal motivation: An updated review of the evidence. *Psychophysiology, 55*(1), e12879.

Hemenover, Scott H, & Bowman, Nicholas D. (2018). Video games, emotion, and emotion regulation: expanding the scope. *Annals of the International Communication Association*, 1-19.

Heraz, Alicia, & Frasson, Claude. (2007). Predicting the three major dimensions of the learner's emotions from brainwaves. *World Academy of Science, Engineering and Technology, 31*, 323-329.

Hershkovitz, Arnon, & Nachmias, Rafi. (2008). *Developing a Log-based Motivation Measuring Tool.* Paper presented at the EDM.

Hocine, Nadia, Gouaïch, Abdelkader, Cerri, Stefano A, Mottet, Denis, Froger, Jérome, & Laffont, Isabelle. (2015). Adaptation in serious games for upper-limb rehabilitation: an approach to improve training outcomes. *User Modeling and User-Adapted Interaction, 25*(1), 65-98.

Horan, William P, Wynn, Jonathan K, Mathis, Ian, Miller, Gregory A, & Green, Michael F. (2014). Approach and withdrawal motivation in schizophrenia: an examination of frontal brain asymmetric activity. *PloS one, 9*(10), e110007.

Huk, Alexander, & Meister, Miriam LR. (2012). Neural correlates and neural computations in posterior parietal cortex during perceptual decision-making. *Frontiers in integrative neuroscience, 6*, 86.

Hull, Clark L. (1935). The conflicting psychologies of learning—a way out. *Psychological Review, 42*(6), 491.

Hunicke, Robin, & Chapman, Vernell. (2004). AI for Dynamic Difficulty Adjustment in Games. 2004. *Association for the Advancement of Artificial Intelligence (AAAI)*, 2-5.

Hunicke, Robin, LeBlanc, Marc, & Zubek, Robert. (2004). *MDA: A formal approach to game design and game research.* Paper presented at the Proceedings of the AAAI Workshop on Challenges in Game AI.

Hunter, Anthony. (2016). *Computational Persuasion with Applications in Behaviour Change.* Paper presented at the COMMA.

i Badia, Sergi Bermudez, Quintero, Luis Velez, Cameirao, Monica S, Chirico, Alice, Triberti, Stefano, Cipresso, Pietro, & Gaggioli, Andrea. (2018). Towards Emotionally-Adaptive Virtual Reality for Mental Health Applications. *IEEE journal of biomedical and health informatics*.

IJsselsteijn, WA, De Kort, YAW, & Poels, Karolien. (2013). The game experience questionnaire.

Isbister, Katherine, & Schaffer, Noah. (2008). *Game usability: Advancing the player experience*: CRC Press.

Jennett, Charlene, Cox, Anna L, Cairns, Paul, Dhoparee, Samira, Epps, Andrew, Tijs, Tim, & Walton, Alison. (2008). Measuring and defining the experience of immersion in games. *International journal of human-computer studies, 66*(9), 641-661.

John, Oliver P, & Srivastava, Sanjay. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research, 2*(1999), 102-138.

Jraidi, Imène, Chaouachi, Maher, & Frasson, Claude. (2013). *A dynamic multimodal approach for assessing learners' interaction experience.* Paper presented at the Proceedings of the 15th ACM on International conference on multimodal interaction.

Kaplan, Avi, & Maehr, Martin L. (2007). The contributions and prospects of goal orientation theory. *Educational psychology review, 19*(2), 141-184.

Kassam, K. S., Markey, A. R., Cherkassky, V. L., Loewenstein, G., & Just, M. A. (2013). Identifying Emotions on the Basis of Neural Activation. *Plos One, 8*(6). doi: 10.1371/journal.pone.0066032

Kellermann, Kathy, Broetzmann, Scott, Lim, Tae-Seop, & Kitao, Kenji. (1989). The conversation MOP: Scenes in the stream of discourse. *Discourse Processes, 12*(1), 27-61.

König, Alexandra, Grippenkoven, Jan, Wegener, Jan, & Pelz, Albert. (2017). SERIOUS GAMES: A PLAYFUL APPROACH TO REDUCE USAGE BARRIERS OF INNOVATIVE PUBLIC TRANSPORT SYSTEMS.

Kors, Martijn JL, Ferri, Gabriele, van der Spek, Erik D, Ketel, Cas, & Schouten, Ben AM. (2016). *A Breathtaking Journey.: On the Design of an Empathy-Arousing Mixed-Reality Game.* Paper presented at the Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play.

Koster, Raph. (2013). *Theory of fun for game design*: " O'Reilly Media, Inc.".

Kosunen, Ilkka, Salminen, Mikko, Järvelä, Simo, Ruonala, Antti, Ravaja, Niklas, & Jacucci, Giulio. (2016). *RelaWorld: neuroadaptive and immersive virtual reality meditation system.* Paper presented at the Proceedings of the 21st International Conference on Intelligent User Interfaces.

Lang, Peter J, Bradley, Margaret M, & Cuthbert, Bruce N. (1990). Emotion, attention, and the startle reflex. *Psychological review, 97*(3), 377.

Lang, Peter J, Bradley, Margaret M, & Cuthbert, Bruce N. (2008). International affective picture system (IAPS): Affective ratings of pictures and instruction manual. *Technical report A-8*.

Lara-Cabrera, Raúl, & Camacho, David. (2019). A taxonomy and state of the art revision on affective games. *Future Generation Computer Systems, 92*, 516-525.

Lazarus, Richard S. (1991). Emotion and adaptation. 1991. *Cité en*, 9.

Lazzaro, Nicole. (2004). Why we play games: Four keys to more emotion without story.

LeDoux, Joseph E. (1992). Brain mechanisms of emotion and emotional learning. *Current opinion in neurobiology, 2*(2), 191-197.

LeDoux, Joseph E. (2017). Semantics, surplus meaning, and the science of fear. *Trends in cognitive sciences, 21*(5), 303-306.

Lee, You-Yun, & Hsieh, Shulan. (2014). Classifying different emotional states by means of EEG-based functional connectivity patterns. *PloS one, 9*(4), e95415.

Lepper, Mark R, & Malone, Thomas W. (1987). Intrinsic motivation and instructional effectiveness in computer-based education. *Aptitude, learning, and instruction, 3*, 255-286.

Lewinski, Peter, den Uyl, Tim M, & Butler, Crystal. (2014). Automated facial coding: Validation of basic emotions and FACS AUs in FaceReader. *Journal of Neuroscience, Psychology, and Economics, 7*(4), 227.

Libet, Benjamin. (2006). Reflections on the interaction of the mind and brain. *Progress in neurobiology, 78*(3), 322-326.

Lin, Jih-Hsuan Tammy, Wu, Dai-Yun, & Tao, Chen-Chao. (2017). So scary, yet so fun: The role of self-efficacy in enjoyment of a virtual reality horror game. *New Media & Society*, 1461444817744850.

Littlewort, Gwen, Whitehill, Jacob, Wu, Tingfan, Fasel, Ian, Frank, Mark, Movellan, Javier, & Bartlett, Marian. (2011). *The computer expression recognition toolbox (CERT).* Paper presented at the Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on.

Liu, Changchun, Agrawal, Pramila, Sarkar, Nilanjan, & Chen, Shuo. (2009). Dynamic difficulty adjustment in computer games through real-time anxiety-based affective feedback. *International Journal of Human-Computer Interaction, 25*(6), 506-529.

Liu, Yisi, Sourina, Olga, & Nguyen, Minh Khoa. (2011). Real-time EEG-based emotion recognition and its applications *Transactions on computational science XII* (pp. 256-277): Springer.

Lloyd-Kelly, Martyn, & Wyner, Adam. (2011). *Arguing about emotion.* Paper presented at the International Conference on User Modeling, Adaptation, and Personalization.

Lough, Sinclair, Kipps, Christopher M, Treise, Cate, Watson, Peter, Blair, James R, & Hodges, John R. (2006). Social reasoning, emotion and empathy in frontotemporal dementia. *Neuropsychologia, 44*(6), 950-958.

Lu, Yifei, Zheng, Wei-Long, Li, Binbin, & Lu, Bao-Liang. (2015). *Combining eye movements and EEG to enhance emotion recognition.* Paper presented at the International Joint Conference on Artificial Intelligence (IJCAI).

Lubar, Joel F. (1991). Discourse on the development of EEG diagnostics and biofeedback for attention-deficit/hyperactivity disorders. *Biofeedback and Self-regulation, 16*(3), 201-225.

Lynch, Teresa, & Martins, Nicole. (2015). Nothing to fear? An analysis of college students' fear experiences with video games. *Journal of Broadcasting & Electronic Media, 59*(2), 298-317.

Malone, Thomas W. (1981). Toward a theory of intrinsically motivating instruction*. *Cognitive science, 5*(4), 333-369.

Malone, Thomas W, & Lepper, Mark R. (1987). Making learning fun: A taxonomy of intrinsic motivations for learning. *Aptitude, learning, and instruction, 3*(1987), 223-253.

Martinez, DC, & Simari, GR. (2012). Emotion-directed argument awareness for autonomous agent reasoning. *Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial, 15*(50).

Maslow, Abraham Harold. (1943). A theory of human motivation. *Psychological review, 50*(4), 370.

Mauchly, John W. (1940). Significance Test for Sphericity of a Normal n-Variate Distribution. 204-209. doi: 10.1214/aoms/1177731915

Mayer, John D, & Geher, Glenn. (1996). Emotional intelligence and the identification of emotion. *Intelligence, 22*(2), 89-113.

McClelland, David C, Atkinson, John W, Clark, Russell A, & Lowell, Edgar L. (1976). The achievement motive.

McClelland, David Clarence. (1984). *Motives, personality, and society: Selected papers*: Praeger Publishers.

McDermott, Drew, Ghallab, Malik, Howe, Adele, Knoblock, Craig, Ram, Ashwin, Veloso, Manuela, . . . Wilkins, David. (1998). PDDL-the planning domain definition language.

Mehrabian, Albert, & Russell, James A. (1974). *An approach to environmental psychology*: the MIT Press.

Mihaly, Csikszentmihalyi. (1990). The psychology of optimal experience. *Harper&Row, New York*.

Murray, Henry Alexander. (1938). Explorations in personality.

Nacke, Lennart Erik, Kalyn, Michael, Lough, Calvin, & Mandryk, Regan Lee. (2011). *Biofeedback game design: using direct and indirect physiological control to enhance game interaction.* Paper presented at the Proceedings of the SIGCHI conference on human factors in computing systems.

Nacke, Lennart, & Lindley, Craig A. (2008). *Flow and immersion in first-person shooters: measuring the player's gameplay experience.* Paper presented at the Proceedings of the 2008 Conference on Future Play: Research, Play, Share.

Nakamura, Jeanne, & Csikszentmihalyi, Mihaly. (2002). The concept of flow. *Handbook of positive psychology*, 89-105.

Nau, Dana, Munoz-Avila, Héctor, Cao, Yue, Lotem, Amnon, & Mitchell, Steven. (2001). *Total-order planning with partially ordered subtasks.* Paper presented at the IJCAI.

Nawwab, Fahd Saud, Dunne, Paul E, & Bench-Capon, Trevor JM. (2010). *Exploring the Role of Emotions in Rational Decision Making.* Paper presented at the COMMA.

Nogueira, Pedro Alves, Rodrigues, Rui Amaral, Oliveira, Eugénio C, & Nacke, Lennart E. (2013). *Guided Emotional State Regulation: Understanding and Shaping Players' Affective Experiences in Digital Games.* Paper presented at the AIIDE.

Ortony, A , Clore, A, & Collins, GL. (1988). The cognitive structure of emotions: Cambridge University Press.

Ortony, Andrew, Clore, Gerald L, & Collins, Allan. (1990). *The cognitive structure of emotions*: Cambridge university press.

Phan, K Luan, Wager, Tor, Taylor, Stephan F, & Liberzon, Israel. (2002). Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in PET and fMRI. *Neuroimage, 16*(2), 331-348.

Picard, Rosalind W. (1997). *Affective computing* (Vol. 252): MIT press Cambridge.

Plutchik, Robert. (1982). A psychoevolutionary theory of emotions: Sage Publications.

Pope, Alan T, Bogart, Edward H, & Bartolome, Debbie S. (1995a). Biocybernetic system evaluates indices of operator engagement in automated task. *Biological psychology, 40*(1), 187-195.

Pope, Alan T, Bogart, Edward H, & Bartolome, Debbie S. (1995b). Biocybernetic system evaluates indices of operator engagement in automated task. *Biological psychology, 40*(1-2), 187-195.

Posner, Jonathan, Russell, James A, & Peterson, Bradley S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and psychopathology, 17*(3), 715-734.

Prendinger, Helmut, Mori, Junichiro, Mayer, Sonja, & Ishizuka, Mitsuru. (2003). Character-based interfaces adapting to users' autonomic nervous system activity.

Prensky, Marc. (2002). The motivation of gameplay: The real twenty-first century learning revolution. *On the horizon, 10*(1), 5-11.

Przybylski, Andrew K, Rigby, C Scott, & Ryan, Richard M. (2010). A motivational model of video game engagement. *Review of general psychology, 14*(2), 154.

Putman, Peter, van Peer, Jacobien, Maimari, Ioulia, & van der Werff, Steven. (2010). EEG theta/beta ratio in relation to fear-modulated response-inhibition, attentional control, and affective traits. *Biological psychology, 83*(2), 73-78.

Quartz, Steven R. (2009). Reason, emotion and decision-making: risk and reward computation with feeling. *Trends in cognitive sciences, 13*(5), 209-215.

Quick, John M, Atkinson, Robert K, & Lin, Lijia. (2012). The gameplay enjoyment model. *International Journal of Gaming and Computer-Mediated Simulations (IJGCMS), 4*(4), 64-80.

Rahwan, Iyad, Madakkatel, Mohammed I, Bonnefon, Jean-François, Awan, Ruqiyabi N, & Abdallah, Sherief. (2010). Behavioral experiments for assessing the abstract argumentation semantics of reinstatement. *Cognitive Science, 34*(8), 1483-1502.

Rahwan, Iyad, & Simari, Guillermo R. (2009). *Argumentation in artificial intelligence* (Vol. 47): Springer.

Roseman, Ira J, Spindel, Martin S, & Jose, Paul E. (1990). Appraisals of emotion-eliciting events: Testing a theory of discrete emotions. *Journal of personality and social psychology, 59*(5), 899.

Rosenfeld, Ariel, & Kraus, Sarit. (2016). *Strategical Argumentative Agent for Human Persuasion.* Paper presented at the ECAI.

Ross, WD. (2010). Rhetoric by Aristotle. *New York: Cosimo Inc*.

Rowe, Jonathan P, Mott, Bradford W, & Lester, James C. (2014). Optimizing Player Experience in Interactive Narrative Planning: A Modular Reinforcement Learning Approach. *AIIDE, 3*, 2.

Rozin, Paul, & Royzman, Edward B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and social psychology review, 5*(4), 296-320.

Russell, James A. (1980). A circumplex model of affect. *Journal of personality and social psychology, 39*(6), 1161.

Ryan, Richard M, & Deci, Edward L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational psychology, 25*(1), 54-67.

Ryan, Richard M, Rigby, C Scott, & Przybylski, Andrew. (2006). The motivational pull of video games: A self-determination theory approach. *Motivation and emotion, 30*(4), 344-360.

Sabourin, Jennifer L, & Lester, James C. (2014). Affect and engagement in Game-BasedLearning environments. *IEEE Transactions on Affective Computing, 5*(1), 45-56.

Salovey, Peter, & Mayer, John D. (1990). Emotional intelligence. *Imagination, cognition and personality, 9*(3), 185-211.

Samsung. (2015). Samsung prototypes brainwave-reading wearable stroke detector. Retrieved from Samsung Tomorrow website: https://news.samsung.com/global/c-lab-engineers-developing-wearable-health-sensor-for-stroke-detection

Sarkheil, Pegah, Goebel, Rainer, Schneider, Frank, & Mathiak, Klaus. (2013). Emotion unfolded by motion: a role for parietal lobe in decoding dynamic facial expressions. *Social cognitive and affective neuroscience, 8*(8), 950-957.

Saxe, Rebecca, & Wexler, Anna. (2005). Making sense of another mind: the role of the right temporo-parietal junction. *Neuropsychologia, 43*(10), 1391-1399.

Scherer, Klaus R. (1986). Vocal affect expression: a review and a model for future research. *Psychological bulletin, 99*(2), 143.

Schoenau-Fog, Henrik. (2011). *The player engagement process–An exploration of continuation desire in digital games.* Paper presented at the Think Design Play: Digital Games Research Conference.

Stemmler, Gerhard. (2004). Physiological processes during emotion. *The regulation of emotion*, 33-70.

Stock, Oliviero, Guerini, Marco, & Pianesi, Fabio. (2016). *Ethical Dilemmas for Adaptive Persuasion Systems.* Paper presented at the AAAI.

Sweetser, Penelope, Johnson, Daniel M, & Wyeth, Peta. (2012). Revisiting the GameFlow model with detailed heuristics. *Journal: Creative Technologies, 2012*(3).

Sweetser, Penelope, & Wyeth, Peta. (2005). GameFlow: a model for evaluating player enjoyment in games. *Computers in Entertainment (CIE), 3*(3), 3-3.

Tato, Ange Adrienne Nyamen, Dufresne, Aude, Nkambou, Roger, Morasse, Frédérick, & Beauchamp, Miriam H. (2018). *Preliminary Evaluation of a Serious Game for Socio-Moral Reasoning.* Paper presented at the Intelligent Tutoring Systems.

Tato, Ange Adrienne Nyamen, Nkambou, Roger, Dufresne, Aude, & Beauchamp, Miriam H. (2017). *Convolutional Neural Network for Automatic Detection of Sociomoral Reasoning Level.* Paper presented at the EDM.

Teigen, Karl Halvor. (1994). Yerkes-Dodson: A law for all seasons. *Theory & Psychology, 4*(4), 525-547.

Teplan, Michal. (2002). Fundamentals of EEG measurement. *Measurement science review, 2*(2), 1-11.

Tomarken, Andrew J, Davidson, Richard J, & Henriques, Jeffrey B. (1990). Resting frontal brain asymmetry predicts affective responses to films. *Journal of personality and social psychology, 59*(4), 791.

Trabelsi, Amine, & Frasson, Claude. (2010). *The emotional machine: A machine learning approach to online prediction of user's emotion and intensity.* Paper presented at the Advanced Learning Technologies (ICALT), 2010 IEEE 10th International Conference on.

Vallerand, Robert J, Blais, Marc R, Brière, Nathalie M, & Pelletier, Luc G. (1989). Construction et validation de l'échelle de motivation en éducation (EME). *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement, 21*(3), 323.

Vallerand, Robert J, & Thill, Edgar E. (1993). Introduction au concept de motivation. *Introduction à la psychologie de la motivation*, 3-39.

Villani, Daniela, Carissoli, Claudia, Triberti, Stefano, Marchetti, Antonella, Gilli, Gabriella, & Riva, Giuseppe. (2018). Videogames for Emotion Regulation: A Systematic Review. *Games for health journal*.

Villata, Serena, Cabrio, Elena, Jraidi, Imène, Benlamine, Sahbi, Chaouachi, Maher, Frasson, Claude, & Gandon, Fabien. (2017). Emotions and personality traits in argumentation: An empirical evaluation 1. *Argument & Computation, 8*(1), 61-87.

Von Stein, Astrid, & Sarnthein, Johannes. (2000). Different frequencies for different scales of cortical integration: from local gamma to long range alpha/theta synchronization. *International journal of psychophysiology, 38*(3), 301-313.

Vuilleumier, Patrik. (2005). How brains beware: neural mechanisms of emotional attention. *Trends in cognitive sciences, 9*(12), 585-594.

Walk, Wolfgang, Görlich, Daniel, & Barrett, Mark. (2017). Design, Dynamics, Experience (DDE): An advancement of the MDA framework for game design *Game Dynamics* (pp. 27-45): Springer.

Winn, Brian M. (2009). The design, play, and experience framework *Handbook of research on effective electronic gaming in education* (pp. 1010-1024): IGI Global.

Wolters, Christopher A. (2004). Advancing Achievement Goal Theory: Using Goal Structures and Goal Orientations to Predict Students' Motivation, Cognition, and Achievement. *Journal of educational psychology, 96*(2), 236.

Yannakakis, Geogios N. (2012). *Game AI revisited.* Paper presented at the Proceedings of the 9th conference on Computing Frontiers.

Yannakakis, Georgios N, & Paiva, Ana. (2014). Emotion in games. *Handbook on affective computing*, 459-471.

Yannakakis, Georgios N, & Togelius, Julian. (2011). Experience-driven procedural content generation. *Affective Computing, IEEE Transactions on, 2*(3), 147-161.

Yee, Nick. (2006). Motivations for play in online games. *CyberPsychology & behavior, 9*(6), 772-775.

Yerkes, Robert M, & Dodson, John D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of comparative neurology and psychology, 18*(5), 459-482.

Zirbel, Esther L. (2014). Learning, concept formation and conceptual change: Tufts University.

Publications

Articles de Journaux

- **Benlamine**, M.S., Dufresne, A., Beauchamp, M., & Frasson, C. (under review 2018). BARGAIN: Behavioral Affective Rule-based Games Adaptation Interface Towards Emotional Intelligent Games: application on a virtual reality environment for socio-moral development. Submitted to the journal *User Modeling and User-Adapted Interaction* at the 21st November 2018, Manuscript number: UMUI-D-18-00119.

- Villata, S., Cabrio, E., Jraidi, I, **Benlamine**, M.S., Chaouachi, M., Frasson, C., & Gandon, F. (2017). Emotions and Personality Traits in Argumentation: an Empirical Evaluation. *Argument & Computation.* [ISSN 1946-2166]

Articles de Conférences

- Doumbouya R., **Benlamine**, M.S., Dufresne, A., & Frasson, C. (2018). Game Scenes Evaluation and Player's Dominant Emotion Prediction. T*he 14th International Conference on Intelligent Tutoring Systems* ***ITS 2018***. Springer International publishing.

- Villata, S., **Benlamine**, M. S., Cabrio, E., Frasson, C., & Gandon, F. (2018). Assessing persuasion in argumentation through emotion and mental states. The *31st Florida Artificial Intelligence Research Society Conference* ***FLAIRS 2018.*** AAAI Press.

- **Benlamine** M.S., Doumbouya R., Frasson C., Dufresne A. (2017). Game Experience and Brain based Assessment of Motivational Goal Orientations in Video Games. The *First International Conference on "Brain Function Assessment in Learning", BFAL 2017*. Springer International publishing.

- **Benlamine**, M. S., Villata, S., Ghali, R., Cabrio, E., Frasson, C., & Gandon, F. (2017). Persuasive Argumentation and Emotions: an Empirical Evaluation with Users. *Proceedings of the Nineteenth international conference on Human-Computer Interaction* ***HCII 2017***. Springer International Publishing.

- **Benlamine** M.S., Chaouachi M., Frasson C., & Dufresne A. (2016). Physiology-based recognition of Facial expressions using EEG and identification of the relevant sensors by emotion. The *3rd International Conference on physiological computing systems **PhyCs 2016***: INSTICC / Springer.
- **Benlamine**, M. S., Chaouachi, M., Villata, S., Cabrio, E., Frasson, C., & Gandon, F. (2015). Emotions in Argumentation: an Empirical Evaluation. *Proceedings of the Twenty- Fourth international joint conference on Artificial Intelligence **IJCAI 2015***. AAAI Press.
- **Benlamine**, M. S., Bouslimi, S., Harley, J. M., Frasson, C., & Dufresne, A. (2015). Brain-based gaming: measuring engagement during gameplay. *Proceedings of the 9th world Conference on Educational Media and technology **EDMEDIA 2015***.

Articles de Workshop

- Harley, J.M., **Benlamine**, M.S., Chaouachi, M., Frasson, C., & Dufresne, A. (2016). Mission accomplished? Measuring gamers' emotion and cognitive engagement during and after a narrative event. Paper accepted *in the 13^{th}International Workshop on Intelligent Tutoring Systems*. Springer International publishing.

Posters

- **Benlamine** M.S., Chaouachi M., Frasson C., & Dufresne A. (2016). Predicting Spontaneous Facial Expressions from EEG. Accepted as Poster in *The 13th International Conference on Intelligent Tutoring Systems **ITS 2016***. Springer International publishing.