



Lutin Game Lab

Dominique Boullier, Audrey Lohard, Antoine Visonneau, Laure Léger, Nicolas Fouquereau, Florent Levillain, Charles Tijus, Sébastien Genvo, Serge Burchardon, Marc Damez, et al.

► **To cite this version:**

Dominique Boullier, Audrey Lohard, Antoine Visonneau, Laure Léger, Nicolas Fouquereau, et al. : Rapport final du projet ANR 2006-2008 (annexes scientifiques). 2009. <halshs-00371331>

HAL Id: halshs-00371331

<https://halshs.archives-ouvertes.fr/halshs-00371331>

Submitted on 30 Mar 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Lutin Game Lab

Projet ANR 06 RIAM 013

Annexes scientifiques au rapport final



Table des matières

SP2 Un expert artificiel pour aider à la décision durant le développement d'un jeu vidéo (Rennes 2)	3
SP3 GameLab : Quick Reality Check (Rennes 2)	33
SP3 : Méthodes en ergonomie des interfaces : oculométrie et méthodes sensorielles	44
SP4 Méthodes d'évaluation en sémiotique du <i>gameplay</i>	133
SP5 Expert artificiel pour le diagnostic	146

NB : les deux autres sous projets relèvent du montage technique (SP1) et de la diffusion auprès des entreprises (SP6) et ne justifient donc pas d'un rapport scientifique

SP2 Un expert artificiel pour aider à la décision durant le développement d'un jeu vidéo

Audrey Lohard, Antoine Visonneau, Dominique Boullier
LAS Université Rennes 2

L'expert artificiel a pour but d'aider à la décision durant le développement d'un jeu vidéo, afin que ce dernier puisse voir ses chances de succès optimisées au moment de sa mise sur le marché. Pour évaluer la qualité des différentes composantes d'un jeu vidéo, les développeurs de jeux vidéo se fondent en général sur leur propre expertise. L'expert artificiel a pour but d'objectiver ces différentes expertises, par type de jeux, pour pouvoir ensuite les simuler et permettre de valider rapidement des choix de développement pendant la phase de conception d'un jeu vidéo.

L'expert artificiel est capable d'apporter trois types de connaissances sur les jeux vidéo. Il aide tout d'abord à comprendre pourquoi les joueurs vont apprécier ou non telle ou telle dimension d'un jeu vidéo. Les joueurs sont souvent dans l'impossibilité d'expliquer eux-mêmes ce qui fait qu'un jeu leur plaît ou leur déplaît. L'expert artificiel va être capable d'expliquer ces jugements en les mettant en relation avec des évaluations réalisées par des experts du domaine. L'expert artificiel peut ensuite utiliser les modèles explicatifs qui permettent de comprendre les jugements des joueurs pour prédire ce que seraient leurs jugements sur un autre jeu. Il agit en quelque sorte comme un testeur virtuel. Les concepteurs de jeux vidéo peuvent donc savoir à l'avance, en fonction du profil défini de leur jeu, comment celui-ci serait jugé par les joueurs. Enfin, l'expert artificiel est également capable de trouver des solutions optimales de conception, en tenant compte des objectifs des concepteurs, mais aussi de leurs contraintes.

Dans le cadre du projet Gamelab, 4 types de jeux ont été sélectionnés : les jeux Action-Aventure, les jeux FPS (*First Person Shooter*), les jeux *casual* et les jeux RTS (*Real Time Strategy*). Nous avons retenus 6 jeux dans chacun de ces genres. A chaque fois, parmi ces 6 jeux, 3 jeux ont été jugés de manière plutôt positive par la critique des revues spécialisées, 3 autres ont été moins bien notés. Cet éventail de jeux est nécessaire pour que l'expert puisse disposer de bases de données faisant référence à de "bons" jeux, mais aussi à de moins bons.

Le rapport suivant présentera

- 1) l'élaboration des bases de données constituées auprès des professionnels du secteur et à partir de tests effectués au laboratoire Lutin,
- 2) le fonctionnement de l'expert artificiel et l'identification des modèles explicatifs et prédictifs robustes,
- 3) l'analyse des modèles et les résultats qu'ils permettent d'obtenir et
- 4) les problèmes rencontrés et les avancées qu'il reste à faire.

1 . L'élaboration des bases de données

Afin que l'expert artificiel dispose de base de données de référence pour pouvoir évaluer des jeux en développement, nous avons dû au préalable l'alimenter avec des évaluations réalisées sur des jeux existants. Notre corpus comportait 6 jeux par genre étudié. Au départ, nous avons donc prévu un corpus total de 42 jeux. L'intégralité des grilles d'évaluation a été remplie finalement pour 3 genres de jeux¹.

Pour permettre à l'expert de créer des modèles explicatifs et prédictifs sur des jeux existants, celui-ci doit être alimenté avec trois types de données :

¹ cf. 4ème partie de ce rapport

- 1) Des données subjectives (données S) qui sont les évaluations faites par des joueurs reçus au laboratoire Lutin pour donner leurs appréciations sur le corpus de jeux Gamelab.
- 2) Des données objectives (données O) qui décrivent certaines caractéristiques objectives des jeux du corpus, comme l'âge minimum requis pour jouer ou encore la présence d'un moteur de physique dans le jeu, etc.
- 3) Des données subjectives-objectivées (données SO), qui correspondent aux évaluations détaillées faites par les professionnels du jeu vidéo sur les propriétés de ce même corpus de jeu (et non une évaluation générale).

Afin de recueillir ces données, nous avons constitué des grilles de description des jeux vidéo et une grille d'entretien avec les testeurs. Les grilles objectives (grille O) et subjectives (grille S) sont identiques, quel que ce soit le genre de jeux traité. En revanche, les grilles subjectives-objectivées (grilles SO), sont adaptées à chacun des genres de jeux².

1.1 . Définir des critères de qualité des jeux vidéo avec les professionnels

1.1.1. Organiser un atelier de travail

Pour chacun des quatre genres de jeux, nous avons organisé un atelier de travail réunissant des professionnels, spécialistes du genre, pour identifier des critères de qualité d'un jeu vidéo susceptibles d'être ensuite évalués et traités par l'expert artificiel. Il s'agissait donc d'identifier des variables caractéristiques du genre en question pour pouvoir décrire finement un jeu et pour pouvoir l'évaluer par rapport aux autres jeux du même genre.

Les game designers représentants des entreprises n'ont pas été difficiles à convaincre car nous avons fait au préalable un travail d'approche : rencontre individuelle de chaque entreprise clé pour présenter le projet et proposer de participer. Nous avons également modifié un des axes d'étude du projet (l'analyse des jeux de simulation a été remplacée par celle des jeux *casual*) pour nous aligner avec ce qui intéresse les entreprises du consortium.

1.1.2. Déroulé de l'atelier

Dans un premier temps, les participants se présentent. Nous exposons ensuite les grandes lignes du projet et l'objectif de la séance.

Phase 1 : brainstorming. Les participants sont mis en confiance et on leur demande de générer un maximum d'idées pour des critères de lecture/analyse/description du type de jeu. Dans cette phase, on privilégie la quantité plutôt que la qualité. L'animateur note les idées, encourage les participants. Après une heure de production de ce type, on décroche toutes les feuilles du paperboard et on les affiche sur le mur du laboratoire.

Phase 2 : on montre la grille réalisée dans un autre atelier pour donner une meilleure idée du résultat final et du type de formulation que l'on recherche. On commente un peu. Ensuite, on reprend la production. On passe en revue chacune des idées affichées au mur et on essaie de trouver la formulation des critères et les plots linguistiques associés.

Un peu avant la fin, on prend du recul pour évaluer la production (faut-il une autre séance ?), faire valider le corpus des jeux qui seront testés et expliquer ce que l'on attend des participants pour la suite des opérations.

1.1.3. Résultat de l'atelier

A la fin de chaque atelier, nous avons obtenu une première version de la grille SO du genre

² Toutes les grilles sont en annexe.

concerné. Pour chacune des variables de la grille, nous avons réfléchi avec les professionnels à une échelle d'évaluation (échelle de Lickert à 7 niveaux). Rappelons que ces variables devront être évaluées de manière subjective-objectivée. Nous avons donc dû définir en commun des plots linguistiques permettant de qualifier les écarts de notation sur l'échelle (avec des exemples de jeux connus des professionnels pour illustrer ces plots). La première version de la grille ne comporte en général pas tous les plots linguistiques qui, on le verra ci-dessous, ont dû être revus avec Intellitech, société conceptrice de l'expert artificiel.

Si chaque grille SO a pour but de permettre le recueil des évaluations subjectives-objectivées pour alimenter l'expert artificiel, elle peut aussi être vue comme un modèle de description fine d'un jeu, prenant en compte les variables que les professionnels du secteur estiment incontournables du genre auquel il appartient.

1.2 . Construire les grilles

1.2.1 Les grilles SO

Après chaque atelier avec un groupe d'experts, nous avons repris la première version de la grille obtenue en fin d'atelier. Nous avons classé les variables figurant dans la grille par grands thèmes, qui peuvent être vus comme autant de dimensions majeures d'un genre de jeu vidéo. Par exemple, pour les jeux FPS, les thèmes identifiés ont été les suivants : Armes, Combats, Vision, Univers, Rythme et Histoire, Hardware. Pour les jeux RTS, on trouve des thèmes tels que : Rapport au terrain de jeu, Cartes, Vision, Unités, Combats.

Pour chaque échelle de réponse, les plots linguistiques ont été affinés. Les experts mobilisés pour remplir les grilles SO finales ne devront pas faire appel, pour répondre, à leur avis personnel sur tel ou tel jeu, mais bien à leur connaissance d'un maximum de jeux du genre pour évaluer les jeux à tester par rapport aux autres jeux du même genre. Les questions posées dans les grilles SO ont donc toutes été formulées (avec Intellitech) de manière à réduire le risque de sollicitation de l'avis personnel, subjectif, de l'expert.

Toutes les questions d'une même grille reflètent le caractère particulier d'un genre de jeu. Pendant l'atelier sur les jeux RTS, un concepteur a parlé de la nécessité de poser une question sur la quantité de connaissances que réclame le jeu. Il a évoqué non pas les mécanismes de jeu, mais les références culturelles que le joueur doit connaître au sujet de l'univers du jeu. Il a en effet expliqué que dans le cas d'un jeu FPS, 80 % des références sont connues. Ex. : si je tire sur un bidon d'essence, il explose. Mais dans le cas du jeu RTS, 90 % des références ne sont pas connues. Ex. : Que veut dire « Couvre le flan sud ! » ? La grille SO consacrée aux jeux RTS pose donc la question suivante : « Caractérisez l'accessibilité. Combien de références culturelles le joueur doit-il apprendre ? »

Nous insistons sur le fait que toutes les grilles SO, avec les questions qu'elles posent, les grands thèmes dans lesquels sont classées ces dernières et les plots linguistiques, ont été entièrement réalisées en collaboration avec les professionnels du secteur. Elles sont le résultat de la mise en commun de leurs expertises respectives. Chaque grille n'a été finalisée qu'après plusieurs itérations entre le laboratoire Lutin, Sébastien Genvo et les studios de développement (mais aussi Intellitech, pour valider le format de la grille).

Enfin, la dernière version de la grille est renvoyée à tous les participants. Chacun teste la grille avec le jeu prévu pour le pré-test. Si aucun problème n'est rencontré, la grille est validée et sera lancée auprès des experts pour être remplie pour chaque jeu.

1.2.2 La grille S

Une unique grille S a été construite pour recueillir les avis des joueurs venus tester le corpus de jeux Gamelab au laboratoire Lutin. Cette grille pose des questions de compréhension, de manipulation et interroge le joueur sur le plaisir qu'il a pris à jouer. Là encore, une échelle de réponses à 7 niveaux est

proposée au joueur³.

Nous avons, avec Intellitech, limité au maximum les possibilités d'erreurs d'interprétation des énoncés. Par exemple, l'énoncé « j'ai trouvé le jeu prenant », trop flou et ouvrant le champ à trop d'interprétations possibles, a été remplacé par « j'ai trouvé le jeu captivant ». Une question sur le prix que les joueurs seraient prêts à payer a été supprimée car les jeux testés ne sont pas tous rattachés à la même plate-forme. Le prix des jeux étant différent selon les plate-formes, nous avons préféré supprimer la question.

1.2.3. La grille O

Une unique grille O a été construite pour décrire les propriétés objectives d'un jeu vidéo. On y trouve, par exemple, le classement PEGI du jeu, les descriptions du système de sauvegarde, de l'époque à laquelle le jeu se situe, du type de graphismes employés, du type de périphérique employé, etc⁴. Cette grille a été réalisée au Lutin et a été relue et corrigée par Sébastien Genvo.

Type de grille	Quantité	Genre concerné	Nombre de questions
SO	1	Action-aventure	48
SO	1	FPS	58
SO	1	Casual games	25
SO	1	RTS	49
S	1	Tous types de jeux	11
O	1	Tous types de jeux	106

Tableau récapitulatif des grilles construites pour l'expert artificiel

1.3 . Remplir les grilles (sur des jeux existants)

1.3.1. Remplissage des grilles S

Pour fournir suffisamment de données S à l'expert artificiel, nous avons fait passer 240 tests. 10 testeurs ont rempli une grille S pour chacun des 24 jeux du corpus Gamelab. Les tests ont tous été réalisés au laboratoire Lutin. Chaque testeur a joué 50 minutes à un jeu avant de l'évaluer au moyen de notre grille S.

Les premiers testeurs ont été recrutés directement au sein de la Cité des Sciences, autour de la Cité des Sciences (par affichage⁵), ou bien à partir d'annonces postées sur des forums de jeux vidéo. Dans le cas des jeux *casual*, le profil de joueur recherché est un profil non expert. Le recrutement a donc été effectué dans la Cité des Sciences. Pour chaque type de tests, nous avons demandé aux testeurs de parler du test autour d'eux. Les testeurs recrutés en vue des jeux RTS ont davantage été recrutés sur des forums.

La construction de l'échantillon a été faite en s'appuyant sur les chiffres de TNS Sofres 2007⁶ et de Médiamétrie 2006⁷. Nous avons réalisé un échantillon spontané. Selon le genre de jeux testés, un niveau d'expérience dans les jeux vidéo plus ou moins élevé a été demandé aux participants. Pour les jeux Action-Aventure et FPS, les testeurs devaient au moins avoir une vague idée du jeu vidéo ; pour les jeux *casual*, aucune expérience n'a été requise. En revanche, pour les RTS, nous avons recruté des joueurs de RTS.

3 Voir la grille S en annexe.

4 Voir la grille O en annexe.

5 Voir l'affiche de recrutement en annexe.

6 TNS Sofres, *Le marché français des jeux vidéo : console, PC, téléphone portable*, 2007

7 Médiamétrie, *Les chiffres clés du jeu vidéo*, 2006

	Joue + 10heures/ semaine		Joue 1 ou 2heures /semaine		Joue quelques fois/mois		Ne joue quasiment jamais
- 15 ans							
15-25 ans	41 h	14 h	6 h		7 h	4 h	
	2 f	3 f	5 f		2 f		6 f
25-35 ans	5 h		1 h		2 h	2 h	1 h
	1 f	1 f			1 f		3 f
+ 35 ans	6 h	3 h	1 h				1 h
			1 f		1 f		

Répartition des testeurs par sexe (h/f), tranches d'âge et expérience du jeu vidéo

Protocole des tests S : chaque joueur a été invité à jouer pendant 50 mn et à remplir ensuite la grille S (pendant 10 mn), puis à rejouer 50 mn à un autre jeu, et à remplir une deuxième grille S (pendant 10 mn). Chaque participant a testé deux jeux dans le but d'alléger les passations.

1.3.2. Remplissage des grilles O

Après avoir testé les 24 jeux du corpus, nous avons rempli une grille O pour chacun d'entre eux, soit 24 grilles O au total.

1.3.3. Remplissage des grilles SO

Avant de pouvoir faire remplir les grilles SO, il a fallu identifier les professionnels pertinents pour chacun des genres de jeux à évaluer et s'assurer de leur disponibilité pour le projet. Parmi les professionnels sollicités figurent à la fois des concepteurs de jeux vidéo issus principalement des entreprises de Capital Games, des étudiants d'écoles de multimédia et de jeux vidéo (Enjmin, SupInfoGames, Gobelins, etc.) et quelques indépendants (journalistes).

Au début du projet, il était prévu que 8 experts remplissent chacun 6 grilles pour un genre donné. Il est très important pour l'expert artificiel que ce soit une seule et unique personne qui remplisse les 6 grilles. Nous avons rencontré plusieurs difficultés de recrutement, principalement en raison des emplois du temps très chargés des entreprises, prises dans leurs impératifs de production. Certains experts qui avaient accepté de participer ont dû abandonner les tests en cours de projet. Nous avons dû revoir le nombre d'experts à mobiliser et avons finalement décidé de recruter 7 experts par genre de jeux. En raison des multiples défections de certains participants, un genre n'a pas pu être évalué : le genre Action-aventure.

Une fois les participants identifiés et la prise de contact effectuée, chacun d'entre eux a reçu la liste des 6 jeux d'un même genre à évaluer, accompagnée de la grille SO à remplir. Au total, chaque expert a donc eu 6 grilles SO à remplir.

1.3.4. Tableaux récapitulatifs des grilles remplies pour l'expert artificiel

Type de grille	Nombre de grilles remplies Action-aventure	Nombre de grilles remplies FPS	Nombre de grilles remplies Casual games	Nombre de grilles remplies RTS	Nombre total de grilles S remplies
S	60	60	60	60	240

Type de grille	Nombre de grilles remplies Action-aventure	Nombre de grilles remplies FPS	Nombre de grilles remplies Casual games	Nombre de grilles remplies RTS	Nombre total de grilles O remplies
O	6	6	6	6	24

Type de grille	Nombre de grilles remplies Action-aventure	Nombre de grilles remplies FPS	Nombre de grilles remplies Casual games	Nombre de grilles remplies RTS	Nombre total de grilles SO remplies
SO	32	42	42	41	157

1.3.5. Tableaux des experts recrutés

Types de jeux	Nombre d'experts mobilisés	Entreprises de Capital Games	Etudiants	Autres
Action-aventure	6	Wizarbox (1)	SupInfoGames (3) Gobelins (1)	Journaliste (1)
FPS	7	Kylotonn (5)	Gobelins (2)	
Casual games	7	Mekensleep (2) Black Sheep Studio (3)	Enjmin (1) Gobelins (1)	
RTS	7	Eugen System (2)	Enjmin (3)	Ubisoft (1) Indépendante (1)

2 . Fonctionnement de l'expert et identification des modèles

L'expert artificiel n'est pas un outil de classement ou de jugement automatisé (comme le serait une nouvelle mesure d'audience), mais un outil de prévision des appréciations précises et d'explication de ces jugements pour pouvoir adapter le jeu avant sa mise sur le marché.

2.1 . Principes

2.1.1. Les données de référence

L'expert artificiel prend en compte trois types de données :

- 1) Les données S : les évaluations subjectives des joueurs sont décomposées en énoncés verbaux flous, et pourtant modélisables et calculables.
- 2) Les données SO : les qualités d'un jeu vidéo sont objectivées grâce à la connaissance de plusieurs experts du domaine validant des descriptions fines en les testant sur un nombre de jeux minimum.
- 3) Les données O : les propriétés objectives d'un jeu sont décrites précisément.

Les grilles S, O et SO, ont été envoyées aux experts au format .xls afin que ceux-ci les remplissent. Pour pouvoir être traitées par l'expert artificiel, ces grilles ont été formatées en .txt.

2.1.2. L'expert artificiel et la logique floue

Les croisements entre ces trois types de données (S, SO et O) sont modélisés à partir d'algorithmes de logique floue qui sont générés automatiquement puis testés sur leur robustesse pour produire des modèles explicatifs à au moins 2 règles et à une ou plusieurs variables.

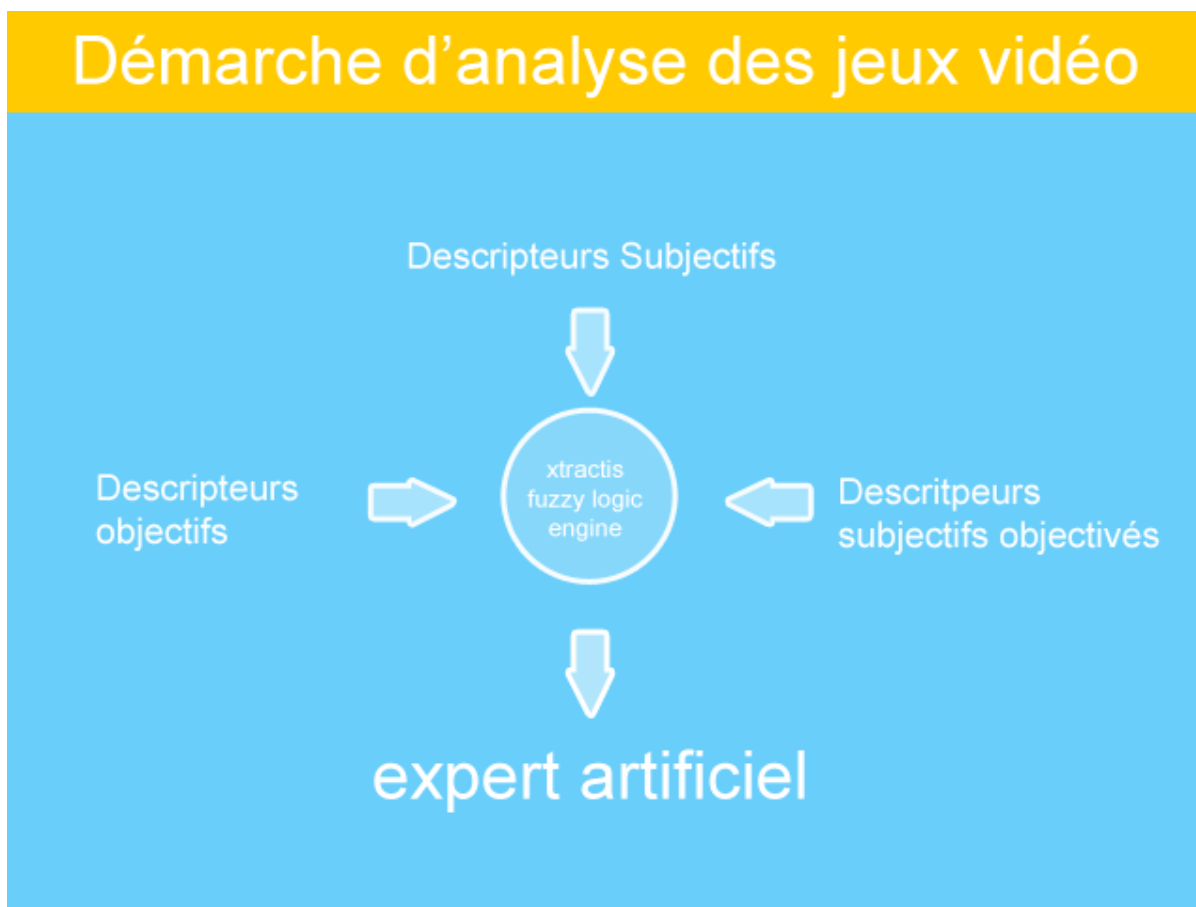


Schéma explicatif du fonctionnement de l'expert artificiel

2.1.3. Le fonctionnement par « projet »

L'expert artificiel fonctionne par « projet ». Chacun des genres de jeux vidéo a été traité séparément des autres dans l'expert artificiel. Traiter tous les genres en même temps n'aurait eu aucun sens. Pour chaque projet, l'expert artificiel a besoin d'être alimenté en données S, O et SO. En raison du nombre insuffisant d'évaluations SO dans le genre Action-aventure, nous n'avons pas généré de projets pour ce genre.

2.2 . Préparer les calculs

Une fois que nous disposons de toutes les grilles (O, S et SO) pour un genre donné, nous pouvons lancer les modélisations avec l'expert artificiel.

2.2.1. Choix du type de projet

L'expert artificiel peut traiter plusieurs types de « projets » : O-SO, SO-S, O/SO-S, etc. Un projet SO-S prendra en compte en entrée les données SO et en sortie les données S. Le projet aura pour objectif de prédire les sorties S en fonction des données SO, c'est-à-dire de prédire les évaluations que feraient les joueurs à partir des évaluations des jeux faites par les experts. Un projet O/SO-S aura pour objectif de prédire les sorties S en fonction des données objectives et des données subjectives-objectivées. Etc.

Dans le cadre du projet Gamelab, nous avons engagé deux types de projets (un projet SO-S et un projet O/SO-S) pour chacun des trois genres de jeux.

2.2.2. Préparation des calculs

Lancer un projet sur l'expert artificiel réclame une phase de préparation des calculs. Pour chaque projet, il faut :

- spécifier les entrées et sorties : choix des bases de référence
- préciser les domaines de définition pour les entrées et les sorties
- créer une base d'apprentissage
- filtrer les évaluations (si nécessaire)
- paramétrer les calculs (choix de la robustesse, corrélation minimale, etc.)

2.2.3. Lancement des calculs

On peut ensuite lancer les calculs. Dans le cas du projet Gamelab, nous avons lancé systématiquement des « baguettes magiques », option de l'expert artificiel qui permet de générer des modèles pour toutes les sorties à la fois, durant une même session.

2.3 . Trier les modèles

Un modèle est composé d'un certain nombre de variables et fonctionne selon un certain nombre de règles. Dans le cadre du projet Gamelab, nous avons défini avec Intellitech que les modèles devaient au minimum être composés d'une variable et au maximum de 8 variables (pour rester intelligibles). Chaque modèle doit également fonctionner avec deux règles au minimum et 4 règles au maximum (pour éviter une trop grande complexité).

2.3.1. Arrêt des calculs

Pour espérer obtenir suffisamment de modèles robustes, il faut que l'expert artificiel ait effectué au minimum 200 générations⁸ par sortie. Une fois ce chiffre minimum atteint ou dépassé, on peut arrêter les calculs et commencer à trier les modèles.

2.3.2. Sélectionner un premier groupe de modèles en fonction de leur précision et de leur robustesse

⁸ Une génération comporte plusieurs modèles.

Une fois les calculs terminés, on sélectionne la sortie qui nous intéresse. L'expert artificiel génère une liste des modèles qu'il a créés. Dans le cadre du projet Gamelab, on a obtenu environ 15000 modèles par sortie, tous projets confondus.

Parmi ces modèles, beaucoup ne sont pas suffisamment performants. On commence par sélectionner ceux qui présentent la meilleure précision associée à la meilleure robustesse possible (la robustesse des modèles est obtenue par validation croisée des méthodes Leave-One-Out et Monte-Carlo). La démarche est la suivante :

- on élimine d'abord tous les modèles qui ne « mappent » pas tous les points (à la fois en Accuracy, en Monte Carlo et en LOO)
- on trie les modèles en corrélation (Monte-Carlo) pour trouver le modèle qui a la meilleure robustesse (on ajoute ce modèle à notre *model list* pour cette sortie)
- on élimine les modèles dont la corrélation et le hamming sont moins bons en précision que ceux du meilleur modèle en robustesse

2.3.3. Sélectionner les meilleurs modèles

Une fois le premier tri effectué, on s'attache à isoler les modèles les plus intéressants, à savoir : ceux qui sont les plus précis tout en étant les plus robustes possibles, et ceux qui sont les moins complexes possibles, tout en restant robustes. La démarche est la suivante :

- On trie les modèles par nombre de variables et de règles.
- On sélectionne les meilleurs modèles (la meilleure corrélation possible associée au *hamming* le plus faible possible) en partant du moins complexe pour aller vers le plus complexe.
- Chacun des modèles sélectionnés est ajouté à la *model list*.
- Parmi les modèles de la *model list*, on sélectionne les *top models*.

3 . Analyses de modèles et résultats

L'expert artificiel permet d'obtenir trois types de connaissances au sujet des jeux vidéo. Dans un premier temps, les modèles générés par l'expert artificiel permettent de comprendre comment se font les jugements des joueurs au sujet des jeux vidéo. Si ces derniers ne peuvent pas expliquer clairement pourquoi ils apprécient ou non telle ou telle dimension d'un jeu, l'expert artificiel, parce qu'il est capable de mettre en relation leurs jugements et les évaluations expertes, est en mesure de fournir des règles explicatives de ces jugements.

Dans un deuxième temps, l'expert artificiel est capable de prédire les jugements que les joueurs feraient s'ils étaient interrogés sur un jeu. Ses modèles explicatifs sont également prédictifs. Si on ajoute aux bases de référence de l'expert la description d'un nouveau jeu, celui-là sera capable de prédire les sorties qui lui correspondent en se fondant sur ses modèles. L'expert artificiel peut donc fonctionner comme un "testeur virtuel" d'un jeu en développement.

Enfin, l'expert artificiel est capable d'inverser ses modèles et de proposer des solutions optimales en fonction d'une requête fixant des objectifs à atteindre (ex. : les joueurs doivent trouver le jeu le plus captivant possible) et imposant des contraintes (la présence d'une *feature* dans le jeu, une obligation *publisher*, etc.).

Avec l'expert artificiel, les professionnels du secteur peuvent disposer d'un outil capable d'expliquer les jugements des joueurs en fonction de catégories qu'ils ont eux-mêmes identifiées comme les plus pertinentes pour un genre de jeux donné puisqu'ils ont participé à la construction des grilles d'évaluation. Ils peuvent aussi connaître à l'avance les jugements que les joueurs rendraient sur leur jeu en

développement, à partir de la description précise de celui-ci. Enfin, ils peuvent optimiser la conception de leur jeu en développement grâce à l'expert artificiel, limitant ainsi les essais-erreurs pendant la phase de développement et augmentant les chances de succès au moment de la mise sur le marché.

3.1 . Comprendre

Les modèles générés par l'expert artificiel permettent de comprendre pourquoi les joueurs apprécient telle ou telle dimension d'un jeu vidéo. L'expert artificiel est capable d'isoler les variables évaluées par les experts qui influent sur le jugement positif ou négatif du joueur au sujet d'une dimension du jeu.

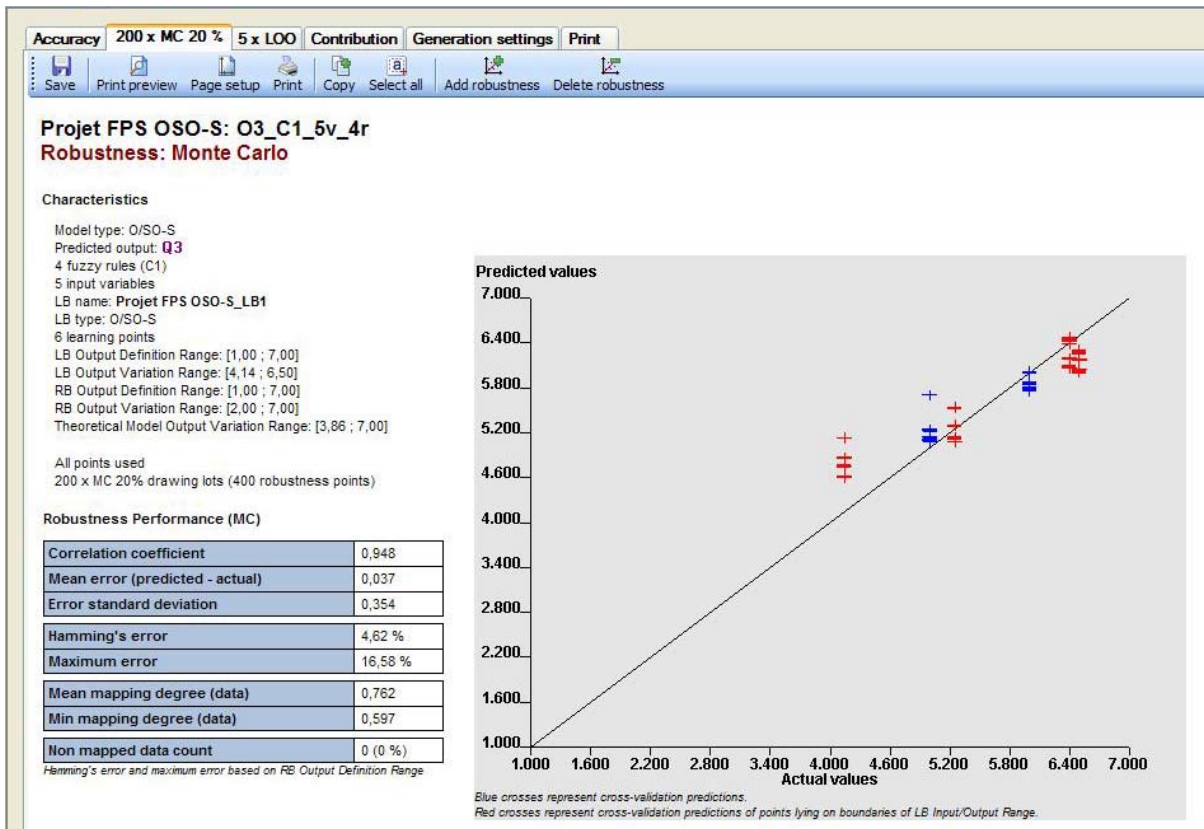
3.1.1. Identifier les variables clés

Prenons un exemple. Nous cherchons à savoir quelles variables ont influé sur le fait que les joueurs ont trouvé un FPS ennuyeux ou captivant. Parmi les modèles générés pour la sortie 3 de la grille S, on sélectionne un modèle pour sa robustesse.

La robustesse du modèle est évaluée en fonction de plusieurs critères. Parmi eux :

- le coefficient de corrélation (notamment en Monte-Carlo)
- le *hamming*
- l'erreur maximale
- l'allure générale du graphique de robustesse

L'un des atouts majeurs de l'expert artificiel est de proposer des modèles précis, mais qui soient aussi fiables en robustesse. Il est nécessaire, pour chaque modèle, de s'assurer que ce celui-ci n'a pas « appris par coeur ». Sur le graphique de robustesse en Monte-Carlo, un modèle qui a « appris par coeur » se reconnaît aisément car la courbe formée par les points de calcul de robustesse se retrouve à l'horizontale. Pour tester la robustesse, l'expert artificiel recommence plusieurs fois ses calculs en ôtant des points (un point correspondant à une évaluation d'un jeu par les experts). Plus les nuages de points sont ramassés, plus le modèle est robuste.



Exemple d'un graphique de robustesse calculée avec la méthode Monte Carlo

Une fois assurés de la robustesse du modèle, nous constatons que ce modèle utilise 5 variables et fonctionne selon 4 règles. Ci-dessous la règle 1 et la règle 4 :

Projet FPS OSO-S: O3_C1_5v_4r

Model rule list (4 rules):

- Rule 1 (all input variables used):**

If	O_Q6 pour accomplir certains objectifs	is	about 1.01 il y a un temps limite
and	O_Q64 le profil de l avatar	is	about 1.01 est accessible
and	Q2 la coherence son arme	is	about 3.29 comporte un faible nombre d incoherences
and	Q3 le gameplay des armes	is	about 1.46 est tres coherent
and	Q33 les vehicules	is	about 0.83 sont inexistant dans le jeu
Then	Q3	is	3.86 jouer a ce jeu est plutot ennuyeux

- Rule 4 (all input variables used):**

If	O_Q6 pour accomplir certains objectifs	is	about 0.01 il n y a pas de temps limite
and	O_Q64 le profil de l avatar	is	about -0.01 est inaccessible
and	Q2 la coherence son arme	is	about 1.80 est forte
and	Q3 le gameplay des armes	is	about 2.80 comporte un tres faible nombre d incoherences
and	Q33 les vehicules	is	about 5.21 apparaissent en fonction des sequences
Then	Q3	is	7.00 jouer a ce jeu est tres captivant

Selon ce modèle, les 5 variables les plus influentes pour comprendre pourquoi les joueurs estiment qu'un FPS est ennuyeux ou captivant sont :

- 2 variables objectives : le temps limité pour accomplir certains objectifs et l'accès au profil de l'avatar
- 3 variables subjectives-objectivées : la cohérence son/arme, le gameplay des armes et les véhicules

Pourtant, avoir identifié les variables qui entrent en jeu dans les évaluations des joueurs n'est pas suffisant.

3.1.2. Observer les valeurs des variables

Si l'on observe les règles explicatives du modèle, on s'aperçoit que dans un cas, une règle aboutit à un jugement selon lequel les joueurs estiment que le FPS est « plutôt ennuyeux » (règle 1), alors que dans l'autre cas, les joueurs estiment que le FPS est « très captivant » (règle 4). Pourtant, ce sont bien les mêmes variables qui sont prises en compte dans les deux règles ! En revanche, ce sont les valeurs que prennent les variables qui sont déterminantes. Cela signifie que même si deux jeux présentent les mêmes *features*, la manière dont ces dernières sont dosées influe considérablement sur le résultat final qui sera proposé aux joueurs. Pour un modèle donné, les mêmes variables peuvent prendre des valeurs différentes et expliquer des jugements opposés sur une même dimension du jeu. Il importe donc, non pas d'avoir seulement identifié les variables clés, mais aussi de connaître l'influence qu'elles ont sur les jugements des joueurs en fonction de la valeur qu'elles prennent.

Pour aller plus loin, il est essentiel de regarder de près les valeurs que prennent les variables. On pourrait faire l'hypothèse qu'en inversant le dosage d'une variable qui donne un résultat négatif, on obtienne un jugement positif. C'est faux ! Parfois, c'est un dosage intermédiaire qui donnerait le jugement le plus favorable. Selon notre exemple, ce n'est pas parce que le fait que les véhicules sont inexistant dans le jeu influe sur le fait que les joueurs trouvent le FPS « plutôt ennuyeux » que le fait de trouver des véhicules partout dans le jeu fera que les joueurs trouveront le FPS « très captivant ». La règle 4 de notre modèle montre bien que le fait que les véhicules apparaissent *en fonction des séquences* contribue au jugement « très captivant ». Il ne suffirait donc pas de maximiser la variable Q33 pour obtenir le meilleur jugement. L'expert artificiel montre que certaines *features* des jeux vidéo doivent être dosées très finement.

3.1.3 Interpréter le modèle

Observons maintenant un modèle explicatif du plaisir à jouer à un jeu RTS.

Projet RTS OSO-S 23JAN: O6_C1T_6v_3r
 Model rule list (3 rules):

- **Rule 1** (all input variables used):

If	O_Q5 rencontrer des boss	is	about 0.00 est impossible
and	O_Q64 avoir acces au profil de son avatar	is	about 0.49 que ce soit possible ou pas
and	O_Q71 modifier les caracteristiques de son avatar	is	about 0.49 que ce soit possible ou pas
and	O_Q95 pendant la partie acceder au menu depuis des boutons sur l ecran	is	about 0.00 est impossible
and	Q18 la mini map	is	about 3.66 est peu presente
and	Q40 accessibilite La quantite de referenes culturelles que le joueur doit apprendre	is	about 1.25 est tres peu c est comme dans le monde reel
Then	Q6 jouer a ce jeu	is	5.54 a plutot plu

- **Rule 2** (all input variables used):

If	O_Q5 rencontrer des boss	is	about 0.00 est impossible
and	O_Q64 avoir acces au profil de son avatar	is	about 1.01 est possible
and	O_Q71 modifier les caracteristiques de son avatar	is	about -0.03 est impossible
and	O_Q95 pendant la partie acceder au menu depuis des boutons sur l ecran	is	about 1.00 est possible
and	Q18 la mini map	is	about 4.56 est plutot presente
and	Q40 accessibilite La quantite de referenes culturelles que le joueur doit apprendre	is	about 3.10 est un peu importante
Then	Q6 jouer a ce jeu	is	5.84 a plutot plu

- **Rule 3** (all input variables used):

If	O_Q5 rencontrer des boss	is	about 0.97 est possible
and	O_Q64 avoir acces au profil de son avatar	is	about 1.01 est possible
and	O_Q71 modifier les caracteristiques de son avatar	is	about 1.01 est possible
and	O_Q95 pendant la partie acceder au menu depuis des boutons sur l ecran	is	about 1.00 est possible
and	Q18 la mini map	is	about 4.56 est plutot presente
and	Q40 accessibilite La quantite de referenes culturelles que le joueur doit apprendre	is	about 5.22 est plusieurs
Then	Q6 jouer a ce jeu	is	6.60 a plu

Observons la variable Q40. Au premier abord, on pourrait être surpris. A priori, on associe plaisir à jouer et facilité à jouer (selon le principe qu'"un bon jeu est un jeu où je gagne"). Pourtant, on remarque que dans le cas de la règle 3, explicative du jugement « jouer à ce jeu m'a plu », la quantité de références culturelles que le joueur doit apprendre est notée 5,22 sur 7 et correspond à « plusieurs références à apprendre », tandis que dans la règle expliquant le jugement « jouer à ce jeu m'a plutôt plu », cette même variable est notée 1,25 sur 7 et correspond à « il y a très peu de références culturelles à apprendre ». On pourrait donc s'étonner de ce résultat. S'il y a plus de références à apprendre, on peut faire l'hypothèse que le jeu est plus difficile et donc, plaît moins. Cependant, le modèle met en évidence le fait qu'un jeu RTS cible un public de joueurs confirmés. Le jeu plairait moins aux joueurs s'il leur demandait moins d'expertise. Ceci se vérifie d'ailleurs avec d'autres variables. L'importance de la présence de la mini-map, la possibilité d'accéder au profil de l'avatar (de l'armée dans le cas d'un jeu RTS), etc., confirment qu'un RTS plaira d'autant plus que les joueurs disposeront des moyens de leur puissance.

3.2. Prédire

Les modèles de l'expert artificiel permettent de prédire l'appréciation des joueurs à partir du profil défini d'un jeu en cours de développement. Si les concepteurs se posent une question sur leur jeu en développement, l'expert artificiel est capable de leur prédire quel serait le jugement des joueurs sur ce problème précis, à condition qu'ils aient au préalable rempli une grille de description de leur jeu (grille SO) et que nous l'ayons ajoutée aux bases de données de référence de l'expert artificiel.

L'exemple ci-dessous n'est pas tiré d'une prédiction faite sur un jeu en développement puisque nous n'avons pas eu le temps de réaliser ce genre de prédictions. Il s'agit d'une simulation à partir d'un produit imaginaire, mais qui permet déjà de se rendre compte des possibilités de l'expert artificiel en termes de prédictions.

3.2.1. Écouter la question que se posent les concepteurs

L'exemple ci-dessous fait l'hypothèse que les concepteurs d'un FPS se demandent s'ils doivent placer la *secondary view* sur toutes les armes, ou seulement sur certaines d'entre elles. En examinant notre banque de modèles pour les jeux FPS, nous pouvons tout d'abord identifier sur quelle(s) dimension(s) du jeu la *secondary view* influe. Sur les 11 sorties de la grille S, 4 sont concernées. La *secondary view* influe donc sur le fait que les joueurs :

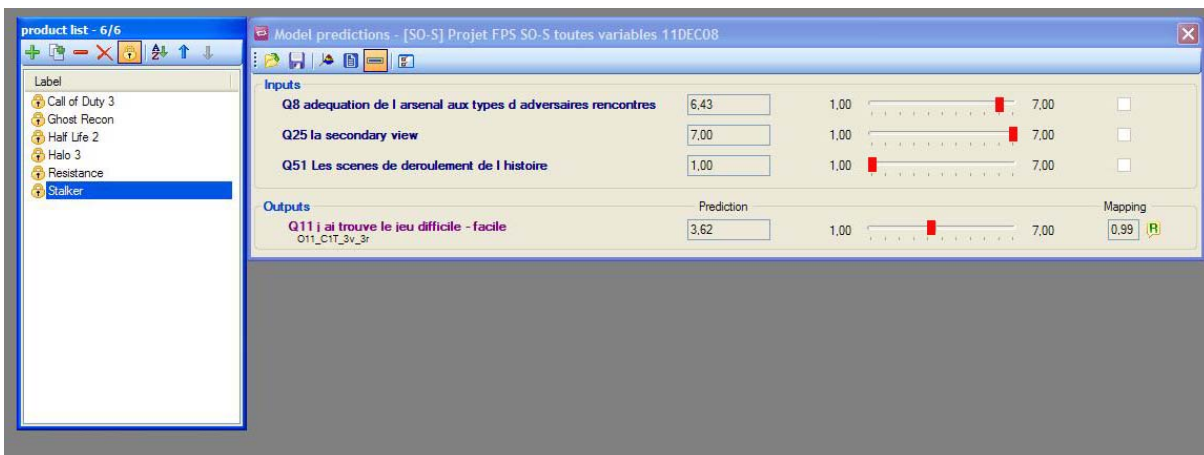
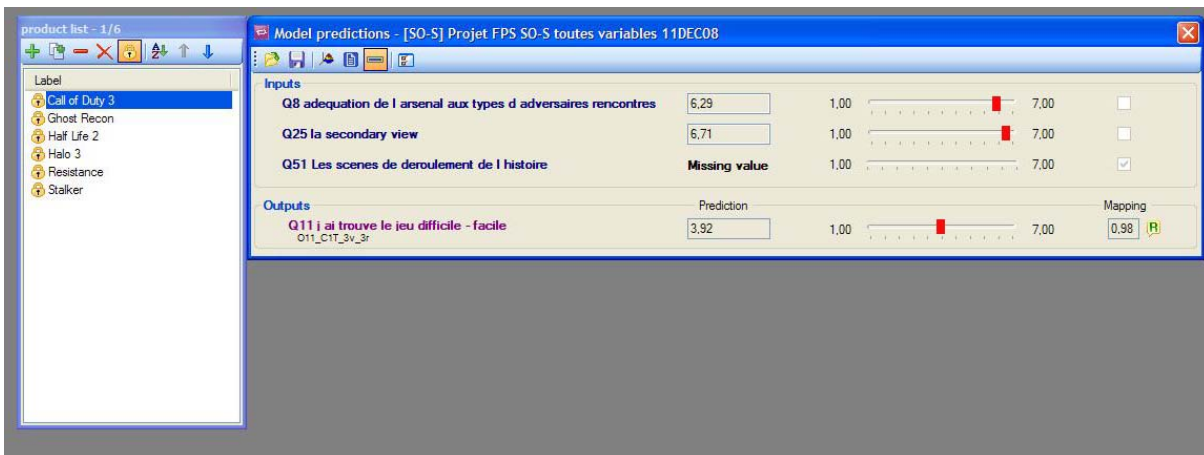
- trouvent le jeu ennuyeux / captivant
- trouvent l'utilisation des contrôles facile / difficile
- estiment que le jeu leur a plu / déplu
- trouvent le jeu difficile / facile

Précisons que nous pouvons répondre aux questions des concepteurs qui interrogent une ou plusieurs dimensions recensées dans nos grilles. Nous ne pouvons donc pas répondre à n'importe quelle question que se posent les concepteurs.

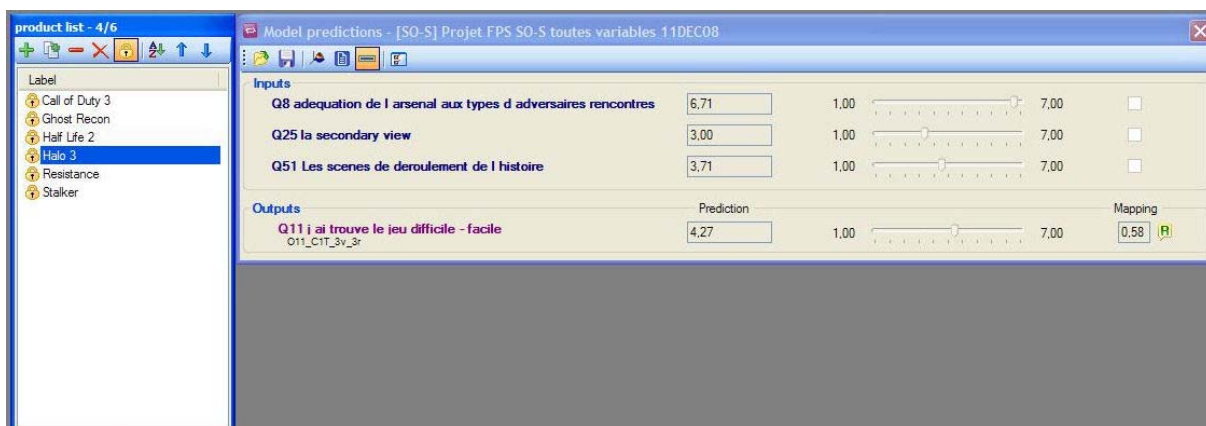
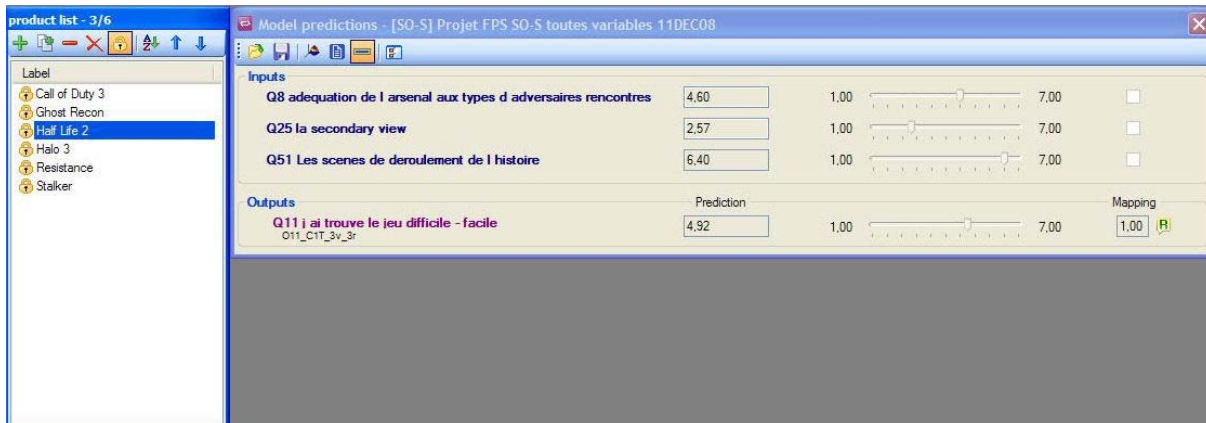
3.2.2. Observer les jugements des joueurs sur le corpus Gamelab

Imaginons que les concepteurs s'intéressent plus particulièrement à la difficulté ressentie par les joueurs. L'expert artificiel est capable de prédire si les joueurs trouveront le jeu difficile ou facile en fonction du type de *secondary view* que les concepteurs ont choisie pour leur jeu.

Voyons d'abord les effets de la *secondary view* sur le corpus Gamelab. Pour cela, on sélectionne les modèles générés pour la sortie 11 de la grille S.



Notons tout d'abord que le modèle qui a permis ces prédictions ne prend pas en compte la seule influence de la *secondary view* sur la difficulté ressentie par le joueur. Le modèle associe à la *secondary view* deux autres dimensions qui doivent être prises en compte dans l'analyse du modèle : l'adéquation de l'arsenal aux types d'adversaires rencontrés et les scènes de déroulement de l'histoire. Dans les deux cas, pour Call of Duty 3 et Stalker, la *secondary view* a été notée proche de 7 par les experts, ce qui correspond à une présence sur presque toutes les armes (note de 6,71), voire sur toutes les armes (note de 7). Ces deux jeux ont été jugés « plutôt difficile[s] » par les joueurs. Examinons maintenant les prédictions pour les jeux qui ne proposent pas la *secondary view* sur toutes les armes, mais seulement sur quelques-unes.

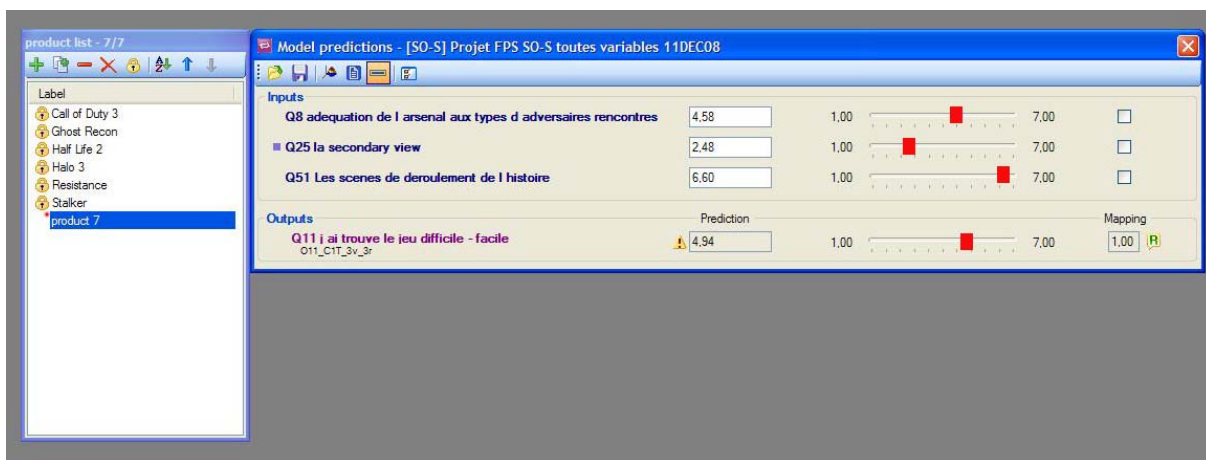
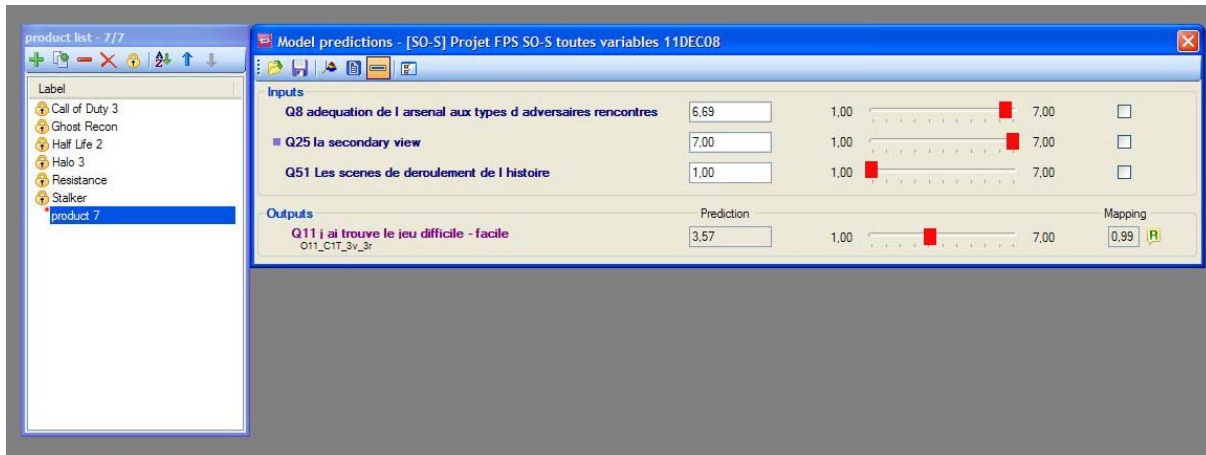


Pour les jeux Half Life 2 et Halo 3, la *secondary view* a été notée respectivement à 2,57 et à 3. La note 3 correspond à la présence de la *secondary view* sur « quelques armes » et la note 2,57 à une présence de la *secondary view* sur peu d'armes. Si nous observons maintenant les jugements sur la difficulté faits par les joueurs, on obtient pour Half Life 2 la note de 4,92 et pour Halo 3 la note de 4,27. Les plots linguistiques de la grille S font correspondre à 4 un jugement qui estime le jeu entre « plutôt difficile » et « plutôt facile ». On en déduit que, dans le cas de Halo 3, le joueur n'a pas jugé que le jeu était ni « plutôt difficile » ni « plutôt facile », un jugement neutre donc. Et dans le cas de Half Life 2, la note est presque de 5, qui correspond au plot linguistique « plutôt facile ».

Si la *secondary view* n'est pas la seule variable qui permette d'expliquer le fait que les joueurs estiment que le jeu est plus facile, puisque l'expert artificiel a construit un modèle à plusieurs variables, pour autant il est intéressant de constater qu'un jeu a plus de chance d'être estimé comme « plutôt facile » si la *secondary view* n'apparaît que sur quelques armes. Si donc les concepteurs souhaitent que leur jeu soit ressenti comme « plutôt facile », nous pouvons a priori leur conseiller de ne mettre la *secondary view* que sur quelques armes.

3.2.3. Prédire les jugements des joueurs

Pour obtenir une prédiction certaine de la note qu'obtiendrait le jeu en développement, nous devons ajouter l'évaluation SO du jeu en développement à nos bases de référence. Voyons ci-dessous deux simulations de jeux, l'un proposant une *secondary view* sur toutes les armes, l'autre seulement sur quelques-unes.



On remarque que, dans le cas d'un jeu proposant une *secondary view* sur toutes les armes, la note prédite est de 3,57, à savoir : le jeu serait jugé « plutôt difficile ». Dans le cas du jeu qui propose une *secondary view* sur quelques armes, la note prédite est de 4,94, à savoir : le jeu serait presque jugé « plutôt facile ».

Rappelons que la *secondary view* n'est pas responsable à elle seule de la note prédite pour le jeu. Les autres variables doivent elles aussi être prises en compte pour obtenir une prédiction précise.

3.3 . Optimiser

Enfin, les modèles générés par l'expert artificiel permettent d'optimiser les composantes d'un jeu en fonction du but recherché (maximiser l'appréciation des joueurs, etc.), tout en tenant compte des contraintes éventuelles (*features* déjà présentes dans le jeu, contraintes venant du *publisher*, etc.). Plus les concepteurs interrogent tôt l'expert artificiel, plus ils auront la possibilité d'ajuster leurs choix aux conseils de l'expert artificiel. Si l'on interroge trop tard l'expert artificiel dans le processus de développement du jeu, certaines dimensions du jeu risquent de ne plus pouvoir être modifiables selon les conseils de l'expert artificiel, alors qu'elles l'auraient été en demandant une optimisation plus tôt dans le développement.

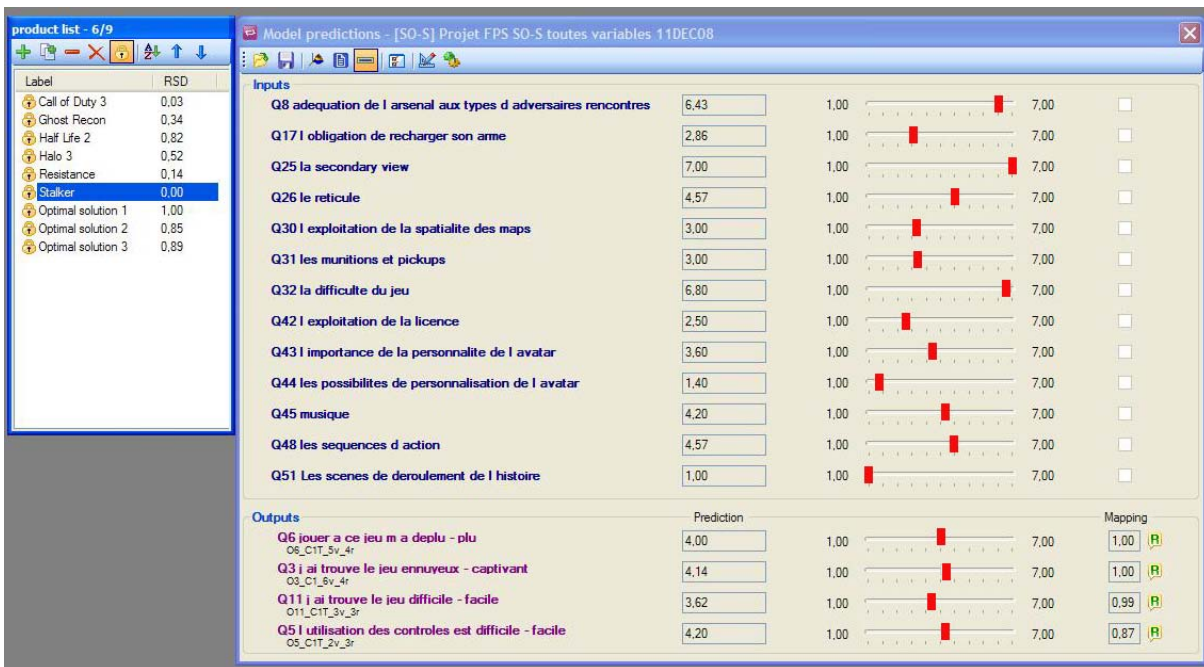
3.3.1. Écouter la requête des concepteurs

Faisons l'hypothèse que les concepteurs souhaitent un jeu FPS qui soit le plus captivant possible, avec des contrôles jugés très faciles, un jeu qui plaise le plus possible et que les joueurs jugent le plus facile possible⁹. L'expert artificiel est capable de prendre en compte plusieurs modèles à la fois et de rechercher des solutions optimales qui répondent à l'ensemble de la requête.

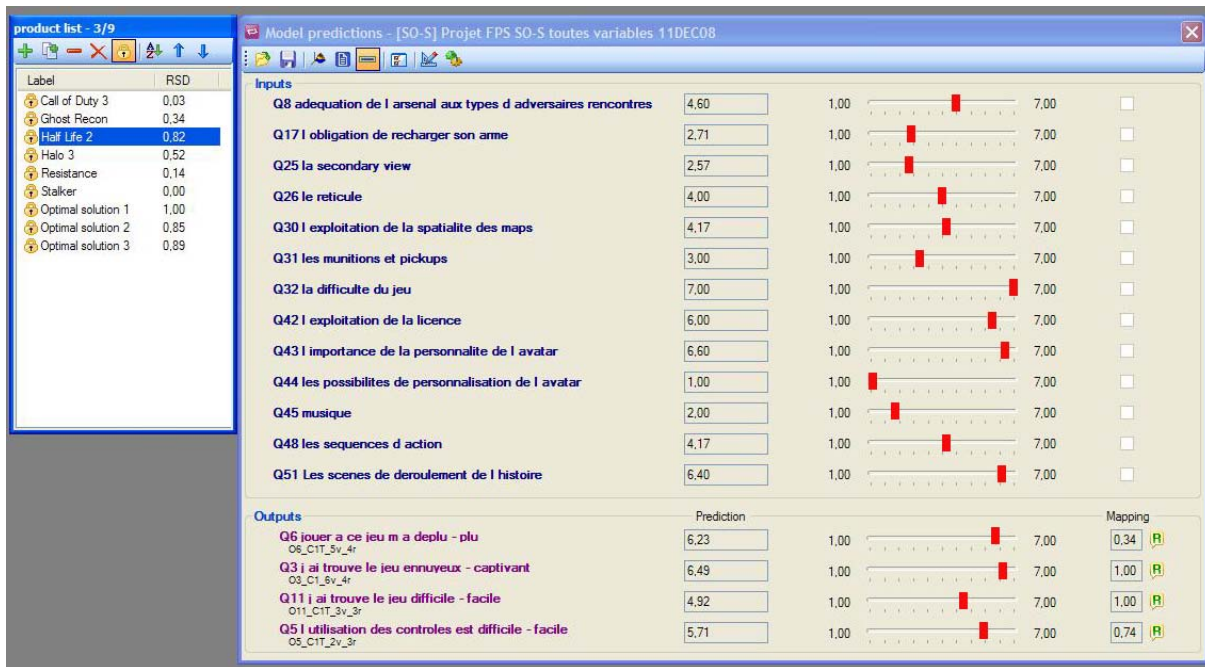
Si l'on peut indiquer à l'expert artificiel les exigences des concepteurs en termes de sorties, on peut également y ajouter des requêtes impliquant des contraintes sur les entrées. Ce cas de figure peut par exemple se présenter si les concepteurs interrogent l'expert artificiel un peu tardivement dans le processus de conception du jeu vidéo ou bien s'ils tiennent particulièrement à une *feature* ou à un « dosage » précis d'une option, etc. On pourrait très bien imaginer que l'ambiance musicale ait déjà été choisie, ou que le *publisher* impose une licence, etc. Dans ce cas, l'expert artificiel chercherait les solutions optimales en fonction de ces contraintes.

3.3.2. Observer le corpus Gamelab

Il est possible d'identifier le jeu du corpus Gamelab qui remplit le mieux possible cette requête. L'exemple que nous allons présenter ci-dessous est à nouveau une simulation puisque, pour des raisons de délais, nous n'avons pas pu tester cette fonctionnalité pour un jeu en développement.



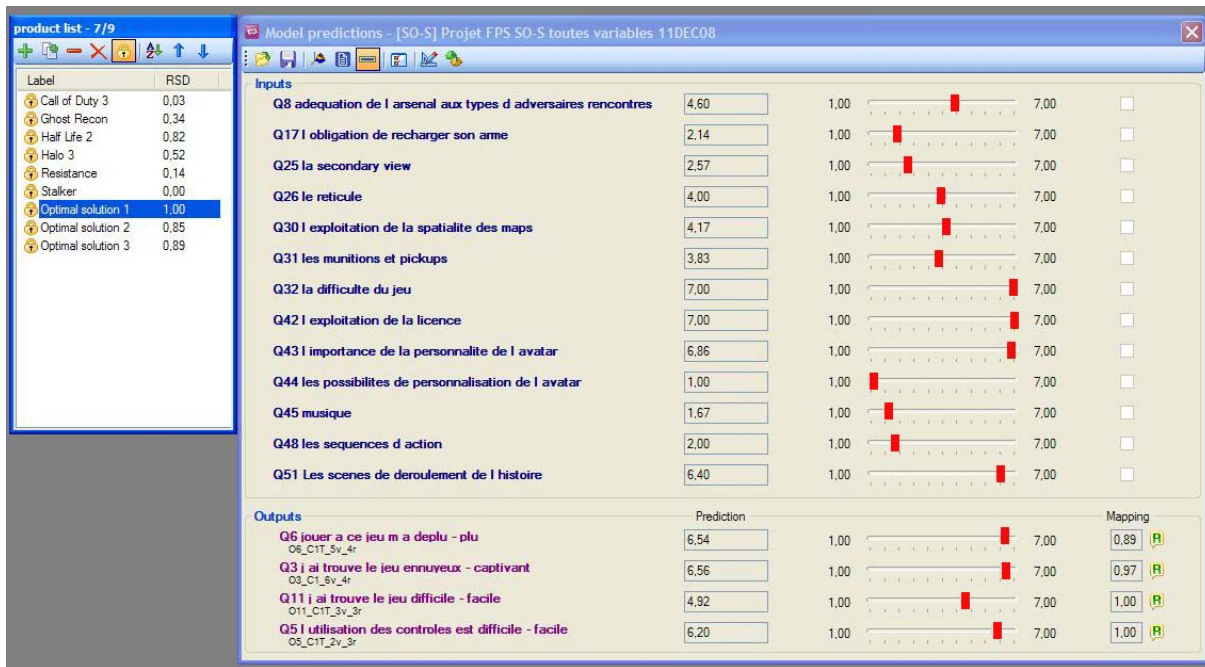
⁹ Là encore, nous n'avons pas pu tester cette fonctionnalité sur un jeu en développement. L'exemple qui va suivre est une simulation.



L'expert artificiel évalue le degré de satisfaction de la requête (colonne RSD = *Request Satisfaction Degree*) pour chacun des jeux du corpus Gamelab. On remarque que le jeu Stalker est celui du corpus Gamelab qui remplit le moins bien la requête. En revanche, Half Life 2, avec 0,82 de RSD, est celui qui répond le mieux à la requête, à savoir : il est le jeu que les joueurs ont trouvé à la fois le plus captivant, avec les contrôles les plus faciles, qui leur a le plus plu et qu'ils ont trouvé le plus facile. Attention, cela ne signifie pas forcément que Half Life obtient à chaque fois la meilleure note. En revanche, Half Life est celui qui répond le mieux à l'ensemble de la requête.

3.3.3. Recherche des solutions optimales

Reprenons notre hypothèse de départ et imaginons que les concepteurs souhaitent concevoir un jeu qui optimise les 4 sorties citées plus haut. L'expert artificiel identifie trois solutions optimales. Il est donc possible de réaliser un FPS qui soit meilleur que Half Life 2 pour les 4 sorties à la fois. Il est même possible, selon l'expert, de concevoir un jeu qui réalise parfaitement cette requête RSD = 1)



L'expert artificiel permet donc de voir comment jouer sur quelles variables pour avoir un jeu qui non seulement remplisse mieux les conditions de la requête que le jeu Half Life 2, mais encore qui permette une satisfaction optimale de la requête pour les sorties indiquées.

4. Problèmes rencontrés et avancées

Au cours du projet, nous avons rencontré plusieurs difficultés de natures différentes. Certaines sont d'ordre pratique, d'autres d'ordre scientifique. Cependant, cette première confrontation de l'expert artificiel au domaine des jeux vidéo a permis des avancées non négligeables dans la manière d'appréhender la qualité des jeux vidéo.

4.1. Difficultés pratiques

4.1.1 Les différents recrutements

Afin de recueillir un nombre suffisant de données pour alimenter les bases de références de l'expert artificiel, nous avons dû procéder à plusieurs vagues de recrutement. Certains d'entre eux n'ont pas été aisés.

Ainsi, il a par exemple été difficile de recruter suffisamment de sujets féminins pour recueillir leurs appréciations subjectives sur le corpus de jeux. Les filles / femmes constituent une population plutôt sous-représentée parmi les *gamers* ; d'autant plus lorsqu'on prend en compte des jeux FPS et RTS, qui sont plus pratiqués par des joueurs que par des joueuses.

Il a également fallu recruter des membres des entreprises de Capital Games soit pour participer aux ateliers, soit pour remplir les grilles d'évaluation. C'est cette dernière étape qui a pris le plus de retard, au point qu'un genre de jeu – Action-Aventure – n'a pas pu être traité, faute de données recueillies suffisantes. En raison de l'emploi du temps très chargé des entreprises pendant l'année 2008, certaines entreprises, malgré leur bonne volonté, n'ont pas pu rendre les grilles d'évaluations des jeux action -aventure dans les délais. Or, pour que l'expert artificiel fonctionne correctement, il est impératif de disposer d'un nombre minimum de grilles. Nous avons donc dû abandonner ce genre de jeu. Aucun modèle n'a donc pu être généré pour ce genre

de jeu.

De la même manière, il aurait été préférable qu'un plus grand nombre d'entreprises remplisse des grilles pour un plus grand nombre de jeux, afin que l'expert dispose de bases de références plus étoffées et d'évaluations expertes plus nombreuses. Pour les raisons indiquées ci-dessus, cela n'a pas été possible.

Enfin, pour les mêmes raisons de délais et d'emploi du temps surchargé, nous n'avons pas pu tester les modèles obtenus avec l'expert artificiel pour des jeux en développement. Pour cela, il aurait fallu que puisse être suivi le protocole de la méthode présentée aux entreprises¹⁰ qui, bien qu'il ait été conçu pour pouvoir être intégré au pipeline de production des entreprises, n'a pas pu être testé.

4.1.2. Un logiciel peu ergonomique

Plusieurs difficultés ont été rencontrées en raison de l'ergonomie pauvre du logiciel. La prise en main a été longue et de nouvelles versions du logiciel ont dûes être installées à plusieurs reprises.

4.1.3. De multiples retards pendant le projet

Toujours en raison de l'emploi du temps très chargé des entreprises durant le projet, la réception des grilles permettant d'alimenter les bases de données expertes de l'expert artificiel a pris beaucoup de retard. Les dernières grilles ont été reçues à la mi-janvier, retardant donc les calculs de modèles.

4.2. Difficultés scientifiques

Pour la première fois, l'objet « jeux vidéo » a été soumis à l'expert artificiel. Ce dernier est potentiellement capable de créer des modèles pour tous types d'objets. De la construction des grilles à l'interprétation des résultats, en passant par la modélisation, les jeux vidéo se sont distingués par leur singularité, et les méthodes ont donc dûes être adaptées en conséquence.

4.2.1. Construire la grille O

Chaque objet étudié par l'expert artificiel doit être décrit à l'aide de trois types de grilles. Concernant les données expertes, nous avons pu discuter avec les professionnels du secteur de ce qui pouvait être évalué de manière subjective-objective (voir plus haut).

Le travail a été un peu plus compliqué concernant la description des propriétés objectives d'un jeu vidéo. Quels genres de données pouvaient comporter la grille O ? Comme nous l'avons dit plus haut, certaines propriétés peuvent facilement être identifiées comme des propriétés objectives : la classification PEGI, les périphériques employés pour jouer, etc.

Un autre genre de données aurait sans doute pu nous être utile mais n'a pas pu être exploité : les logs. En effet, le corpus d'origine étant uniquement constitué de jeux déjà commercialisés, nous n'avons pas pu poser de logs pendant les parties. Ces derniers nous auraient sans doute permis de récolter des données que nous avons envisagées au départ, comme par exemple les durées respectives des premières cinématiques, etc. D'autres propriétés, comme l'existence ou non de passages secrets dans le jeu, ne nous étaient pas réellement accessibles. Il aurait fallu, pour être certains des réponses à valider dans la grille, pouvoir demander aux game designers de ces jeux de nous renseigner sur ces sujets car, même en observant et testant nous-mêmes les jeux, nous risquions de passer à côté de ces données pouvant être qualifiées d'« objectives » en ne les découvrant pas par nous-mêmes lors de nos tests. Certaines données qui auraient pu servir à une évaluation des propriétés objectives du jeu ont donc dûes être écartées, non pas en raison de leur non-pertinence, mais par manque de certitudes à leur sujet.

¹⁰ Voir le powerpoint « Un expert artificiel pour aider à la décision durant le développement d'un jeu vidéo »

4.2.2. Les jugements subjectifs impossibles à expliquer

En étudiant la liste des modèles générés, nous nous sommes aperçus que certaines questions S, correspondant aux sorties de l'expert artificiel, ne trouvaient pas de modèles suivant robustes pour être considérés comme explicatifs (et donc prédictifs). A chaque fois, les modèles débouchaient sur des corrélations trop faibles. Dans le cas des jeux FPS, il s'agit des questions concernant le niveau de difficulté, la compréhension des règles du jeu ou encore l'intérêt ressenti pour l'histoire. Pourquoi ces appréciations subjectives n'ont-elles pas trouvé de règles explicatives ?

Nous pouvons envisager plusieurs raisons pour lesquelles ces sorties n'ont pu être expliquées. En voici des exemples :

- notre manière de poser la question a été inadaptée : l'évaluation est impossible à exprimer verbalement de manière aussi synthétique
- les questions qui sont en entrées ne sont pas explicatives de ces jugements subjectifs

4.2.3. Les entrées non utilisées

De même, nous nous sommes aperçus que certaines données SO ou O étaient soit jamais sollicitées pour expliquer une sortie, soit considérées comme nulles car ne présentant pas une variabilité suffisante pour être explicatives.

Là encore, nous avons forgé des hypothèses :

- certaines questions ont été mal posées et ont induit les game designers en erreur lors de leurs réponses
- certaines variables sont mal codées
- certains facteurs ne peuvent pas affecter la satisfaction des joueurs

4.2.4. Nuances¹¹

Si nous prenons en compte les trois types de jeux à la fois, nous observons que seule la question 7 de la grille S « L'ambiance visuelle de ce jeu : M'a tout à fait déplu <=> M'a beaucoup plu » ne donne aucun résultat. Les autres sorties trouvent des explications pour au moins un type de jeux. On peut donc considérer que seule la question S n°7 a réellement été mal posée. Cependant, si nous avons évalué les jeux Action-Aventure ou si nous avons construit les autres grilles différemment, nous aurions peut-être obtenu des résultats différents. La question S n°7 est peut-être pertinente malgré tout, mais nous n'avons pas pu le déterminer pendant la durée du projet.

4.3. Avancées

Malgré les difficultés mentionnées ci-dessus, l'utilisation de l'expert artificiel pour évaluer la qualité des jeux vidéo a permis plusieurs avancées qui ouvrent la voie pour des recherches futures.

4.3.1. Les résultats obtenus

La confrontation de l'expert artificiel au domaine des jeux vidéo a permis que soit testée pour la première fois une approche originale de mesure de la qualité des jeux vidéo, par la logique floue. Une méthode d'aide à la décision durant le développement d'un jeu vidéo¹² a pu être mise en place, pour anticiper sur les chances de succès d'un jeu, avant sa mise sur le marché.

L'étape de construction des grilles a favorisé non seulement la mise en commun des expertises des professionnels du secteur, mais aussi leur objectivation, puisqu'elles étaient jusqu'alors essentiellement

¹¹ Rappel : les grilles S et O sont communes à tous les types de jeux.

¹² Voir « Un expert artificiel pour aider à la décision durant le développement d'un jeu vidéo »

effectuées de manière informelle. Les ateliers organisés au Lutin ont favorisé les échanges entre professionnels qui ont pu, à cette occasion, se mettre d'accord sur des critères de qualité et un vocabulaire commun précis. Il convient de souligner les effets de constitution d'une culture professionnelle commune formalisée à travers ce type d'exercices, qui ne peut que contribuer à faire progresser la qualité d'ensemble de la production française, en donnant des outils de réflexivité sur leur activité aux entreprises de jeux vidéo.

Cette expertise a pu être simulée pour trois types de jeux. Dans chacun des cas, nous avons obtenus des modèles explicatifs et prédictifs pour plusieurs sorties¹³. Parmi les modèles robustes, nous avons sélectionnés les meilleurs d'entre eux. Cependant, des modèles un peu moins robustes ont été conservés pour au moins deux raisons :

- Certains modèles un peu moins performants utilisent en entrées des données que les modèles plus robustes n'utilisent pas. Ces modèles, même s'ils ne sont pas totalement fiables, donnent tout de même des indications, indiquent des tendances ; ce qui est toujours mieux que le hasard pour la prise de décision.
- Ces modèles pourraient peut-être s'avérer plus performants avec un corpus de jeu plus étendu.

	Jeux FPS	Jeux <i>casual</i>	Jeux RTS
Modèles déjà disponibles	Environ 70	Environ 35	Environ 35
État des calculs	En cours	En cours	En cours

Tableau récapitulatif du nombre de modèles robustes obtenus par type de jeux

4.3.2 Ce qui reste à accomplir

La plus grosse tâche qui reste à accomplir est de tester les modèles de l'expert artificiel sur des jeux en développement, puisque la méthode proposée aux professionnels est destinée à aider à la prise de décision durant le développement d'un jeu vidéo. A cette occasion, il sera sans doute possible d'améliorer la grille O en collaboration avec les professionnels du secteur.

Les grilles d'évaluation, et par extension les modèles, sont représentatifs d'une manière d'évaluer les jeux vidéo à un temps t. Une mise à jour régulière des grilles d'évaluation est à prévoir pour que l'expert artificiel simule l'expertise des professionnels de manière synchronique et produise des modèles qui soient toujours représentatifs de l'expertise des professionnels à un moment donné.

De même, il faudra alimenter les bases de références de l'expert artificiel en ajoutant les évaluations de nouveaux jeux. Plus on enrichira ces bases, plus l'expert saura expliquer, prédire et optimiser avec précision.

De la même manière encore, les appréciations subjectives des joueurs doivent continuer à être évaluées. Les goûts changent avec le temps et ce que l'on recherche dans un jeu vidéo peut également évoluer. Cette mise à jour des appréciations est nécessaire pour que les futurs modèles de l'expert artificiel puissent expliquer et prédire ce que les joueurs apprécient réellement dans un jeu vidéo. Nous considérons, en fonction de la vitesse d'évolution des jeux vidéo, que les modèles que nous avons extraits auront une validité de 3 ans du point de vue des goûts des joueurs, puis une validité de moins en moins grande. Ce qui devrait nous obliger à reprendre des tests tous les 3 ans, avec de nouveaux jeux et de nouveaux joueurs, sans pour autant devoir réviser toute la phase de construction des grilles ou de façon modérée. Mais l'ensemble de l'évaluation par les experts devrait être cependant reprise avec le nouveau corpus de jeux, pour pouvoir produire de nouveaux modèles adaptés. Cette opération doit donc être programmée une fois que nous aurons validé l'utilisation de l'expert artificiel durant l'année 2009.

Enfin, le tri et l'analyse des modèles doivent eux aussi continuer.

¹³ Voir les tableaux récapitulatifs par genre en annexe.

Annexes techniques

Grille O

Grille S

Exemple de questions de grille SO

(la liste complète des questions doit rester confidentielle pendant le temps de la validation auprès des entreprises)

Exemple de résultats pour un genre de jeux

(Voir fichier Excel joint)

(contrainte identique à la précédente)

Oui Non Pas de réponse

Rythme et histoire

- Q1 le joueur progresse dans une histoire pré-définie
- Q2 Il y a des passages secrets
- Q3 Il n'y a qu'une seule fin possible au jeu
- Q4 Le jeu permet au joueur de gagner de façon définitive
- Q5 Le joueur peut rencontrer des boss
- Q6 il y a un temps limité pour accomplir certains objectifs
- Q7 Il y a des casse-tête à résoudre
- Q8 on a un score
- Q9 il y a des power-up qui augmentent les capacités de l'avatar (pouvoirs magiques, armes, objets)
- Q10 il y a des power-down (qui pénalisent le joueur)
- Q11 il y a des cinématiques
- Q12 Le joueur a des quêtes à accomplir
- Q13 on peut choisir parmi plusieurs lignes de dialogue

Vision

- Q14 on voit l'avatar (dans RTS, avatar = armée)
- Q15 Une ocularisation interne est proposée au joueur (ce que voit l'avatar ex fps)
- Q16 Une ocularisation zéro est proposée au joueur (ce n'est pas le point de vue d'un avatar)
- Q17 On peut faire tourner la caméra autour de l'avatar
- Q18 Le joueur peut influencer sur la position de la caméra

Mouvement

- Q19 L'avatar peut se déplacer

Q20	Le joueur contrôle la position de l'avatar
Q21	Le joueur dirige plusieurs avatars simultanément
Q22	Le joueur incarne des avatars différents à tour de rôle
Q23	L'avatar peut sauter
Q24	L'avatar peut nager dans l'eau
Q25	L'avatar peut voler (sans véhicule)
Q26	Le joueur peut contrôler des véhicules volants
Q27	L'avatar a au moins deux vitesses de déplacement
Q28	L'avatar peut viser et tirer
Q29	L'avatar peut utiliser des moyens de locomotion (voiture, cheval, bateau) en mode jeu (vs cinématique)

Univers

Q30	L'environnement est photoréaliste
Q31	L'environnement est cartoon
Q32	L'environnement est abstrait
Q33	le contexte temporel est :Heroic fantasy
Q34	le contexte temporel est :Contemporain
Q35	le contexte temporel est :SF
Q36	le contexte temporel est :imaginaire
Q37	Autre contexte temporel
Q38	Le mécanisme du jeu utilise uniquement deux dimensions
Q39	le mécanisme du jeu utilise les 3 dimensions
Q40	Il y a un moteur de physique dans le jeu
Q41	Il y a des décors destructibles
Q42	Il y a une alternance jour / nuit
Q43	Le jeu fait partie d'une licence (cinéma ou épisode antérieur à cette version)
Q44	Il y a une carte
Q45	Il y a un système pour stocker des items collectés

Avatar

- Q46 Le joueur dirige un avatar
- Q47 Le joueur peut incarner un homme
- Q48 Le joueur peut incarner une femme
- Q49 Le joueur peut incarner une forme humanoïde (non humaine)
- Q50 Le joueur peut incarner une forme animale
- Q51 L'avatar peut se transformer
- Q52 L'avatar peut remporter des points d'expérience

Personnages Non Joueurs

- Q53 PNJ : aucun = 0, 1 = 1, 2 à 4 = 2, 5 à 10 = 3, 11 à 100 = 4, 101 et + = 5
- Q54 Un même personnage fournit régulièrement des informations qui permettent d'avancer
- Q55 Les IA parlent (on entend leur voix)

Violence

- Q56 Présence de sang dans le jeu
- Q57 Le joueur peut attaquer des PNJ non menaçants

Gestion des compétences du joueur et accompagnement dans son expérience de jeu

- Q58 Nombre de niveau de difficulté accessible dans le menu
- Q59 Le jeu commence par une page d'explication du jeu
- Q60 Présence d'un didacticiel
- Q61 Il est nécessaire de savoir lire pour jouer à ce jeu
- Q62 Il y a un journal de quêtes
- Q63 Des hint et conseils fournissent des indications pour avancer (de façon automatique)
- Q64 J'ai accès au profil de mon avatar
- Q65 Il y a un tableau des meilleurs scores / classement
- Q66 Je peux suivre ma progression sur un calendrier

- Q67 On peut avoir plusieurs sauvegardes
- Q68 Le jeu repose sur un motif de gameplay unique (répétition de la même action)
- Q69 certaines actions requierent de la dextérité (ex : servir dans wii tennis)
- Q70 il y a des modalitrés d'action à découvrir (combos,etc)

Personnalisation du jeu

- Q71 Le joueur peut modifier les caractéristiques de son avatar
- Q72 Le joueur a la possibilité de créer des niveaux
- Q73 Le joueur a le choix de l'environnement dans lequel va se dérouler sa partie
- Q74 Le joueur peut personnaliser les contrôles

Pour jouer le joueur utilise :

- Q75 une manette
- Q76 un clavier
- Q77 une souris
- Q78 un stylet
- Q79 un volant
- Q80 un micro
- Q81 une guitare
- Q82 une WiiMote
- Q83 un nuntchuk

Sauvegarde

- Q84 le joueur peut sauvegarder à tout moment
- Q85 la partie sauvegardée reprend exactement où le joueur a sauvegardé

Classification PEGI

- Q86 Age minimum requis par PEGI
- Q87 Classé "Gros mots"
- Q88 Classé "Discrimination"
- Q89 Classé "Stupéfiants"

Q90 Classé "Peur"
Q91 Classé "Jeu de hasard"
Q92 Classé "Teneur sexuelle"
Q93 Classé "Violence"

Contrôles

Q94 Pendant la partie, j'accède au menu depuis la manette
Q95 Pendant la partie, j'accède au menu depuis des boutons sur l'écran
Q96 Pendant la partie, j'accède au menu depuis le clavier
Q97 un système de menu arborescent permet d'effectuer des réglages
Q98 Il y a un pointeur
Q99 la jauge de vie est toujours visible
Q100 le score est toujours visible
Q101 le temps est toujours visible

Infos éco

Q102 Le jeu a reçu cette note de le part de Gameranking.com
Q103 Jeux vidéo.com note générale de la rédaction
Q104 Gamekult

Collaboratif

Q105 On peut jouer à plusieurs
Q106 On peut jouer en ligne

Grille S

Lutin Gamelab - Grille S

Testeur

Jeu

date

Instructions

Cocher l'une des cases en mettant un **X**
 Vous avez la possibilité de répondre entre deux propositions
 Vous avez la possibilité de ne pas donner de réponse

Comprendre le but du jeu a été	Très difficile		Plutôt difficile		Plutôt facile		Très facile	Ne veut/peut pas répondre	Q
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Q1
Comprendre les règles du jeu a été	Très difficile		Plutôt difficile		Plutôt facile		Très facile	Ne veut/peut pas répondre	Q2
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Q2
J'ai trouvé le jeu	Très ennuyeux		Plutôt ennuyeux		Plutôt captivant		Très captivant	Ne veut/peut pas répondre	Q3
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Q3
Pour commencer à jouer, les informations ont été	Très insuffisantes		Insuffisantes		Suffisantes		Parfaitement suffisantes	Ne veut/peut pas répondre	Q4
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Q4
L'utilisation des contrôles du jeu est	Très difficile		Plutôt difficile		Plutôt facile		Très facile	Ne veut/peut pas répondre	Q5
	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Q5

		M'a tout à fait déplu		M'a plutôt déplu		M'a plutôt plu		M'a beaucoup plu	Ne veut/peut pas répondre	
	Jouer à ce jeu									Q6
		M'a tout à fait déplu		M'a plutôt déplu		M'a plutôt plu		M'a beaucoup plu	Ne veut/peut pas répondre	
	L'ambiance visuelle de ce jeu									Q7
		Pas du tout intéressante		Peu intéressante		Plutôt intéressante		Très intéressante	Ne veut/peut pas répondre	
	J'ai trouvé l'histoire									Q8
		Certainement pas		Sans doute pas		Sans doute		Sûrement	Ne veut/peut pas répondre	
	Je rejouerais à ce jeu									Q9
		M'a tout à fait déplu		M'a plutôt déplu		M'a plutôt plu		M'a beaucoup plu	Ne veut/peut pas répondre	
	L'ambiance sonore et musicale									Q10
		Très difficile		Plutôt difficile		Plutôt facile		Très facile	Ne veut/peut pas répondre	
	J'ai trouvé le jeu									Q11

SP3 GameLab : Quick Reality Check™

Dominique Boullier, Audrey Lohard, Antoine Visonneau

Ce document présente Quick Reality Check, une méthode de test en ergonomie et gameplay des jeux vidéo en cours de développement mise au point dans le cadre du projet ANR Lutin Gamelab.

Le document est organisé en 4 parties correspondant aux sous projets (SP)

- Inventaire des méthodes
- Tests sur des jeux existants
- Tests sur des jeux en cours de développement
- Conclusion

Inventaire des méthodes

La bibliothèque d'articles de référence

La première partie du travail s'est concentrée sur la réalisation d'un inventaire des méthodes traditionnelles utilisées pour l'évaluation de l'ergonomie des interfaces. Nous avons mis en place un site de travail collaboratif (site spip) pour coordonner les efforts de Paris 6, Paris 8 et Rennes 2. Chaque article considéré comme important a fait l'objet de la rédaction d'une synthèse d'un paragraphe, l'article en question a été stocké dans le site collaboratif accompagné de la synthèse.



Dans un deuxième temps, nous avons entamé la production de fiches méthodes. Nous avons commencé par définir le plan de fiche, qui devait servir à l'élaboration des fiches méthodes. Ce plan permet d'avoir un canevas commun et d'évaluer les méthodes les une par rapport aux autres. Ce plan contient une série d'informations pratiques concernant l'effort nécessaire, le résultat que l'on peut atteindre grâce à cette méthode mais aussi, une liste d'articles de références (issues du site spip) que l'on peut consulter pour avoir plus d'information.

Plan complet d'une fiche :

- Ce qui est testé
- Ce qui n'est pas testé

- Points forts
- Points faibles
- Effort nécessaire
- Conseil pratiques
- Exemple de protocole
- Livrable
- Particulièrement adapté pour
- Sources et références
- Ont participé à l'élaboration de cette fiche

Ces fiches constituent donc un premier livrable du projet Gamelab. Il s'agit d'outils de travail opérationnels, directement applicables mais ces fiches constituent aussi les fondations du travail d'élaboration de nouvelles méthodes dans la mesure où nous avons travaillé en combinant et complétant des méthodes existantes.

Les 20 fiches méthodes

- Analyse de l'activité
- Auto Confrontation
- Recueil Des Mesures Physiologiques
- L'analyse Heuristique
- Cognitive Walkthrough
- Entretiens
- Focus Group
- Guidelines
- Analyse des Logs
- Playtests Sur Prototypes Papier
- Playtests Scénarises Sur Prototypes Interactifs
- Playtests Ouverts Sur Prototypes Interactifs
- Questionnaires
- Thinking Aloud – Protocoles Verbaux
- Oculometrie
- Digital Script Doctor

Enseignements tirés pour la mise au point des méthodes

L'inventaire des méthodes nous a permis de tirer de nombreux enseignements. Nous avons pu analyser les méthodes les plus récentes développées chez Bungie (Halo3) qui reposent en grande partie sur l'analyse de Log. Pour compléter cette approche, nous avons souhaité conserver une part de verbalisation dans le protocole de test (par exemple, pour identifier la cause d'un problème lorsqu'il survient). Durant l'inventaire des méthodes, nous avons pu analyser les différentes approches et, en comparant l'effort nécessaire à déployer pour chaque méthode, nous avons été en mesure d'identifier les briques (fragments de méthode) compatibles avec les objectifs du projet. (cost/benefit analysis).

SP3 Expérimentation sur des jeux existants

L'expérimentation sur des jeux existants a été l'occasion de mettre en pratique les méthodes analysées dans l'inventaire des méthodes.



En résumé la mise au point de la méthode à porté sur :

- 2 jeux Aventure/action (10 testeurs, 5 par jeu)
12 indicateurs, interruptions pour verbalisation, questionnaire standard Gamelab 11 questions + questionnaire spécifique – enregistrement vidéo. Enregistrement vidéo et annotation avec Dartfish. Trois expérimentateurs.
- 2 jeux de type FPS (3 testeurs, qui jouent aux 2 jeux)
14 indicateurs, 9 exercices, questionnaire *exagéré* 11 questions, questionnaire spécifique. Enregistrement vidéo et annotation avec Dartfish. Annotation des évènements en temps réel. Trois expérimentateurs.
- 3 jeux casuels (3 testeurs ont joué aux 3 jeux).
7 variables observées, jeu libre+6 exercices, questionnaire standard gamelab 11 questions + questionnaire de comparaison. Filmé avec la caméra de l'eye tracker mobile. Deux expérimentateurs.

Cette étape a présenté de nombreuses difficultés mais a finalement permis de poser les fondations de la méthode QRC. Il y a eu deux difficultés principales à surmonter : la dimension exploratoire (1) et la difficulté pratique liée au côté boîte noire du jeu commercialisé(2).

La dimension exploratoire a présenté quelques difficultés pour les tests : en effet, nous n'avions pas accès aux gamedesigners des jeux que nous allions tester, nous n'avions donc pas de commanditaire qui nous guidait sur les points du jeux sur lesquels il désirait avoir une réponse. Nous avons donc du procéder en deux temps : émettre des hypothèses et établir un protocole de test qui valide ou infirme ces hypothèses.

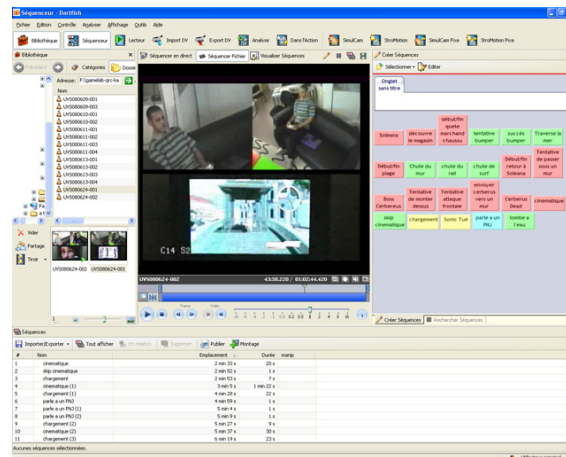
L'enseignement que l'on retire de cette étape se situe à plusieurs niveaux. A un premier niveau, il apparaît qu'il faut commencer toute démarche de test par une phase d'inspection de l'ergonomie et du gameplay du jeu. A un second niveau, cette phase a permis de tester un grand nombre d'indicateurs et d'évaluer leur pertinence. A titre d'exemple au travers de différents tests, nous avons cherché à identifier ce qui peut-être mesuré dans un jeu.

- **Chronométrer** : le temps nécessaire pour accomplir une tâche, pour terminer le niveau. Cette information est relativement aisée à obtenir, cependant, il est difficile d'en tirer des conclusions dans l'absolu. C'est donc un critère qui acquiert de la valeur dans la comparaison.
- **Nombre de tentatives** : si le nombre de tentatives est trop élevé, le joueur commence à ressentir de la frustration, cet indicateur permet de mettre en lumière un problème d'équilibrage de la difficulté dans le jeu. Par exemple, ce point a permis de mettre en lumière la différence entre difficulté ludique et difficulté de compréhension. (une difficulté ludique génère du challenge, du fun, tandis qu'une difficulté de compréhension génère de la frustration). Cet indicateur est à mettre en relation avec les indicateurs de succès ou d'échec à une tâche ou un objectif du jeu.
- **Nombre d'ennemis tués** : nous cherchions à évaluer l'intensité du jeu. Il est très difficile de tirer des conclusions sur la qualité du jeu en fonction de ce seul critère, mais cela donne une idée précise de l'intensité d'une séquence (ce n'est pas pareil de tuer 30 ennemis ou 3). Une séquence intense peut appeler une séquence plus lente pour obtenir un meilleur rythme et conserver le joueur immergé dans l'expérience du jeu.



- **Score** : pour les casual games, le score est un indicateur synthétique qui mesure la performance (en combinant parfois plusieurs critères). Ce qui est intéressant ici, c'est de mesurer la progression dans le temps. Si un joueur progresse, d'une partie à l'autre, son score reflétera cette progression et il est probable qu'il passe un bon moment. Dans tous les cas, on devra valider ces hypothèses par de la verbalisation.
- **Petits exercices** : pour évaluer une primitive du gameplay (ex combat ou déplacement), nous avons élaboré une série de petits exercices que les joueurs devaient effectuer en début puis en fin de session. Nous voulions évaluer la progression (le joueur a-t-il progressé dans sa maîtrise des contrôles ?)
- **Comparaison des jeux** : en fin de partie, on demande au joueur d'évaluer le jeu selon une grille. Ensuite, il joue à un autre jeu comparable et on lui demande d'évaluer le second jeu avec la même grille. Enfin, on lui demande de produire plusieurs classements des jeux en fonction de critères (graphisme, son, prise en main etc...) Cela permet d'inscrire le jeu testé dans le contexte concurrentiel, ce qui correspond à la réalité commerciale du jeu (il est positionné en rayon à côté des jeux concurrents).
- **L'interruption** : durant le jeu, on demande au joueur de mettre sur pause le temps de répondre à une question, par exemple une question de compréhension sur l'interface. Cette technique fonctionne bien, il faut avoir une check list pour vérifier que tous les joueurs auront répondu à toutes les questions. Pendant que le testeur joue, il faut attendre le bon moment adéquat pour vérifier la compréhension de tel ou tel élément d'interface.

- **Questionnaires** : dans le cadre du montage de l'expert artificiel, nous avons élaboré un premier questionnaire pour l'évaluation des jeux par les joueurs. Durant la phase de test sur les jeux en cours de développement, nous avons, dans un premier temps réutilisé ce questionnaire. Nous avons reformulé plusieurs questions dans l'optique d'obtenir un questionnaire plus contrasté, qui devait permettre de faire ressortir plus clairement les lacunes d'un jeu. Les expressions étaient volontairement exagérées (par exemple : « c'est le meilleur jeu auquel j'ai joué » ou « les graphismes sont époustouflants »). Bien que les résultats soient encourageants, nous avons décidé de revenir au premier questionnaire afin de pouvoir se laisser la possibilité de comparer les réponses du joueur avec l'information recueillie dans le cadre du montage de l'expert artificiel. Cette piste n'a pas été explorée mais cette base de données d'appréciation subjective sur le corpus de jeu reste un actif du gamelab sur lequel il sera possible de capitaliser par la suite.



- **La vidéo** : les premiers tests ont largement utilisé les équipements vidéo du laboratoire lutin. Nous cherchions à avoir une image qui assemble trois flux vidéo : le visage du joueur, le contexte, l'écran de la console. Ce dispositif est relativement satisfaisant dans la mesure où l'information collectée est de bonne qualité, et facilement réutilisable pour une analyse ultérieure à la séance de test. Cependant, l'enregistrement vidéo pose le problème du temps d'analyse. Si on considère qu'il faut 3 heures d'analyse pour une heure d'enregistrement, et qu'il y a 10 testeurs (X1H) cela signifie 30 heures d'analyse vidéo, hors synthèse, ce n'est pas compatible avec le rythme de l'industrie du jeu vidéo. Nous avons essayé de solutionner ce problème en utilisant plusieurs techniques d'annotation des séquences vidéo (en utilisant la suite Dartfish par exemple). Bien que ce logiciel soit de bonne qualité, l'annotation de vidéo en temps réel est une tâche très fatigante pour l'opérateur et nécessite d'immobiliser une personne supplémentaire pour toute la durée du test. Enfin, pour les raisons de fatigue évoquée ci-dessus, il faut limiter le nombre d'éléments à relever. Pour toutes ces raisons, nous avons décidé de ne pas faire reposer la méthode QRC sur l'analyse vidéo.

Difficultés et limites :

- N'ayant pas accès au code source du jeu, nous ne pouvions pas automatiser ces mesures. Il a donc fallu beaucoup de travail pour compter le nombre d'ennemis tués, chronométrer le nombre de tentatives, etc. Cela rend ce type de mesure difficilement applicable aux développeurs de jeu vidéo si cela n'est pas automatisé. La méthode QRC que nous proposons repose donc sur l'utilisation extensive de capteurs intégrés au jeu pour automatiser toutes les mesures des actions des joueurs.

- Il est difficile de comparer précisément les jeux. Lorsqu'on chronomètre le temps qu'il faut pour effectuer une action d'un jeu à l'autre, c'est rarement exactement la même chose qu'on mesure (les décors sont différents, il est donc difficile de garantir qu'une distance dans le jeu A est équivalente à une distance dans le jeu B par exemple). Il est donc préférable de se limiter à des comparaisons exprimées explicitement par le joueur.

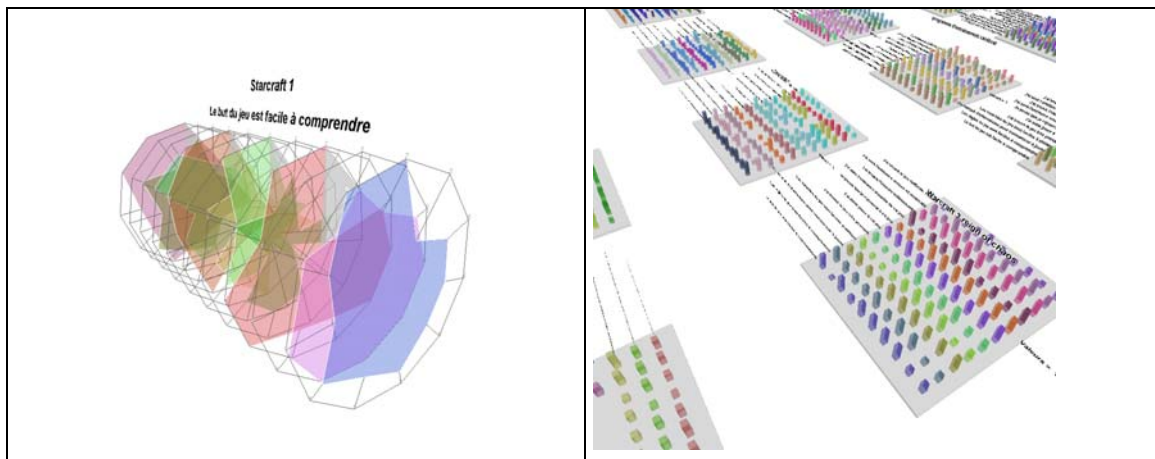
Conclusion : une structure générique de test « QRC »

Le protocole que nous avons mis au point repose en grande partie sur les enseignements tirés de cette étape du projet. De façon générique, le protocole QRC se compose des étapes suivantes

- Prise en main (5mn)
- Exercices - première itération (10mn)
- Phase de jeu libre avec quelques interruptions (environ 25mn)
- Exercices - 2eme itération (10mn)
- Questionnaire standard Gamelab (5mn)
- Entretien ouvert (5mn)

Visualisation des résultats

Un axe de recherche exploratoire a été » conduit sur la restitution des résultats des tests. L'objectif était de faciliter la compréhension des données recueillies. A titre d'exercice, nous avons essayé de visualiser les données destinées à alimenter l'expert artificiel. Le résultat n'a pas été convaincant mais ce travail a servi de base à l'élaboration des logiciels de restitution graphiques de la méthode QRC. Ces graphiques ont été produits avec l'environnement de développement Processing.



Expérimentation sur des jeux en cours de développement



Mise au point du protocole

L'expérimentation et mise au point de la méthode QRC a été réalisée sur Crusaders, un jeu en cours de développement par le studio Kylotonn, membre Capital Games.

Le protocole repose largement sur le travail effectué sur les existants, cependant, le fait de travailler sur un jeu en cours de développement a révélé de nouvelles questions et de nouveaux axes de travail présentés ci-dessous.

- **Axe 1 : Le gamedesigner et le publisher.**

Pour fournir une méthode de test adaptée à l'industrie des jeux vidéo, il a été nécessaire de passer du temps à bien comprendre les acteurs et leurs problématiques respectives. Dans un premier temps, il s'agit de comprendre la problématique du gamedesigner. Le gamedesigner est le créatif qui porte le jeu. Il a pour objectif de concevoir une œuvre unique qui saura captiver le joueur. Il entretient donc une relation affective forte avec son jeu et il est très important de bien comprendre ses intentions créatives pour que le test soit orienté vers l'amélioration de son concept, plutôt qu'un test-sanction qui chercherait à démontrer que l'objectif n'est pas atteint. Il faut donc consacrer du temps à comprendre l'intention créative puis fournir un travail de conseil pour imaginer le protocole de test qui saura révéler les problèmes tout en donnant une orientation constructive à la démarche. Par exemple, si l'analyse ergonomique du prototype donne à penser qu'il y a un problème au niveau des contrôles, il peut être intéressant de travailler avec le gamedesigner en amont pour développer une seconde option. Ainsi, le test pourra mesurer objectivement si la solution proposée est plus performante que la solution initiale. La difficulté de l'exercice consiste à évaluer ce qui est une difficulté ludique, voulue par le concepteur par opposition à une difficulté collatérale qui n'est pas issue d'une volonté créative. Enfin, le fait de demander au designer de décrire précisément son intention créative permet d'imaginer des exercices qui valideront ou invalideront les solutions proposées. (exemple : combien de fois penses tu qu'un joueur doit utiliser la parade dans une partie de 30 minutes ?). Ce type de question permet d'identifier une variable à observer durant le test. Le résultat apportera des réponses précises au gamedesigner sur le niveau d'accomplissement de ses objectifs de conception. On voit donc ici que la mise en place du test se fera de façon itérative. Il doit y

avoir un échange entre le gamedesigner qui est le garant de concept créatif et le responsable du test qui va imaginer un protocole adapté.

Un autre facteur important de contexte est le publisher. Nous n'avons pas travaillé sur un jeu en cours de production mais sur un prototype que Kylotonn développe avec l'objectif de convaincre un publisher d'investir pour produire l'intégralité du jeu. L'utilisateur final du prototype a donc un profil particulier – très différent du casual gamer par exemple.

- **Axe 2 : identifier les fonctions clés**

Dans le cas de Crusader, nous avons identifié de nombreuses fonctions clés ou features qui forment le core gameplay et/ou les éléments à forte valeur ajoutée, facteur de différenciation. Attendu qu'un joueur ne peut pas les connaître avant de jouer, nous avons recommandé de développer un tutoriel – qui avait pour objectif de transmettre la connaissance des mécanismes du jeu avant de tester la maîtrise de ces mécanismes par le joueur. Ce tutoriel n'a pas pour objectif d'être intégré au jeu final tel quel. Il s'agit d'une approche permettant finalement de faire des tests unitaires sur l'ergonomie des fonctionnalités clés du jeu. De plus, avec le tutoriel, nous étions assurés que la phase de transmission de la connaissance du jeu vers le joueur avait eu lieu, il devenait donc pertinent de mesurer ensuite, dans une phase de jeu libre l'utilisation réelle des features par le joueur. (capacité à réappliquer les connaissances).

- **Axe 3 core gameplay**

Crusader est un jeu d'action aventure de type Hack/slash. Une dimension clé du jeu est la gestion des combats. Nous avons donc recommandé d'implémenter une mini arène de combat (practice) dans lequel le joueur affronte un ennemi. Ici donc pas d'histoire, l'objectif est de pouvoir travailler la mise au point de la mécanique de combat. Dans le cadre du test, ce développement permettait aussi d'évaluer la progression du joueur dans une situation exactement similaire avant et après la phase de jeu libre (soit 30 minutes plus tard)

- **Axe 4 Level design**

Kylotonn avait développé un niveau représentatif de ce qu'allait être l'ensemble du jeu (graphiques/son/intrigue). Pour la mise au point de ce niveau (level design), nous avons proposé de mesurer :

- l'utilisation précise des fonctions clés du jeu (statistiques : features / controls)
- les parcours dans le niveau (trace du joueur en XY dans la map)
- la résolution des énigmes (succès/échec)

Les séances de test

Description du protocole

- Prise en main (5mn)
- Tutorial (19 étapes correspondants aux compétences élémentaires et avancées)
- Exercice : le joueur affronte 3 ennemis dans une arène
- Phase de jeu libre avec quelques interruptions (environ 25mn)
- Exercice : le joueur affronte 3 ennemis dans une arène (2eme itération)
- Questionnaire standard Gamelab (5mn)
- Entretien ouvert (5mn)

Il ya eu un pré-test pour ajuster le protocole puis 5 tests (environ 1H30 à chaque fois) effectués au Laboratoire Lutin. Le gamedesigner a assisté au premier test.

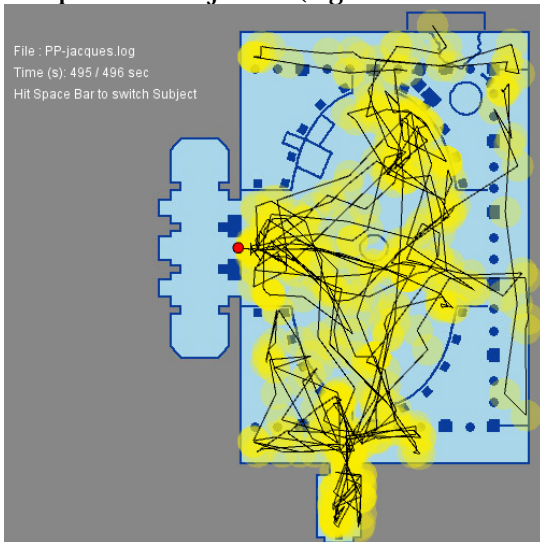
Pendant le test, l'expérimentateur prend des notes (essentiellement la verbalisation par les joueurs, les réponses (succès/échec) aux questions posées pendant les interruptions.

Game Doctor : analyse des résultats.

Les notes des tests sont réorganisées par thème plutôt que par séance de test afin de faire ressortir les thèmes problématiques ou les succès rencontrés.

Les logs (recueil automatique de données) sont reformatés pour pouvoir être utilisés dans les logiciels de restitution graphique. Lorsque tout est prêt, le responsable du test organise la séance de gamedoctor : on analyse toutes les données avec le gamedesigner pour tirer des conclusions et imaginer des solutions. Pour cette étape, les supports de travail sont :

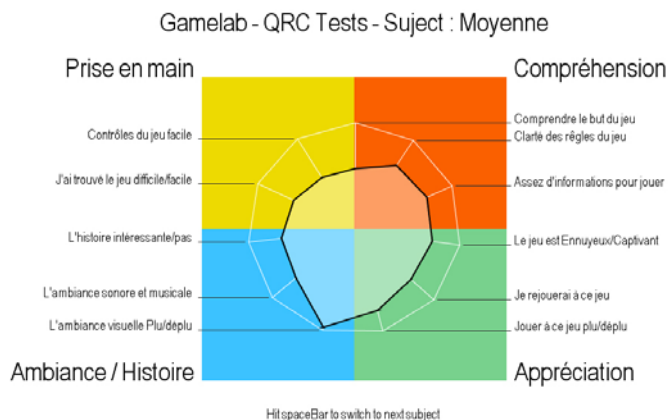
- Les notes prises lors des tests (réorganisées par thème)
- Le parcours du joueur (logiciel de restitution graphique)



- La table des événements (visualisation graphique des statistiques d'utilisation des features)



- L'appréciation subjective par le testeur (visualisation graphique des résultats du questionnaire).



Chaque point soulevé par les tests a fait l'objet d'une discussion : le problème est-il réel, faut-il le résoudre (priorité) ? Quelles sont les solutions. A la suite de ces tests, le studio a fait évoluer son prototype.

Conclusion

Si l'on regarde l'ensemble de la démarche, le projet a permis de faire une avancée significative dans les méthodes de test de l'ergonomie et du gameplay des jeux vidéo en cours de développement. Nous avons désormais un protocole composé d'une batterie de briques de méthodes applicables à de nombreux jeux dans quatre genres représentatifs du jeu vidéo.

La mise au point de la méthode a porté sur 8 jeux et a fait intervenir 21 testeurs, sans compter les pré-tests qui ont été effectués sur chaque protocole. Les premières approches mobilisaient jusqu'à 3 expérimentateurs pour dérouler le protocole qui était assez lourds (1) conduite du test, (2) annotation vidéo et (3) prise de note en temps réel.

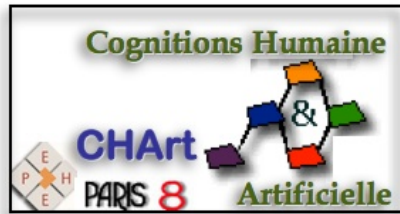
La version finale de la méthode repose sur une ou deux séries de 5 tests et ne requiert qu'un seul expérimentateur. Il y a donc eu un gain réel en termes de productivité (besoin de moins d'expérimentateurs) qui s'explique par l'automatisation du recueil de données (installation de capteurs dans le jeu) et le développement d'outils de restitution graphique qui réduisent considérablement le temps d'analyse. La démarche a permis de montrer les avantages de d'une approche plus légère que les méthodes traditionnelles (ex : pas d'utilisation de la vidéo) qui est plus alignée avec les contraintes de délais tendus de l'industrie du jeu vidéo. Cependant, la méthode n'est pas exhaustive, et pourra être complétée par d'autres protocoles selon les spécificités des jeux et leur avancée dans le cycle de développement. Il faut aussi noter que temps de traitement des résultats a été largement automatisé, mais il reste une marge de progression pour avoir la chaîne complète. Plus d'automatisation permettrait également d'obtenir un retour plus précis des testeurs : si on était en mesure d'obtenir les résultats graphiques en temps réel, on serait en mesure d'interroger le testeur en fin de séance à la lumière de ces informations.

Prochaines étapes

La prochaine étape consiste à faire connaître cette méthode auprès des acteurs du jeu vidéo en Ile de France (par l'intermédiaire de Capital Games) pour une mise en application sur d'autres jeux en cours de développement. Pour favoriser la dissémination et la communication des méthodes, nous avons développé une page qui présente les méthodes et leurs bénéfices pour les entreprises (accessible depuis le site du laboratoire Lutin <http://www.lutin-userlab.fr>)



The screenshot shows the website for Lutin GameLab, an observatory for video games, ergonomics, and gameplay. The page features a navigation menu with links for Accueil, Plateforme, Recherche, Services, Projets, Membres, Partenaires, and Info pratiques. The main content area is titled 'Lutin GameLab observatoire des jeux vidéo, ergonomie • gameplay'. It includes three sections: 'Description du projet' (Lutin GameLab: an ANR-funded research project for operational evaluation of video game quality), 'GameRoom' (a dedicated testing environment with multiple consoles and cameras), and 'Inventaire des méthodes' (a list of methods for ergonomic evaluation). The page also includes logos for Lutin Userlab, ANR, and Capital Games.



Lutin Game Lab

SP3 : Méthodes en ergonomie des interfaces

Livrable de l'étude des méthodes en ergonomie des interfaces, dont celles qui sont basées sur le recueil de données comportementales, oculométriques, et physiologiques, des jeux vidéos : présentation de la méthode, des différents tests utilisateurs effectués et limites du travail effectué

Renseignements de référence	
Identifiant de projet	ANR LUTIN GAMELAB
Auteur(s)	Laure Léger, Nicolas Fouquereau, Florent Levillain, Charles Tijus
Partenaire(s)	LUTIN, LIP 6
Participants	Julie Benoist, Ivan Bigorgne, Nicolas Fouquereau, Stéphane Hélène, Laure Léger, Florent Levillain, Charles Tijus, Alexandra Wang

Version	Date	Motivation
Finale		
Etat	2 /03/2009	Validé

Résumé

Sous Projet 3 - Méthodes en ergonomie des interfaces

Laboratoire Cognitions Humaine et Artificielle (ex. Cognition & Usages)

1. Enjeux et problématique, état de l'art

L'étude des IHM pour le jeu vidéo est un nouveau champ de recherche. Pour l'élaboration de préconisations aptes à améliorer la qualité ergonomique du jeu vidéo, deux axes de recherches sont possibles. Le premier est d'adapter les méthodes et outils qui existent pour les IHM. Le second est de développer des méthodes et des outils spécifiques en s'appuyant sur l'existant. Ces deux axes de recherche ont été retenus. Le premier axe, objet d'un rapport intermédiaire, permet au concepteur d'appliquer au jeu vidéo les méthodes classiques d'évaluation des IHM. Le deuxième axe a été développé autour de deux techniques, l'oculométrie cognitive, la mesure physiologique, et de leur intégration, tirant partie des travaux qui utilisent la première technique pour étudier *l'attention*, la *recherche d'information* et la *profondeur du traitement de l'information perçue*, et ceux qui utilisent la deuxième pour *l'émotion* et *l'engagement dans la tâche*, c'est-à-dire pour des variables comportementales liées au Gameplay. Ces techniques offrent une granularité plus fine que les méthodes classiques, permettent de repérer des transitions d'états du jeu, et l'impact d'événements de jeu.

2. Matériel et méthodes

Pour *l'axe 1*, les méthodes classiques, à base de questionnaires portant sur des heuristiques, ont été recensées et intégrées. Pour *l'axe 2*, partant des heuristiques de l'axe 1, les traces oculaires et physiologiques (activité électrodermale, fréquence cardiaque, fréquence respiratoire) ont été analysées pour définir des indicateurs de jouabilité : *l'utilisabilité*, la *compréhensibilité*, *l'apprenabilité*, et le *degré d'immersion*, en considérant la prise en main, l'exploration et la localisation de l'information et la typicalité du jeu. L'observation est écologique, dans le contexte naturel de jeu, selon des conditions de variation expérimentale, et avec des procédures de segmentation des traces des épisodes de jeu.

3. Résultats

Un *premier résultat* est la mise en relation des variables expérimentales (type de jeu, expertise, situation du jeu) avec les variables oculaires et physiologiques. Contrairement aux méthodes classiques, subjectives et statiques (axe 1), nos méthodes permettent ainsi d'évaluer tout ou partie d'un jeu vidéo à partir de données objectives liées à la dynamique du jeu. Par exemple, pour mesurer l'immersion, on met en relation l'état du jeu et l'état du joueur, en considérant les alternances entre périodes de préparation à l'action, d'engagement dans l'action et de relâchement. Un *deuxième résultat* est que nous n'avons pas trouvé de correspondance univoque entre les données subjectives et les mesures objectives.

4. Discussion sur le degré de réalisation des objectifs initiaux, les verrous restant à franchir, les ruptures, les élargissements possibles, les perspectives ouvertes par le projet, l'impact scientifique, industriel ou sociétal des résultats.

Les résultats sont encourageants. Toutefois, l'instrumentation de l'observation et la masse de données ont rendu la recherche très lourde et conduit à limiter les types de jeux considérés. La généralisation à d'autres types de jeux est nécessaire, tout comme la nécessité de comprendre la relation entre variables subjectives et objectives (travaux du LIP6), en faisant aussi appel à l'évaluation subjective experte des jeux.

5. Conclusions et recommandations sur l'exploitation et la dissémination des résultats.

Les résultats sont issus d'une recherche exploratoire posant les fondations d'une méthode objective d'évaluation de la qualité du jeu vidéo. Elle a été utilisée pour mesurer les différences entre TVHD et TVSD (www.tdf.fr/592.html). Elle pourra participer à la définition d'une méthode pour l'attribution d'un label de Qualité du Jeu Vidéo. Des articles sont en préparation pour la revue *Media Psychology*.

L'objectif de SP3 est de proposer une méthode d'évaluation ergonomique des jeux vidéo et des expérimentations avec des jeux existant déjà sur le marché. Cette méthode d'évaluation est basée essentiellement sur l'analyse des données comportementales, oculométriques et physiologiques et l'idée est de voir dans quelle mesure les analyses, qu'on effectue habituellement sur des interfaces du type logiciel ou site Web, peuvent s'appliquer à une interface dynamique et interactive comme celles des jeux vidéo.

DONNEES OCULOMETRIQUES

L'ANALYSE DES MOUVEMENTS OCULAIRES POUR LES JEUX VIDEOS

Les études portant sur l'évaluation ergonomique des jeux vidéo sont très récentes et très peu nombreuses. Celles qui utilisent les mouvements oculaires le sont encore plus. De ce fait il nous fallu mettre au point une méthode qui soit applicable aux jeux vidéo. Cette méthode s'appuie essentiellement sur une comparaison expert-novice en regardant dans quelles mesures les patterns des novices rejoignent les patterns des experts qui sont censés être optimum pour remplir les différentes actions nécessaires dans le jeu. Pour identifier quels indicateurs analysés et comment interpréter ces analyses nous nous sommes appuyés sur les méthodes déjà utilisées dans l'ergonomie des interfaces type logiciels ou site Web. Ainsi cette méthode se base essentiellement sur l'analyse des fixations oculaires (surtout leur durée) et des amplitudes saccadiques.

Etant face à un environnement dynamique, l'analyse de ces paramètres sera plus compliquée que pour l'utilisation d'un logiciel ou d'un site Web. En effet il ne sera pas possible de définir à priori des zones de l'interface comme celles qui ont trait au menu, à l'aide contextuelle..., sauf si ces zones sont toujours visibles à l'écran quel que soit le moment du jeu. Ce qui n'est pas le cas par exemple de l'aide contextuelle, qui n'est visible que lorsqu'on l'active. Ce qui fait que les analyses devront être globales avec des segmentations fait en termes de temps et non en termes d'événements. En effet, le but de ce sous-projet est d'obtenir une méthode qui pourra être utilisée pour évaluer des jeux en cours de développement. De ce fait les tests utilisateurs qui comprennent à la fois les passations et les analyses qui résultent de ces passations devront être rapides. Une analyse se basant sur les événements effectués pendant le jeu nécessiterait un découpage vidéo et une indexation manuelle de ces différents événements. Un tel découpage n'est possible qu'à la suite de la passation et peut se révéler très long : il faut repasser toutes les vidéos enregistrées lors des passations : augmenter les temps de jeu et le nombre de joueurs par passation augmente sensiblement les temps de découpage. De ce fait il est très difficile de fournir des résultats très rapides, avant la sortie du jeu sur le marché par exemple, ou pour répondre à une demande particulière.

Pour répondre à cette demande de rapidité, un logiciel d'analyse globale des mouvements oculaires a été mis au point (Lutin MOC). Ce logiciel peut analyser simultanément plusieurs fichiers présentant les données oculaires. L'analyse fournit pour tous les fichiers traités plusieurs indices comme le nombre de fixations, les durées de ces dernières, les amplitudes saccadiques. L'analyse peut porter sur l'intégralité de la passation ou sur un moment déterminé du jeu, les indices peuvent être analysés par zones d'écran ou sélectionnés en fonction des événements utilisateur.

Cette application fut développée en Visual Basic Application sous Microsoft Excel, par soucis

de simplicité et d'évolutivité des fonctions au fil du temps et des nouvelles analyses envisagées. Une version indépendante est en développement actuellement.

Grâce à ce logiciel, nous pouvons en un temps très court (dépendant du nombre de fichiers à traiter) traiter les données oculométriques qui sont ainsi prêtes pour une analyse très peu de temps après toutes les passations. Sans ce logiciel pour chaque passation, il faudrait compter plusieurs heures pour extraire des fichiers oculaires les indicateurs nécessaires pour les analyses. Avec ce logiciel pour par exemple 20 fichiers oculaires il faut compter moins de 10 minutes pour traiter les données. Le revers de cette automatisation et globalisation du traitement des fichiers oculaire, bien utiles, est la non-segmentation au cas par cas (passation par passation) des fichiers. Ce n'est pas que le logiciel ne le permet pas, mais le temps pour effectuer ce type de segmentation est très long ; il faut déjà repérer pour chaque joueur les marqueurs de la segmentation. De ce fait même avec ce logiciel, le traitement prendra plus de temps : visionnage de la vidéo de la passation pour repérer les moments à segmenter et traitement individuel de chaque fichier oculaire. Cette contrainte de découpage permet d'expliquer pourquoi la méthode que nous avons choisi de développer pour l'évaluation ergonomique reste très globale.

Dans une première partie, nous présentons les principes de cette méthode qui va tester les jeux vidéo sur plusieurs critères : apprenabilité, compréhensibilité, facilité de prise en main du jeu, typicalité du jeu, localisation de l'information, exploration de l'interface. Puis nous présentons trois expérimentations qui nous ont permis d'évaluer notre méthode.

METHODE ENVISAGEE

DESCRIPTION DES MOUVEMENTS OCULAIRES

Deux grands types de mouvements oculaires sont répertoriés : les fixations et les saccades. Chacun de ces types répond à une fonction précise : les fixations assurent la stabilité du regard sur un point A de l'environnement et les saccades assurent l'orientation du regard d'un point A vers un point B. Les mouvements oculaires sont une succession de fixations et de saccades, les unes précédant les autres.

Les fixations : stabilisation du regard

La stabilisation du regard sur la scène visuelle est assurée par un système de fixation qui assure le maintien de l'objet regardé dans la fovéa (partie de l'œil qui permet une prise d'information visuelle avec le maximum de précision). Elle s'effectue par l'intermédiaire de mouvements de très faibles amplitudes (micro-saccades) pour maintenir le point de fixation, constitué d'un ensemble de points échantillonnés distants de moins 0.5 degrés d'amplitude les uns des autres sur une durée minimale de 100ms.

Une fixation se caractérise principalement par sa durée qui est censée dépendre de la difficulté relative du traitement de l'information fixée. Plus cette information est difficile à traiter, plus le temps de fixation sur cette partie de la scène visuelle sera long (Just & Carpenter, 1980). Dans une tâche de lecture, une fixation se caractérise par un temps de pause d'environ 250 ms sur la partie du texte considérée. Ce temps de pause correspondrait à l'analyse et au filtrage des informations visuelles, à leur intégration aux informations précédemment traitées et à la préparation de la prochaine saccade (Jacobs & Lévy-Schoen, 1987).

Les saccades : orientation du regard

L'orientation du regard sur une scène est assurée par le système saccadique. La saccade consiste en un mouvement rapide et conjugué des yeux. Le système saccadique permet l'orientation du regard d'un point A vers un point B. Les saccades assurent un changement volontaire de direction de l'axe visuel dans l'espace en réponse à des signaux externes sur la position d'une cible dans le champ visuel ou à des codes internes. Leur rôle serait de permettre d'amener l'image de la cible sur la fovéa. Une saccade peut être décrite selon plusieurs caractéristiques : son amplitude (la distance séparant les points A et B), sa durée (entre 30 et 60 millisecondes) et sa vitesse (rapport entre l'amplitude et la durée).

Les saccades sont souvent jugées trop rapides pour permettre une quelconque prise d'information.

Dans la lecture, les saccades peuvent être de deux types :

- soit progressives, lorsque le regard se dirige sur une partie de la scène visuelle qui n'a pas encore été explorée,
- soit régressives, lorsque le regard revient sur une partie de la scène qui a déjà été explorée (retour en arrière).

Le parcours oculaire d'un individu sur une scène visuelle se caractérise donc par une succession de saccades et de fixations, les unes précédant les autres. Cet ensemble de saccades et de fixations nous permet de reconstruire en temps réel le parcours oculaire (ou scanpath, en anglais) de l'individu sur la scène visuelle.

ANALYSE DES MOUVEMENTS OCULAIRES DU JOUEUR

Dans l'évaluation ergonomique des jeux vidéo, trop peu d'études analysent les mouvements oculaires des participants lors d'une séance de jeu. Pourtant l'extraction visuelle de l'information est importante pour ce type d'activité. Ainsi analyser les mouvements oculaires lorsqu'un individu est en train de jouer peut nous permettre d'étudier dans quelles mesures le design visuel affecte le gameplay, d'analyser la perception des informations transmises au joueur pendant l'action.

Les quelques études qui utilisent l'enregistrement des mouvements oculaires étudient la possibilité de synchroniser l'enregistrement des mouvements oculaires avec le flux vidéo (Kenny, Koesling, Delaney, McLoone & Ward, 2005, Sennersten & Lindley, 2008), d'utiliser la position du regard comme input pour le jeu vidéo (Jönsson, 2005), de décrire les patterns oculaires lors de l'apprentissage dans un jeu vidéo (Alkan & Cagiltay, 2007), ou selon le type de jeu (El-Nasr & Yan, 2006). Un seul article tente d'expliquer l'intérêt de l'analyse de mouvement oculaires pour l'évaluation de l'utilisabilité des jeux vidéo sans pour autant mener de tests avec enregistrement des mouvements oculaires (Johansen, Noergaard & Rau, 2008).

Pourtant les méthodes oculométriques pour évaluer l'utilisabilité des interfaces hommes – machine existent (Goldberg & Kotval, 1999, Jacob & Karn, 2003). Seulement elles se basent sur une interface en 2D le plus souvent statique (site Web ou logiciel). Evaluer la jouabilité d'un jeu par l'analyse des mouvements oculaires revient à décrire et qualifier les patterns oculaires des participants lors de la visualisation d'un film où le spectateur définit les actions et les déplacements réalisés à l'écran. De ce fait, des contraintes techniques rendent plus difficile l'analyse des mouvements oculaires.

Dans une étude précédente (présentée dans le livrable intermédiaire de Novembre 2008 mis en ANNEXE 1) en étudiant deux jeux issus de la même catégorie nous avons tenté de mettre en évidence des corrélations entre les notes que les joueurs ont attribuées aux jeux vidéo sur plusieurs critères ergonomiques (critères de Bastien & Scapin) et les mouvements oculaires. Seulement les résultats attendus ne sont pas très nets : nous n'avons pu mettre en évidence de corrélations fortes entre ces deux variables. Ce manque de corrélation peut être expliqué par deux facteurs : soit elles n'ont pas lieu d'être, soit le nombre de participants sur lequel s'effectuaient les corrélations n'était pas assez important (10). En effet pour obtenir des résultats nets dans une analyse de corrélation un ensemble de 50 couples minimum est nécessaire. Or il n'est pas envisageable d'analyser les tests utilisateurs de 50 joueurs : volume de données à traiter trop important, temps de passation et temps d'analyse trop long pour promettre une analyse rapide du jeu vidéo. Pour cette raison, nous avons réfléchi à définir une autre méthode pour analyser les mouvements oculaires.

Cette nouvelle méthode est basée principalement sur le paradigme expert-novice. Utiliser ce paradigme va nous permettre de juger les jeux sur différents critères de jouabilité comme la compréhensibilité, l'apprenabilité, la localisation de l'information, l'exploration de l'interface en prenant comme groupe de référence les experts. L'idée est alors d'étudier dans quelle mesure le pattern oculaire des novices rejoint le pattern oculaire des experts. Moins de temps est nécessaire pour obtenir cette ressemblance et meilleur est le jeu. Par ailleurs l'analyse des patterns oculaires va permettre d'identifier des signatures oculaires des différentes catégories de jeux et de ce fait qualifier la différence entre le jeu testé et les jeux qui font partie de la même catégorie.

Avant de présenter la méthode nous définissons les deux types de mouvements oculaires (fixations et saccades) à partir desquels nous avons élaboré notre méthode que nous développons dans une seconde partie. Dans une troisième partie, nous présentons les expérimentations basées sur des tests utilisateurs qui ont permis de mettre à l'épreuve notre méthode. Enfin dans une dernière partie nous pointons les limites de nos résultats et les modifications qu'il faudrait apporter à notre méthode pour qu'elle puisse évaluer au plus juste la jouabilité d'un jeu vidéo.

Apprenabilité du jeu

Le principe d'apprenabilité du jeu est basé sur les caractéristiques des patterns oculaires observées chez des utilisateurs d'interface homme-machine de niveaux d'expertise différents. Les patterns oculaires ont des signatures différentes entre des utilisateurs reconnus novices et des utilisateurs reconnus experts. Cette différence de signature est due au plus grand besoin d'intégration de l'information visuelle des novices, qui ne connaissent pas le fonctionnement de l'interface et du gameplay du jeu.

Cette méthode propose d'observer l'évolution des caractéristiques des patterns de joueurs novices, en prenant pour référence les caractéristiques des patterns des joueurs experts. Les joueurs novices, en pratiquant le jeu vont peu à peu mémoriser les fonctionnements de celui-ci, ce qui devrait augmenter la proximité de leurs patterns avec ceux des joueurs experts, pour qui l'apprentissage – et l'évolution des patterns oculaires qui en résulte – sera bien moindre.

- **Evolution des fixations oculaires**

De nombreuses études (Golberg & Kotval, 1999) portant sur l'utilisation des interfaces homme-machine ont observé une diminution des durées de fixations oculaires liées à la maîtrise qu'un utilisateur avait d'une interface. Les durées des fixations oculaires sont dues au besoin en temps d'acquisition de l'information suscité par le stimulus

soumis à l'individu ; plus le stimulus va contenir d'information pour l'observateur, et plus la durée d'arrêt du regard (fixation) nécessaire au traitement de cette information va être longue.

Selon le Théorie de Détection du Signal (Weil-Barais et Coll., 1993 ; Bagot, 1996), la quantité d'information représentée par un stimulus est inversement proportionnelle au niveau d'attente qu'en a le récepteur.

Un joueur novice, qui ne connaît pas le fonctionnement d'un jeu, va nécessairement devoir traiter beaucoup plus d'informations qu'un joueur expert, ce qui augmentera substantiellement ses durées de fixation.

- **Evolution des amplitudes saccadiques**

Les amplitudes saccadiques sont elles aussi dépendantes du niveau d'expertise d'un utilisateur ; on observe une évolution des amplitudes des saccades oculaires à mesure qu'un individu prend connaissance d'une interface et mémorise sa composition spatiale (Golberg & Kotval, 1999). Les saccades oculaires sont des mouvements musculaires très rapides, limitant de ce fait les corrections possibles en cours d'exécution. C'est pourquoi la saccade oculaire est considérée comme un mouvement ayant une trajectoire balistique, estimée à l'avance pendant la fixation oculaire précédente (Rayner, 1998). Ce principe implique que le traitement de la position d'arrivée du mouvement s'appuie sur la vision para-fovéale ou la vision périphérique (pour les saccades longues), toutes deux peu discriminantes (Sanders, 1993). Ce qui suggère la place d'un traitement top-down pour permettre à l'observateur de cibler correctement un stimulus relativement éloigné, difficile à identifier perceptivement (Everatt, Bradshaw & Hibbard, 1998).

Nous estimons donc que les joueurs experts, ayant une bonne connaissance de la situation spatiale des zones du jeu, effectueront des saccades d'amplitudes plus élevées que les joueurs novices, qui ne pourront bénéficier du même traitement top-down de l'information difficile à percevoir.

Facilité de prise en main du jeu

L'objectif de cette méthode est de pouvoir situer un jeu en fonction de la rapidité de prise en main qu'il permet. Cette méthode est basée sur une comparaison de l'évolution des caractéristiques des patterns de joueurs reconnus novices avec les patterns de joueurs reconnus experts.

En se basant sur les caractéristiques des patterns oculaires détaillées précédemment (durées de fixation et amplitudes saccadiques), on estime qu'un jeu dont la prise en main est rapide doit proposer :

- une très rapide évolution des patterns novices, rejoignant les patterns experts
- une relative stagnation des patterns après un temps assez court

En effet, un novice confronté à un jeu facile à prendre en main va très rapidement développer des connaissances sémantiques et procédurales portant sur les commandes du jeu. Ainsi, plus la prise en main est aisée et plus cette première évolution doit être rapide.

La relative stagnation des patterns après cette première phase suppose simplement que les joueurs novices comme les joueurs experts n'apprennent plus - ou presque plus - du jeu après cette phase. Cette définition doit être considérée comme le pôle supérieur de l'échelle de qualité de la prise en main du jeu, en pratique des comparaisons avec d'autres jeux du même genre sont nécessaires.

Typicalité du jeu (et transfert de connaissances)

La typicalité du jeu est une méthode basée sur la comparaison du jeu évalué avec d'autres jeux, issus d'un échantillon de jeux représentatifs du même genre. Cette évaluation permet de situer le jeu sur une échelle de typicalité qui va d'«atypique» (ou innovant) à « typique ». Le principe est de comparer les données oculométriques de joueurs experts (nécessaire pour obtenir les patterns oculaires correspondant au gameplay du jeu) à partir de passations, qui peuvent être courtes si le moment de test correspond à l'action principale du jeu. Les données retenues sont les indices de densité spatiale des fixations oculaires, évalués à partir d'un outil de cartographie des fixations. Grâce à cet outil, nous obtenons une description des fixations détaillée sous forme de cartes topographiques des fréquences ou des durées de fixations réparties sur l'écran du jeu (cartes de chaleur).

En comparant la carte de chaleur du jeu évalué à un ensemble de cartes de chaleur de jeux évalués précédemment, il nous a été alors aisé de pouvoir situer le jeu entre un jeu reconnu comme fortement typique du genre et un jeu fortement atypique du genre.

Des comparaisons inter-genre sont également envisagées pour permettre d'étudier des jeux hybrides, de plus en plus usités par les développeurs.

Des résultats quantitatifs peuvent appuyer ces données descriptives, le logiciel Lutin MOC permet de calculer la répartition des fixations sur une matrice représentant des zones d'écran. Le pas de la matrice de calcul peut être décidé en fonction de la finesse d'analyse souhaitée. Le logiciel calcule les indices statistiques quantitatifs nécessaires à la description chiffrée permettant de placer le jeu sur l'échelle de typicalité.

Outre l'intérêt intrinsèque de la méthode, il nous paraît important de préciser que cette dernière évaluation représente un bon indicateur de la capacité d'un jeu à pouvoir permettre un **transfert de connaissances** ; en effet, le jeu peut être considéré comme une activité de résolution de problème (Simon, 1973 - 1995) permettant une résolution partielle ou totale par transfert analogique. On peut considérer qu'un joueur aillant développé une expertise dans un jeu proche du nouveau jeu acheté devrait pouvoir effectuer des analogies entre les deux jeux, ces inférences devraient lui permettre de jouer plus facilement et de prendre en main le jeu plus rapidement.

Les études portant sur le raisonnement analogique (Holyoak & Gentner, 1998, Bonnardel, 2000) précisent que la distance séparant le domaine source (jeu pratiqué) du domaine cible (jeu nouveau) détermine la difficulté du transfert de connaissances, un domaine éloigné nécessitant une transformation plus en profondeur des propriétés adaptés d'un problème à l'autre.

Ces résultats suggèrent qu'un jeu très typique devrait générer des transferts analogiques plus importants et plus faciles pour les joueurs experts, qui devraient se sentir à l'aise bien plus rapidement avec le jeu typique. Ainsi cette méthode pourrait fournir aux développeurs une indication préalable sur l'accueil réservé par les joueurs à leur jeu selon leur niveau d'expertise.

Enfin, certains genres de jeux comme les jeux RTS ou les jeux FPS nécessitent un certain

transfert de connaissance grâce à la complexité de leur gameplay ou de leur interface de commande, d'autres genres comme les jeux Casual doivent au contraire pouvoir être joués aussi bien par des novices que par des experts, dès la phase d'accostage.

Comprenabilité de l'interface

La comprenabilité de l'interface est une méthode d'évaluation basée sur l'analyse de deux caractéristiques des mouvements oculaires : les durées de fixation, qui représentent le besoin d'acquisition de l'information visuelle, telles que décrites précédemment, et les saccades régressives qui représentent les retours immédiats sur des fixations précédentes, pour traiter à nouveau l'information (Rayner, 1998 ; Baccino, 2004).

En limitant l'analyse oculométrique aux zones de l'interface du jeu (intérêt particulier pour les jeux RTS), les données oculaires enregistrées dans les premières phases de la partie nous donnerons un élément de comparaison fiable concernant la facilité de compréhension de l'interface de commande du jeu.

Une interface intuitive doit générer, d'une part des durées de fixation oculaire plus courtes, puisque le joueur aura moins d'information à acquérir, et ceci plus rapidement qu'une interface difficile à comprendre. D'autre part les saccades régressives devraient être présentes en moins grand nombre et décroître plus rapidement. Les fixations régressives sont observées notamment dans les activités de lecture lorsqu'un texte est difficile à intégrer, ou lorsque la cohérence locale du texte est défaillante (Albrecht & O'Brien, 1993). Le lecteur va dans un premier temps traiter le mot fixé et continuer sa lecture par une pour plusieurs fixations progressives, pour revenir sur le mot fixé précédemment afin de le traiter de nouveau, dans le but de compléter l'intégration du mot dans son modèle mental (Johnson-Laird, 1995) ou chercher à résoudre l'incohérence en effectuant des inférences. Les fixations régressives sont observées également dans les activités portant sur l'utilisation d'interfaces homme-machine telles que la découverte d'une interface logicielle ou la recherche d'informations sur le Web.

En somme, plus le joueur peut s'appuyer sur les connaissances issues de son expérience de joueur ou pragmatiques (intuitivité de l'interface) et plus il pourra identifier rapidement les codes de l'interface (durées de fixation), sans effectuer de traitements régressifs (fixations régressives).

Localisation de l'information

Les données oculométriques peuvent nous permettre de déterminer la rapidité de localisation de l'information sur l'écran de jeu. En se basant sur une tâche de recherche d'information scénarisée (on demande au participant de localiser rapidement un élément), nous devrions observer une évolution des patterns oculaires représentatifs de la facilité de localisation offerte par la scène du jeu. Nous devrions observer tout d'abord des saccades oculaires dites ambiantes, qui sont caractérisées par une forte amplitude et une forte dispersion, pour très rapidement évoluer vers de saccades focales caractérisées par des amplitudes faibles et une dispersion très faible.

Les saccades ambiantes sont utilisées pour explorer rapidement l'information visuelle, ce traitement de surface permet de localiser les grands ensembles du stimulus afin de sélectionner le plus pertinent dans la situation demandée. Les saccades ambiantes précèdent les saccades focales, qui, une fois l'information localisée, permettent de traiter l'information en profondeur.

Selon ce principe de fonctionnement du système perceptif visuel, plus tôt les saccades ambiantes laissent place aux saccades focales, plus la localisation de l'information au sein d'un stimulus est aisée pour le participant. Ainsi un élément de jeu dont la recherche donne lieu à une période de saccades ambiantes plus courte qu'un autre élément équivalent, pourra être qualifié de plus facile à localiser. Ce principe peut être appliqué à l'évaluation d'éléments de jeux différents, toutes choses étant égales par ailleurs.

Exploration de l'interface

Une dernière méthode proposée vise à évaluer la facilité d'exploration de l'interface d'un jeu en utilisant la répartition des fixations dans l'espace au cours du jeu. En se basant sur les propriétés des saccades ambiantes énoncées dans la partie précédente, nous faisons l'hypothèse qu'un jeu dont l'interface ou l'écran de visualisation est facile à explorer, devrait comporter des fixations oculaires plus dispersées (avec des amplitudes saccadiques plus grandes) et moins nombreuses. En effet, les saccades ambiantes ont pour but d'acquiescer superficiellement l'information visuelle, celles-ci étant plus amples et moins nombreuses suggèrerait que le traitement de l'écran de jeu a pu s'effectuer plus facilement puisque le joueur a eu besoin de moins de prise d'information. L'exploration de l'interface du jeu doit être évaluée dès les premières secondes d'une nouvelle séquence de jeu ou dès le début de la première partie du joueur.

Récapitulatif de la méthode oculométrique (graph)

- Apprenabilité

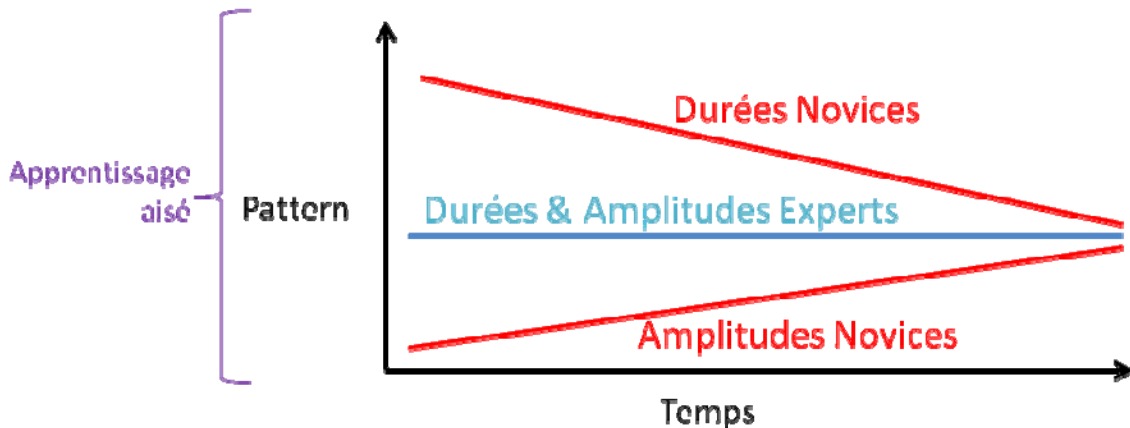


Figure 1. Schéma de principe de l'évolution des patterns oculaires pour un jeu proposant un apprentissage aisé

- Prise en main du jeu

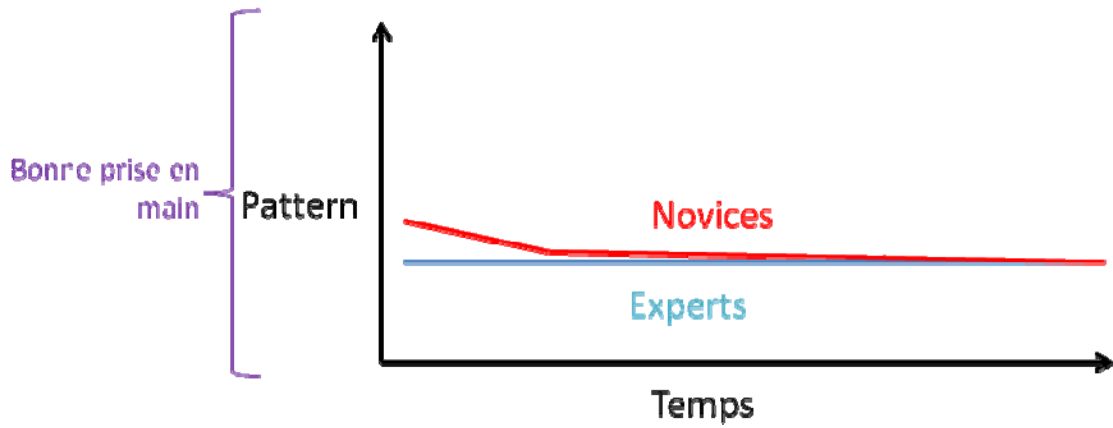


Figure 2. Schéma de principe de l'évolution des patterns oculaires pour un jeu offrant une bonne prise en main

- Typicalité du jeu (et transfert de connaissances)

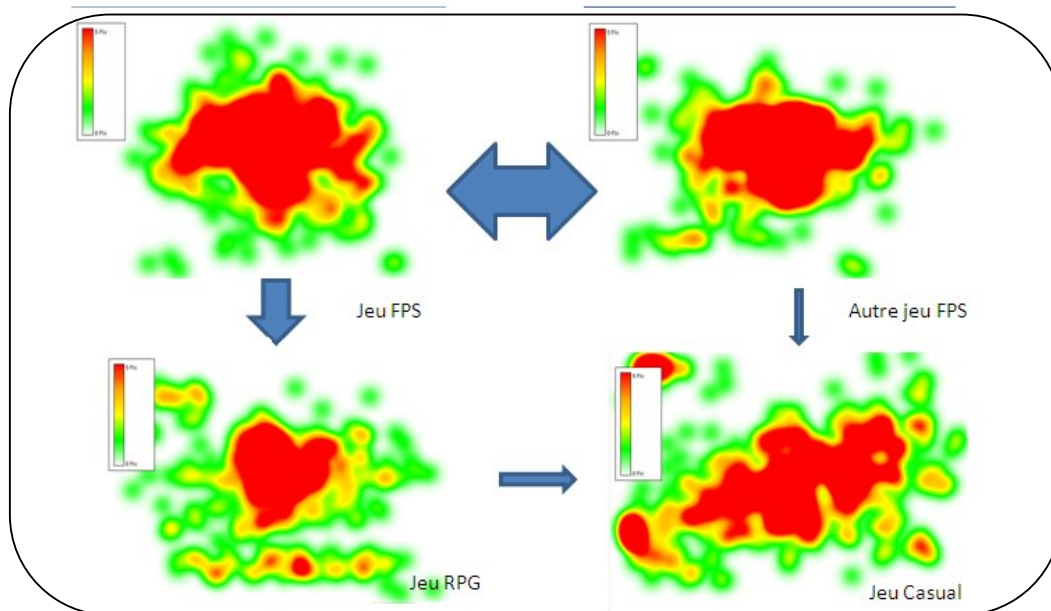


Figure 3. Principe de comparaison des cartes topographiques des fixations (cartes de chaleur) entre quatre jeux : deux d'un même genre (en haut) et deux de genres différents (en bas).

- Comprehensibility of the interface

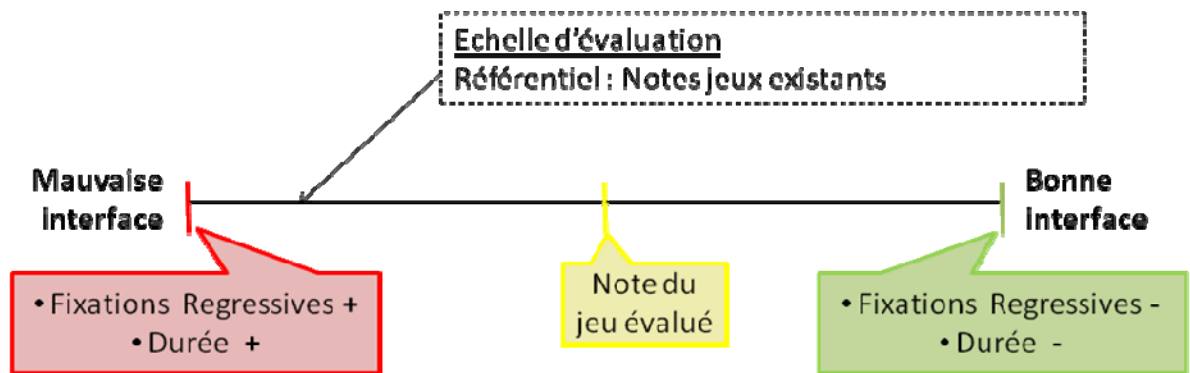


Figure 4. Echelle de d'évaluation de la « compréhensibilité » de l'interface d'un jeu vidéo.

- Localization of information

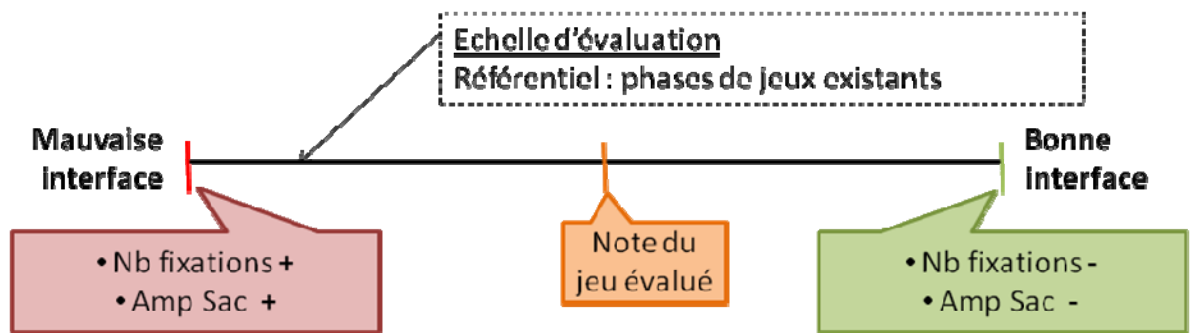


Figure 5. Echelle d'évaluation de la facilité de localisation de l'information dans le jeu testé.

- Exploration of the interface

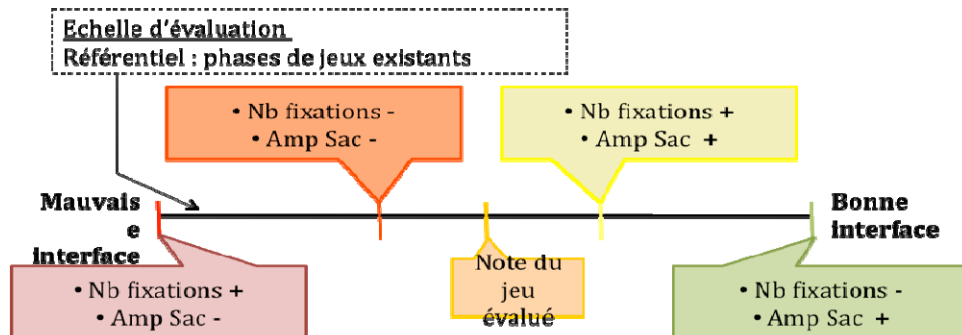


Figure 6. Echelle d'évaluation de la facilité d'exploration de l'interface du jeu testé.

QUESTIONNAIRES D'UTILISABILITE

Parallèlement à l'analyse des mouvements oculaires, nous avons mis au point un questionnaire d'utilisabilité du jeu vidéo. Le but de ce questionnaire est de recueillir – le plus exhaustivement possible - le ressenti subjectif des joueurs concernant les qualités du jeu évalué, en termes d'ergonomie et de gameplay.

Cette évaluation devait nous permettre de qualifier plus précisément les résultats oculométriques obtenus lors des tests, en nous fournissant un point de vue différent du jeu : celui du joueur venant de réaliser le test.

Le questionnaire d'utilisabilité fut élaboré à partir de deux sources : des critères d'ergonomie des interfaces de Bastien et Scapin pour l'évaluation ergonomique et des heuristiques issues de la revue de littérature de Federoff pour l'évaluation du gameplay. Les critères de Bastien et Scapin sont organisés en huit grandes catégories :

- ✓ **Guidage** : critères destinés à guider l'utilisateur dans son activité
- ✓ **Charge de travail** : critères destinés à réduire la charge perceptive et mnésique
- ✓ **Contrôle explicite** : critères destinés à donner le contrôle à l'utilisateur
- ✓ **Adaptabilité** : critères destinés à adapter le système à l'utilisateur
- ✓ **Gestion des erreurs** : critères destinés à prévenir les erreurs
- ✓ **Homogénéité / Cohérence** : critères destinés à l'interface cohérente dans son ensemble, pour que l'utilisateur puisse trouver des invariants
- ✓ **Signifiante des codes et dénominations** : critères destinés à ce que tout symbole, référent ou abréviation doit être parfaitement compréhensible par l'utilisateur
- ✓ **Compatibilité** : critères visant à ce que les caractéristiques de l'interface soient compatibles avec les caractéristiques de l'utilisateur

La revue de littérature de Federoff présente seize items de gameplay que l'on peut regrouper en six grandes catégories :

- ✓ **Esthétique** : items en rapport avec le plaisir visuel qu'apporte le jeu
- ✓ **Apprentissage** : items en rapport avec l'acquisition de compétences que permet le jeu
- ✓ **Rythme du jeu** : items en rapport avec le rythme narratif et le rythme de l'action du jeu
- ✓ **Gratifications** : items en rapport avec les gratifications fournies au joueur pendant le jeu
- ✓ **Difficulté/Challenge** : implique la gestion de la difficulté du jeu, qui peut être positive (challenge) ou négative (impossibilité d'avancer dans le jeu)
- ✓ **Liberté d'action** : regroupe les items décrivant la liberté d'action que permet le jeu

Nous avons extrait de ces deux sources un ensemble de quarante cinq items, que nous avons retravaillés en quarante et une questions contextualisées afin que le joueur puisse répondre à la suite du test sans être dérouté ou avoir besoin d'une expertise dans les domaines évalués.

Les joueurs répondaient à des affirmations sur une échelle de Lickert en sept points, allant de un : « Je suis tout à fait d'accord » à sept : « Je ne suis pas du tout d'accord ». L'ordre des questions était contre-balancé et la valence positive ou négative de l'affirmation évaluant le jeu était présentée aléatoirement. Ainsi, pour l'évaluation de l'ergonomie du jeu, les joueurs ont eu à donner leur avis par une note à l'affirmation du type « *Le graphisme de l'interface vous a aidé à comprendre les différentes fonctions* » pour évaluer le critère « guidage » ou à une affirmation du type « *le déroulement du jeu vous a étonné par ses rebondissements* » pour évaluer le critère « Rythme » du jeu.

Les questionnaires d'utilisabilité étaient présentés aux participants après chaque test de jeu, l'expérimentateur s'assurait que le jeu était encore accessible au participant afin de l'aider à répondre au questionnaire. Les questionnaires étaient présentés sous forme de formulaire à remplir sur un écran avec une souris, ainsi le recueil des réponses était automatisé.

TESTS JOUEURS REALISES

Les tests se sont déroulés en trois grandes parties : Une session expérimentale pour comparer deux jeux RTS, une autre session pour comparer deux jeux FPS et une dernière session pour comparer plusieurs jeux de quatre catégories de jeux. Les deux premières sessions ont pour but d'évaluer notre méthode en comparant pour chaque catégorie un jeu reconnu bon et un jeu reconnu mauvais (sur les forums et par l'évaluation ergonomique d'un des expérimentateurs). La dernière session a pour but de comparer les densités spatiales des jeux selon leur catégorie.

JEUX RTS

Méthode

Participants. Nous avons recruté un ensemble de 26 participants pour réaliser la série de passations oculométriques. Le recrutement a été effectué en majorité parmi les visiteurs de la Cité des Sciences et de l'Industrie de Paris, lieu où se situe le laboratoire. Les participants étaient informés qu'ils allaient devoir jouer à des jeux vidéo, l'expérimentateur leur expliquait qu'il fallait qu'ils aient déjà joué avant pour pouvoir participer. Nos participants étaient âgés de 21 à 35 ans (moyenne 24,6 ans), en majorité des hommes (5 femmes) et de profession diverses. L'expertise des joueurs était évaluée une fois le joueur recruté, par un questionnaire sondant leurs habitudes de joueurs (cf. [Annexe](#)). Nous avons veillé au cours des passations à compléter nos groupes d'expertises. Le groupe d'experts fut constitué via un forum de « hard gamers ». Chaque participant a été rétribué par une place de cinéma à la Géode (cinéma sphérique de la cité des Sciences et de l'Industrie).

Matériel : les Jeux. Les deux jeux testés étaient « Perimeter » et « War Hammer 40000 : Dark Crusade », tous deux sont des jeux de Stratégie en Temps Réelle ou « RTS ». Dans ce genre de jeu, le but est d'acquérir des ressources pour construire une armée suffisamment puissante pour conquérir de nouveau territoire. Le joueur prend le rôle d'un général d'armée qui doit gérer sa base et commander ses troupes en vue de dessus, à l'aide d'une interface de commande assez présente, où sont situées les fonctions nécessaires au commandement.

Nous avons choisi ces deux jeux en nous basant tout d'abord sur les notes obtenues dans la presse, en visant tout particulièrement les notes concernant l'ergonomie et le gameplay.

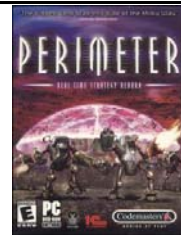


Warhammer 40000 / Dark Crusade

Note Gamekult : **8/10**

Critiques :

- Peu scénarisé
 - Le moteur (graphique) accuse son âge
- L'IA et le pathfinding perfectibles



Perimeter

Note Gamekult : **5/10**

Critiques :

- Gestion des troupes un peu loupée
 - Combats brouillons
 - **Ergonomie médiocre**
- Bien trop gourmand

Puis nous avons testé ces jeux en effectuant une expertise ergonomique sur leur interface de commande et évalué des mécanismes de gameplay qui régissent la prise en main de chaque jeu (cf. [Annexe](#)). Nous avons obtenu 21 points positifs et 3 négatifs pour le jeu Dark Crusade contre 1 point positif et 23 points négatifs pour Perimeter.

Dispositif oculométrique. Les mouvements oculaires des participants étaient enregistrés avec le dispositif Tobii 1750 qui échantillonne la position des deux yeux toutes les 20 ms et fait une moyenne de ces deux positions pour déterminer les coordonnées cartésiennes (x,y) du regard.

Procédure

Passation. Le participant s'asseyait devant un bureau sur lequel étaient placés un clavier, une souris et l'oculomètre Tobii (cf. [Annexe](#)). Après lui avoir expliqué le fonctionnement de l'oculomètre et le type de données récupérées à l'issue de son test, il était invité à remplir un formulaire de participation librement consentie.

Lorsque le participant était prêt, la consigne lui était donnée à l'écrit (cf. [Annexe](#)) et à l'oral. L'expérimentateur lui demandait s'il avait bien compris, lui annonçait qu'il allait être conduit à lui poser des questions sur le jeu au cours de la partie, puis lançait la calibration de l'oculomètre pour enfin lancer la passation.

Pendant son test, le joueur est invité à avancer suffisamment dans le cours de la partie pour pouvoir construire un certain type de véhicule, nécessitant la construction de plusieurs bâtiments et le développement de plusieurs technologies.

Le choix de cette tâche fut motivé par la possibilité offerte d'observer l'évolution des joueurs de façon écologique, au cours des premières minutes d'utilisation des deux jeux. Cette tâche fut calibrée pour proposer une stricte équivalence d'action et de difficulté pour les deux jeux.

Trois questions (cf. [Annexe](#)) étaient posées pendant le test, nécessitant la localisation de trois éléments importants de l'interface de jeu. Ces trois questions étaient les mêmes pour les deux jeux, leur ordre de présentation était contrebalancé.

A l'issue du premier test, le participant évaluait le jeu testé en remplissant un questionnaire d'appréciation subjective (cf. [Annexe](#)), il pouvait à tout moment consulter le jeu encore présent à l'écran.

Les participants testaient chacun des deux jeux pendant quinze minutes. L'ordre de présentation des jeux était contrebalancé.

Après le test du deuxième jeu, le participant était remercié et informé de la problématique de la recherche.

Identification des fixations oculaires. Une fixation oculaire a été définie par un minimum de 5 points successifs dans un rayon de 20 pixels

Résultats

Les résultats vont être présentés en deux parties : une première partie sur l'analyse du questionnaire d'utilisabilité et une seconde partie sur l'analyse des mouvements oculaires effectués lors du jeu.

- Questionnaires d'utilisabilité

Evaluation de l'ergonomie

La figure 7 laisse apparaître une légère supériorité de résultats pour le jeu Dark Crusade, mais l'amplitude de cette variation reste peu marquée et évolue conjointement pour les deux jeux. L'égalité observée pour le critère « Gestion des erreurs » est à pondérer par le fait qu'aucune gestion des erreurs n'est apparente dans ces deux jeux.

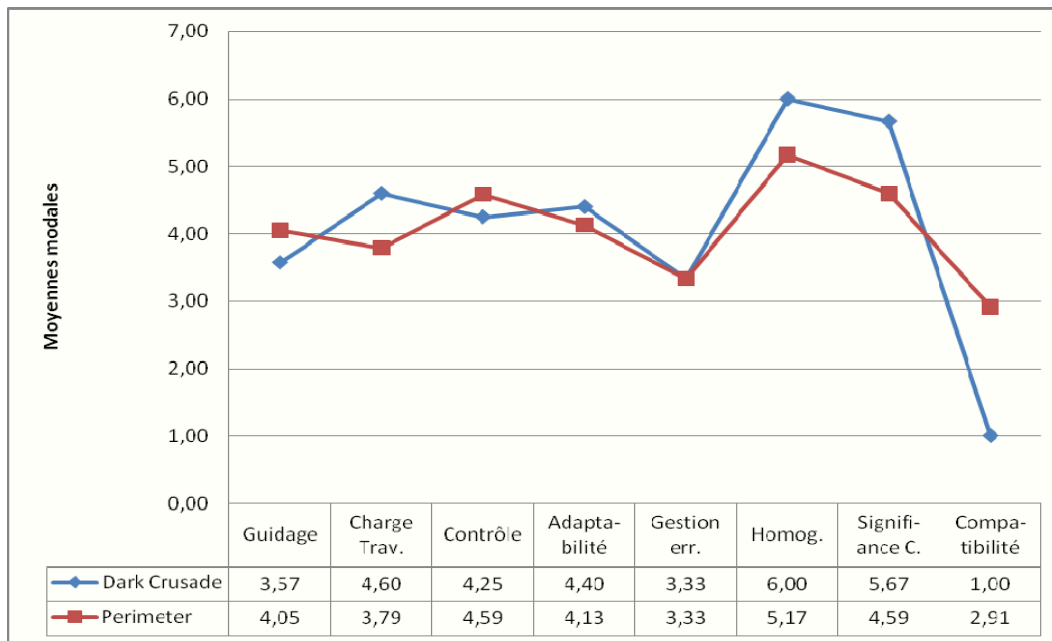


Figure 7. Notes modales moyennes obtenues par les deux jeux RTS pour les critères ergonomiques

Evaluation du gameplay

La figure 8 montre comme pour les critères ergonomiques une évaluation légèrement supérieure pour le jeu Dark Crusade, mais là encore l'amplitude des variations entre les deux jeux n'est que peu marquée dans l'ensemble.

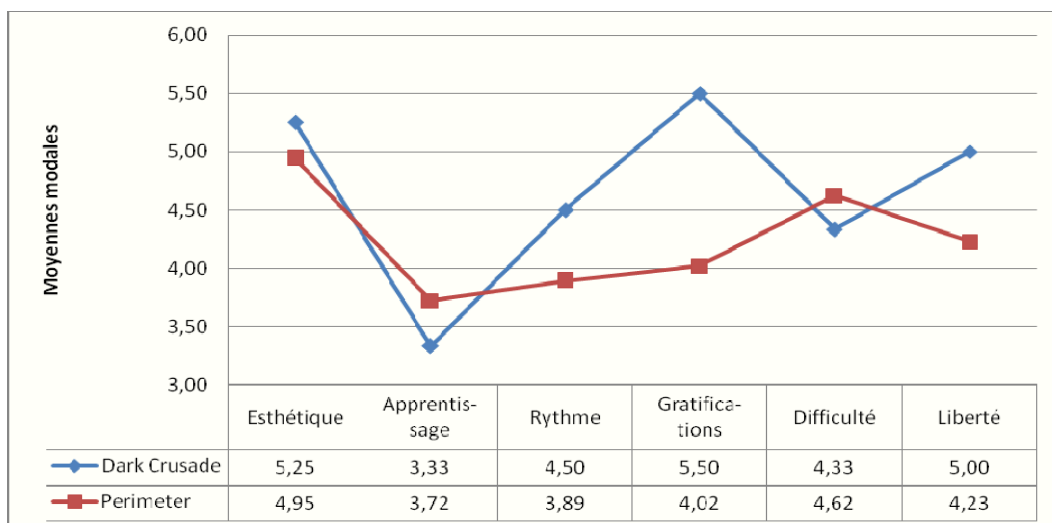


Figure 8. Notes modales moyennes obtenues par les deux jeux RTS pour les critères de gameplay

Mouvements oculaires

Apprenabilité : évolution des durées moyennes de fixation et des amplitudes saccadiques.

La Figure 9 nous indique que les patterns concernant l'évolution des durées moyennes de fixations ne varient pas trop pour Dark Crusade et ceci quelle que soit l'expertise. En revanche ce n'est pas le cas pour Perimeter.

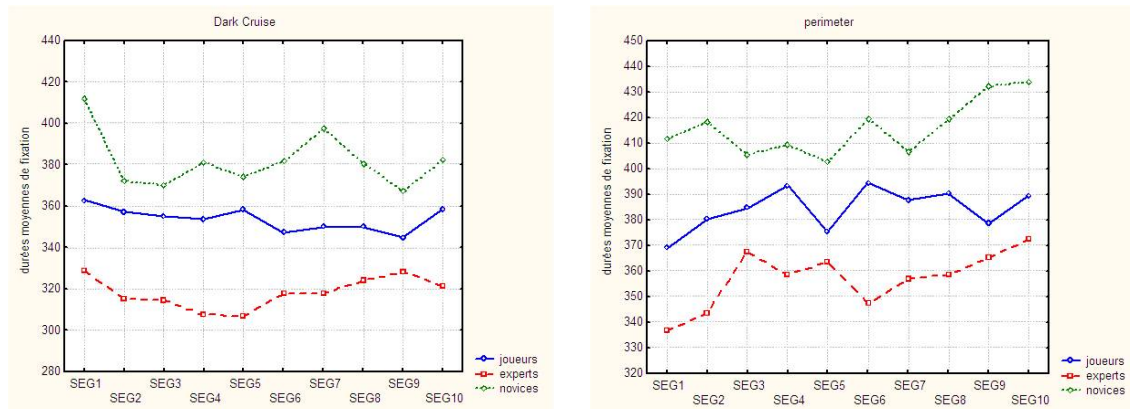


Figure 9. Evolution des durées moyennes de fixation selon l'expertise du joueur pour chaque jeu

Pour les amplitudes saccadiques quels que soient le jeu et le niveau d'expertise nous observons des variations importantes selon le moment du jeu (Figure 10).

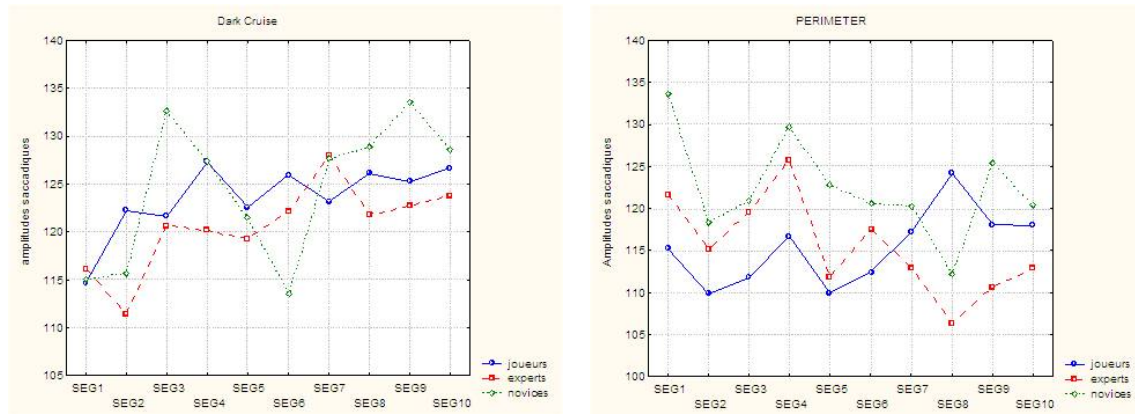


Figure 10. Evolution des amplitudes saccadiques selon l'expertise du joueur pour chaque jeu

Comprenabilité : saccades régressives et durée moyenne de fixation

La Figure 11 montre que le pourcentage de saccades régressives est plus important pour Perimeter que pour Dark Crusade, et ceci quelle que soit l'expertise des joueurs. Par ailleurs Perimeter présentait des durées de fixation plus longues que Dark Crusade (Figure 12). Ces résultats suggèrent que les éléments dans Dark Crusade sont plus compréhensifs que les éléments dans Perimeter.

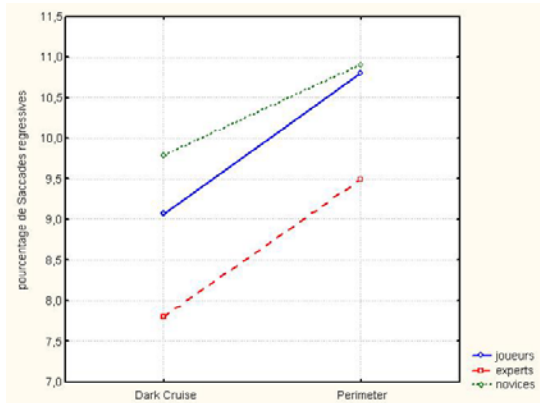


Figure 11. Pourcentage de saccades régressives selon le jeu et l'expertise du joueur

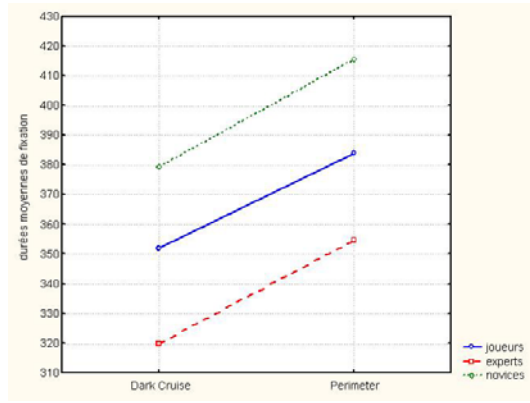


Figure 12. Durée moyenne de fixation selon le jeu et l'expertise du joueur

Localisation de l'information.

Pour répondre aux différentes questions posées les joueurs font moins de fixations oculaires (Figure 13) et des saccades plus amples (Figure 14) sur Dark Crusade que sur Perimeter. Ces résultats suggèrent que Dark Crusade permettrait de mieux localiser l'information sur l'interface que Perimeter.

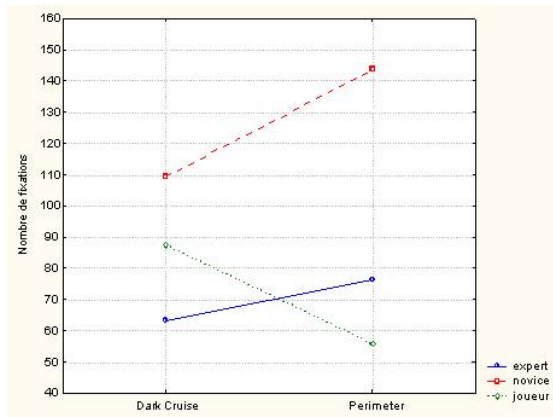


Figure 13. Nombre de fixations nécessaires pour répondre aux questions selon le jeu et l'expertise du joueur.

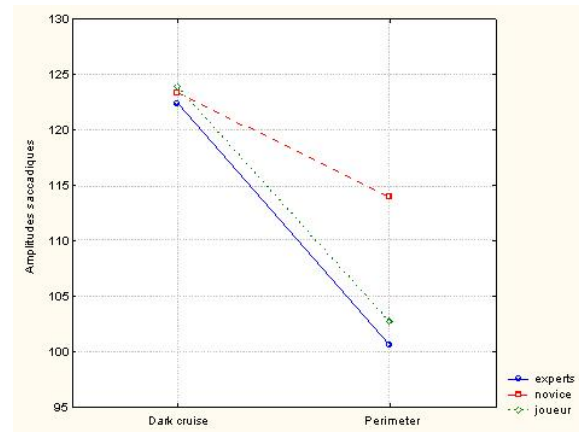


Figure 14. Amplitudes saccadiques observées pour répondre aux questions selon le jeu et l'expertise du joueur.

Exploration de l'interface

La Figure 15 et la Figure 16 nous suggèrent que Perimeter permet une meilleure exploration que Dark Crusade : bien que le nombre de fixations soit équivalent entre les jeux quelle que soit l'expertise des joueurs, Perimeter a engendré des saccades oculaires plus amples que Dark Crusade.

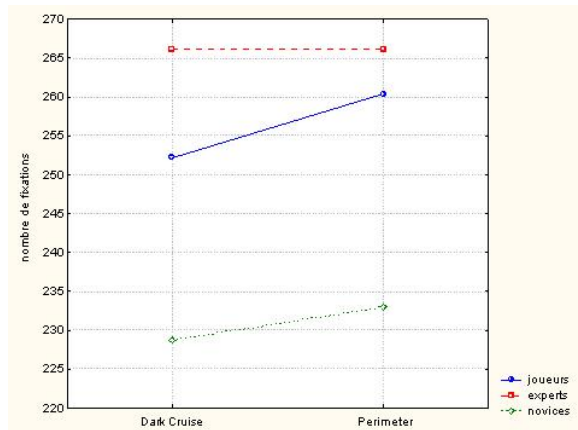


Figure 15. Moyennes du nombre de fixations observées dans les deux premières minutes pour chaque jeu selon l'expertise du joueur

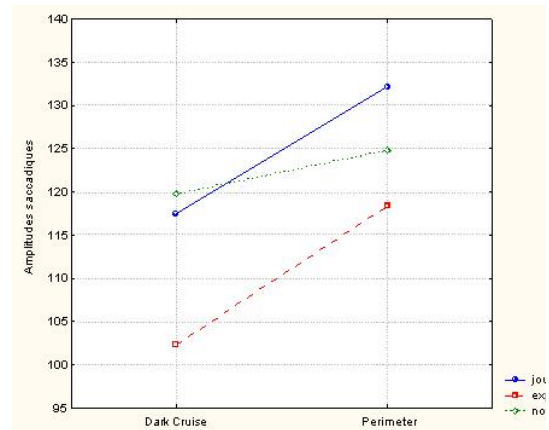


Figure 16. Moyennes amplitudes saccadiques observées dans les deux premières minutes pour chaque jeu selon l'expertise du joueur

En conclusion sur les mouvements oculaires.

Les différentes analyses nous indiquent que Dark Crusade permet une meilleure apprenabilité, une meilleure compréhension et une meilleure localisation de l'interface que Perimeter. En revanche Perimeter permettrait une meilleure exploration de l'interface que Dark Crusade. Ce résultat étonnant est certainement dû au fait Dark Crusade présente une interface avec un niveau d'arborescence du menu d'interface supplémentaire, là où Perimeter possède l'intégralité de ses commandes au premier niveau. Notre méthode d'analyse de la facilité d'exploration de l'interface étant basée en grande partie sur des saccades amples, la configuration des jeux dans ce cas précis pourrait ne pas se prêter à cette analyse.

JEUX FPS

Méthode

Participants. Nous avons recruté 23 participants pour réaliser la série de passations oculométriques. Le recrutement a été effectué à l'aide de la base de données « Faites de la science » construite par le RISC. Les participants étaient informés par courriel qu'ils allaient devoir jouer à des jeux vidéo, l'expérimentateur leur demandait au préalable s'ils avaient une bonne pratique des jeux FPS en leur expliquant que l'on recherchait plusieurs profils. Seuls les joueurs se qualifiant de bon joueur de FPS étaient retenus. Ils étaient âgés entre 18 et 41 ans (moyenne 24, 4 ans), composés d'une majorité d'hommes (3 femmes) et de professions diverses. L'expertise des joueurs était également évaluée une fois le joueur recruté, par un questionnaire détaillé sondant leurs habitudes de joueurs (cf. [Annexe](#)). Nous avons veillé au cours des passations à compléter nos groupes d'expertises. Les participants ont été dédommés par un bon d'achat d'une valeur de 10 euros.

Matériel : les Jeux. Les deux jeux testés étaient « Halo3 » et « History Channel : Battle for the pacific ». Ces deux jeux sont du genre « Tire à la première personne » (FPS). Dans ce type de jeux, le joueur est amené à incarner un soldat en vue subjective, c'est-à-dire que le jeu est conçu de telle sorte que le joueur puisse se mettre dans la peau du personnage. L'écran de jeu

présente alors la scène vue des yeux du personnage (un soldat en l'occurrence), les éléments d'interface sont nécessaires, mais doivent être savamment dosés pour se faire oublier et permettre au joueur en bonne immersion.

Nous avons opté pour ces deux jeux en nous basant sur les notes obtenues dans la presse tout en ayant le souci de garder une bonne équivalence de genre entre les deux. De plus Halo 3 est reconnu dans le monde du jeu vidéo pour être un jeu finement calibré, pour lequel de nombreux tests utilisateur ont été faits.



Halo 3

Note Gamekult : **7/10**

Critiques :

- Beaucoup trop court
- Peu d'innovations, pas de

surprises

- Level design souvent paresseux
- Graphiquement décevant
- Le boss de fin, nullissime
- Pas de bots en multi
- I.A. toujours incapable de conduire



The History Channel : Battle for the Pacific

Note Gamekult : **2/10**

Critiques :

- SoF : Payback avec une skin Seconde Guerre mondiale
- Zéro I.A.
- Fil conducteur incompréhensible
- C'est pas bien beau et ça rame
- Multi déserté et insipide
- Environnements très répétitifs
- Deux heures de jeu à tout casser
- Intégralement en anglais

Les deux jeux choisis nous ont également permis de tester notre dispositif de mesure oculométrique sur console de jeu, puisque ces deux jeux s'exécutaient sur la Xbox 360 et se jouaient avec une manette. Pour pallier au fait que Battle for Pacific était en anglais, nous avons choisi une séquence très pauvre en dialogue et nous avons traduit les quelques paroles prononcées pendant le jeu.

Procédure

Le test se déroulait exactement de la même manière que pour le test précédent sur les jeux RTS (réalisation d'un scénario, [Annexe 10](#), répondre à trois questions pendant le jeu, [Annexe](#) , et répondre à un questionnaire d'appréciation subjective pour chaque jeu, [Annexe](#)), à l'exception du périphérique utilisé pour agir dans le jeu : une manette de la Xbox 360 au lieu de la souris et du clavier. Les participants testaient chacun des deux jeux pendant dix minutes. Après le test du deuxième jeu, le participant était remercié et informé de la problématique de la recherche.

La définition d'une fixation oculaire est exactement la même que celle qui est utilisée dans le test précédent.

Résultats

Tout comme dans le test précédant les résultats vont être présentés en deux parties : analyse des questionnaires d'évaluation et analyse des mouvements oculaires.

- Questionnaires d'utilisabilité

Evaluation de l'ergonomie

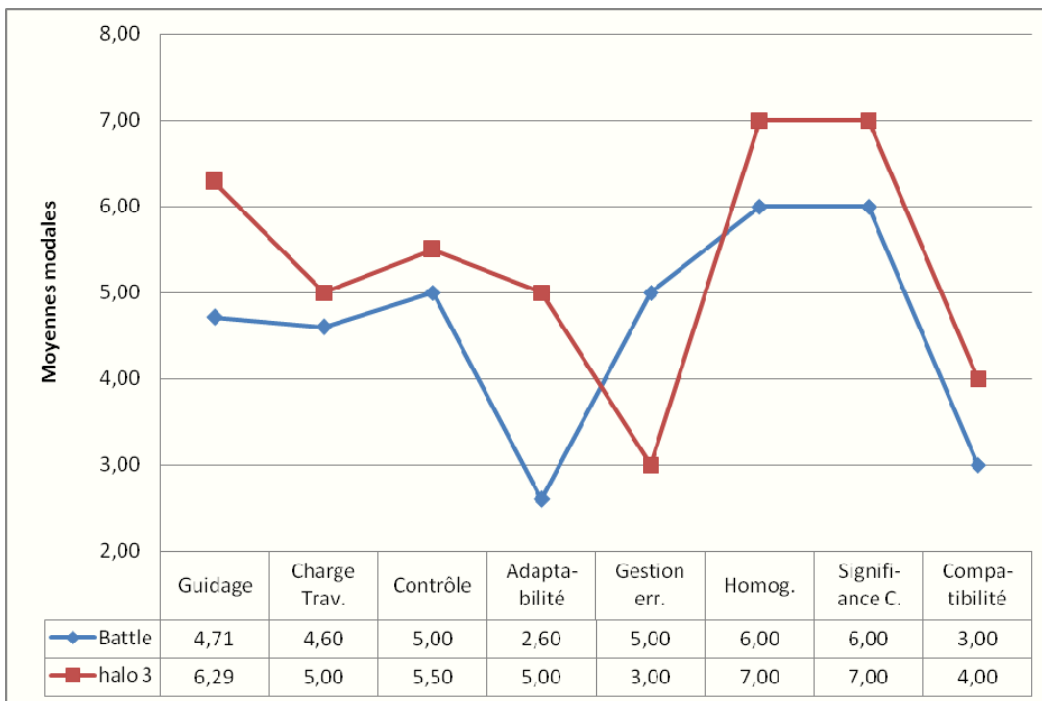


Figure 17. Notes modales moyennes obtenues par les deux jeux FPS pour les critères ergonomiques

La Figure 17 montre une meilleure évaluation de Halo 3 pour tous les critères ergonomiques, sauf pour le critère « gestion des erreurs ». Il faut noter que ce dernier critère est difficile à évaluer dans

la mesure où les deux jeux ne présentent pas de gestion d'erreurs visible du point de vue de l'utilisateur.

Evaluation du gameplay

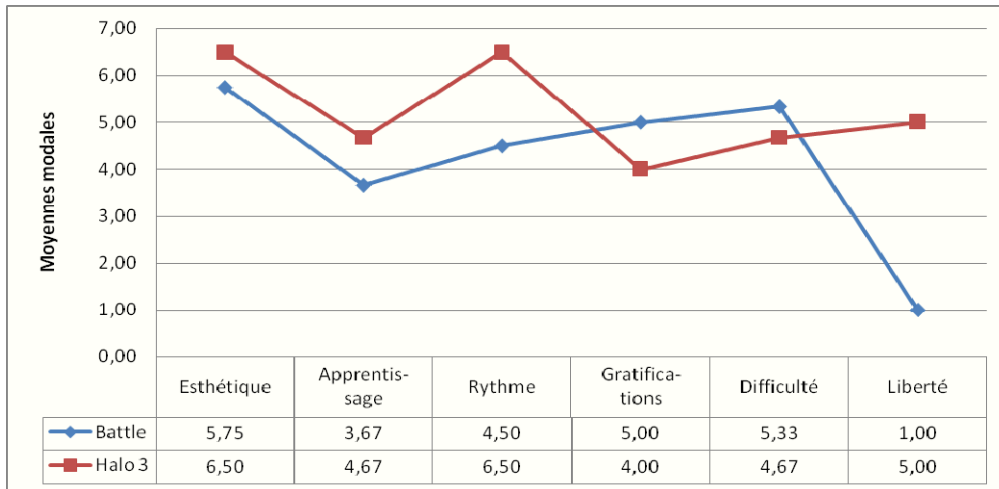


Figure 18. Notes modales moyennes obtenues par les deux jeux FTS pour les critères de gameplay

La Figure 18 montre globalement une évaluation supérieure du jeu halo 3 pour les critères de gameplay. Seules les catégories « Gratifications » et « Difficulté » ont été évaluées en faveur de Battle, avec une amplitude de variation faible.

- Mouvements oculaires

Apprenabilité : évolution des durées moyennes de fixation et amplitudes saccadiques

L'analyse de l'évolution des durées moyennes de fixations nous indiquent que pour Halo 3 le pattern des novices tend vers le pattern des experts ce qui n'est pas le cas de Battle (Figure 19).

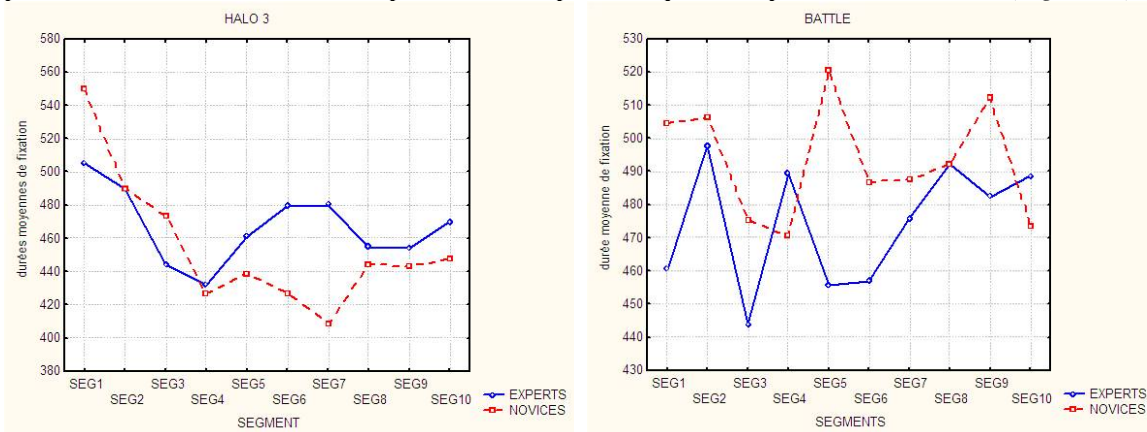


Figure 19. Evolution des durées moyennes de fixations selon l'expertise des joueurs pour chaque jeu

En revanche, il n'y a pas de telles tendances pour les amplitudes saccadiques (Figure 20).

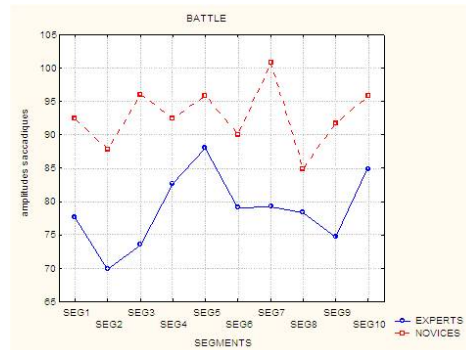
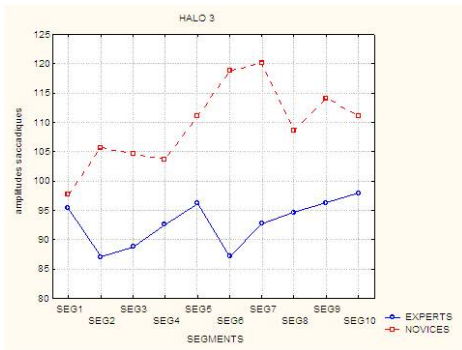


Figure 20. Evolution des amplitudes saccadiques selon l'expertise des joueurs pour chaque jeu

Comprenabilité : saccades régressives et fixations oculaires.

Comme le montre la Figure 21, le jeu Halo 3 présentait moins de saccades régressives que le jeu Battle. Par ailleurs la Figure 22 montre une augmentation franche des durées de fixations pour les novices dans le cas de Battle.

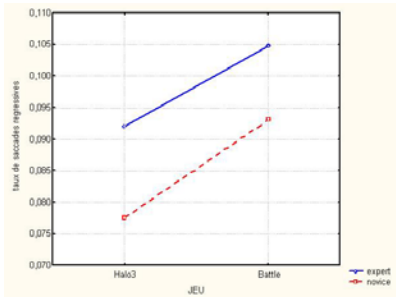


Figure 21. Taux de saccades régressives pour chaque jeu FPS selon l'expertise du joueur

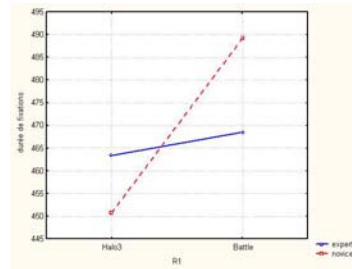


Figure 22. Durées moyennes de fixation pour chaque jeu FPS selon l'expertise du joueur

Localisation de l'information

Les joueurs ont fait moins de fixations oculaires pour trouver la cible dans le jeu Battle que dans Halo3 (Figure 23). Par ailleurs ils ont fait des saccades plus faibles pour Battle que pour Halo 3 (Figure 24).

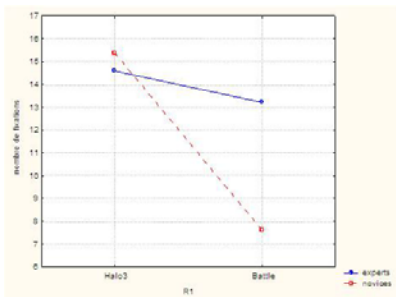


Figure 23. Nombre de fixations nécessaires pour trouver la cible pour chaque jeu FPS selon l'expertise du joueur

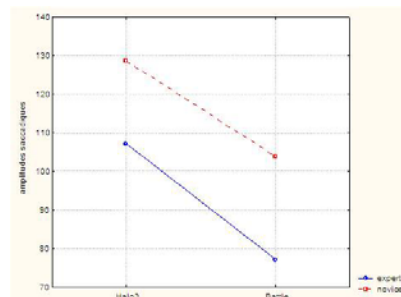


Figure 24. Moyenne des amplitudes saccadiques observées pour trouver la cible pour chaque jeu FPS selon l'expertise du joueur

Exploration de l'interface

La Figure 25 et la Figure 26 nous suggèrent que Halo 3 permet une meilleure exploration de

l'interface que Battle : Les joueurs ont certes effectué plus de fixations oculaires sur Halo 3 que sur Battle (Figure 25) mais des saccades plus amples sur Halo 3 que sur Battle (Figure 26).

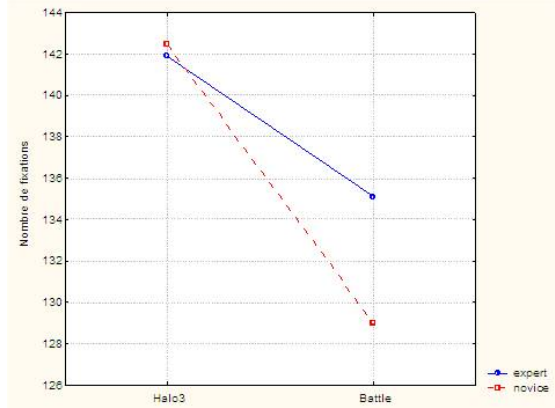


Figure 25. Nombre de fixations observées sur les deux premières minutes pour chaque jeu FPS selon l'expertise du joueur

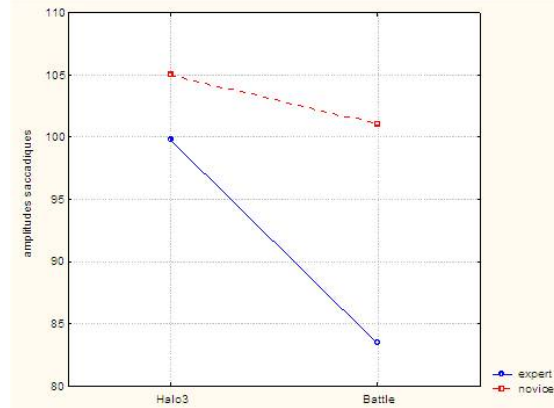


Figure 26. Moyenne des amplitudes saccadiques observées sur les deux premières minutes pour chaque jeu FPS selon l'expertise du joueur

En conclusion

Les différentes analyses effectuées nous suggèrent que Halo 3 permet une meilleure apprenabilité, meilleure comprenabilité et une meilleure exploration que Battle. En revanche ces mêmes analyses nous indiquent que Battle permet une meilleure localisation de l'information que Halo3. Ce dernier résultat peut être expliqué par le fait que nous avons privilégié l'équivalence du temps initial d'apprentissage de l'interface. Ce qui nous a amené à considérer les éléments fixes de l'interface visuelle comme cibles de nos questions. Même si les questions sont strictement les mêmes et portent donc sur des éléments d'égale importance, l'interface fixe d'Halo 3 est bien plus fournie en commandes que celle de Battle. On peut donc penser que les deux jeux n'étaient pas assez équivalents de ce point de vue pour être comparés de cette façon.

Typicalité des jeux

Le but de ce test est de voir dans quelle mesure des jeux définis comme faisant partie de la même catégorie (FPS, RTS, Casual Games, RPG...) permettaient de développer les mêmes patterns de fixations oculaires. Pour ce faire, nous avons pour chaque jeu testé, étudié comment se répartissent les fixations oculaires des participants. Le but était alors d'avoir une sorte de signature d'une catégorie de jeu qui permettra par la suite d'évaluer le degré d'appartenance d'un nouveau jeu à une catégorie de jeu et quantifier la bonne ou mauvaise catégorisation du jeu en question.

- **Méthode**

Participants. Nous avons recruté quatorze participants pour réaliser l'échantillon de signatures nécessaire à l'évaluation de la typicalité d'un jeu nouveau. Ces participants ont été recrutés en partie dans les Universités de Paris 8 Vincennes Saint-Denis et de Paris 10 Nanterre la Défense. Nous avons recruté notre population de cette manière pour nous permettre de pouvoir faire de nombreux tests avec un même participant. Comme cette méthode impose de jouer à de nombreux jeux, il fallait pouvoir garder un contact suffisant avec notre échantillon pour qu'il puisse revenir plusieurs fois. Les participants étaient donc des étudiants et enseignants, âgés de 21 à 33 ans et composés d'une majorité d'hommes (2 femmes). Les niveaux d'expertise étaient volontairement contrastés ; avec des joueurs novices et des joueurs très experts. Chaque participant a été dédommagé avec un bon d'achat d'une valeur de 10 euros.

Matériel : les Jeux. Nous avons évalué douze jeux se répartissant dans les quatre genres étudiés :

- 4 jeux de rôle (RPG)
- 5 jeux de tire à la première personne (FPS)
- 3 jeux casual

Ces jeux ont été choisis parce qu'ils pouvaient être typiques de leur genre (comme *Oblivion*, *Half Life 2*, *Command & Conquer 3* et *Amplitude*) ou parce qu'ils pouvaient être très atypiques (comme *Sims 2*, *Farhenheit*, *Stalker* et *Suprem Commander*).

Procédure.

Après avoir rempli un formulaire de participation librement consentie, les participants étaient invités à se placer devant le bureau où se situaient un clavier, une souris et l'oculomètre Tobii. Le fonctionnement de l'oculomètre ainsi que le type de fichiers de données recueillis lui était rapidement expliqué. Après une période de calibration, l'expérimentateur invitait les participants à jouer librement quelques minutes pour se familiariser avec le jeu testé, ce temps ne dépassait pas dix minutes (pour les novices en particulier). Lorsque le participant arrivait dans une séquence du jeu prédéfinie au préalable, l'expérimentateur lançait l'enregistrement oculométrique. L'enregistrement était arrêté au bout de cinq minutes, temps nécessaire au recueil de la signature oculométrique du jeu.

Le participant pouvait continuer quelques minutes s'il le désirait. A la fin de la série de tests, l'expérimentateur remerciait le participant et lui expliquait la problématique et les enjeux de l'étude.

- **Résultats**

Comme le montre la *Figure 27*, les joueurs experts fixent moins de zones de l'interface que les joueurs novices ou intermédiaires. Par ailleurs comme supposé, les jeux type FPS présentent une densité spatiale plus faible que les autres types de jeux. La *Figure 28* illustre ces différences de densité spatiale selon le type de jeu.

Nous pouvons remarquer par exemple que pour les jeux casual comme *les Sims* les fixations oculaires se répartissent sur une bonne partie de l'interface sur le menu, en revanche pour les jeux FPS comme *Crisis* les fixations oculaires sont surtout dans le centre de l'interface. Les jeux RPG comme *Oblivion* quant à eux ont une répartition proche des FPS mais avec une utilisation plus intense de l'interface de commande à l'écran.

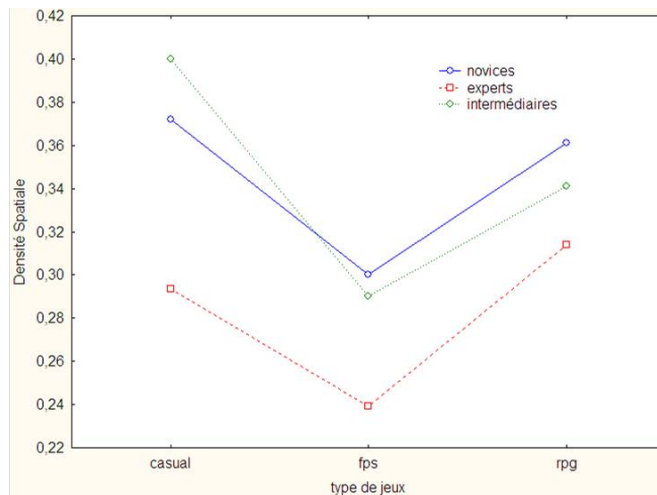


Figure 27. Moyennes des densités spatiales selon le type de jeux et le niveau d'expertise.

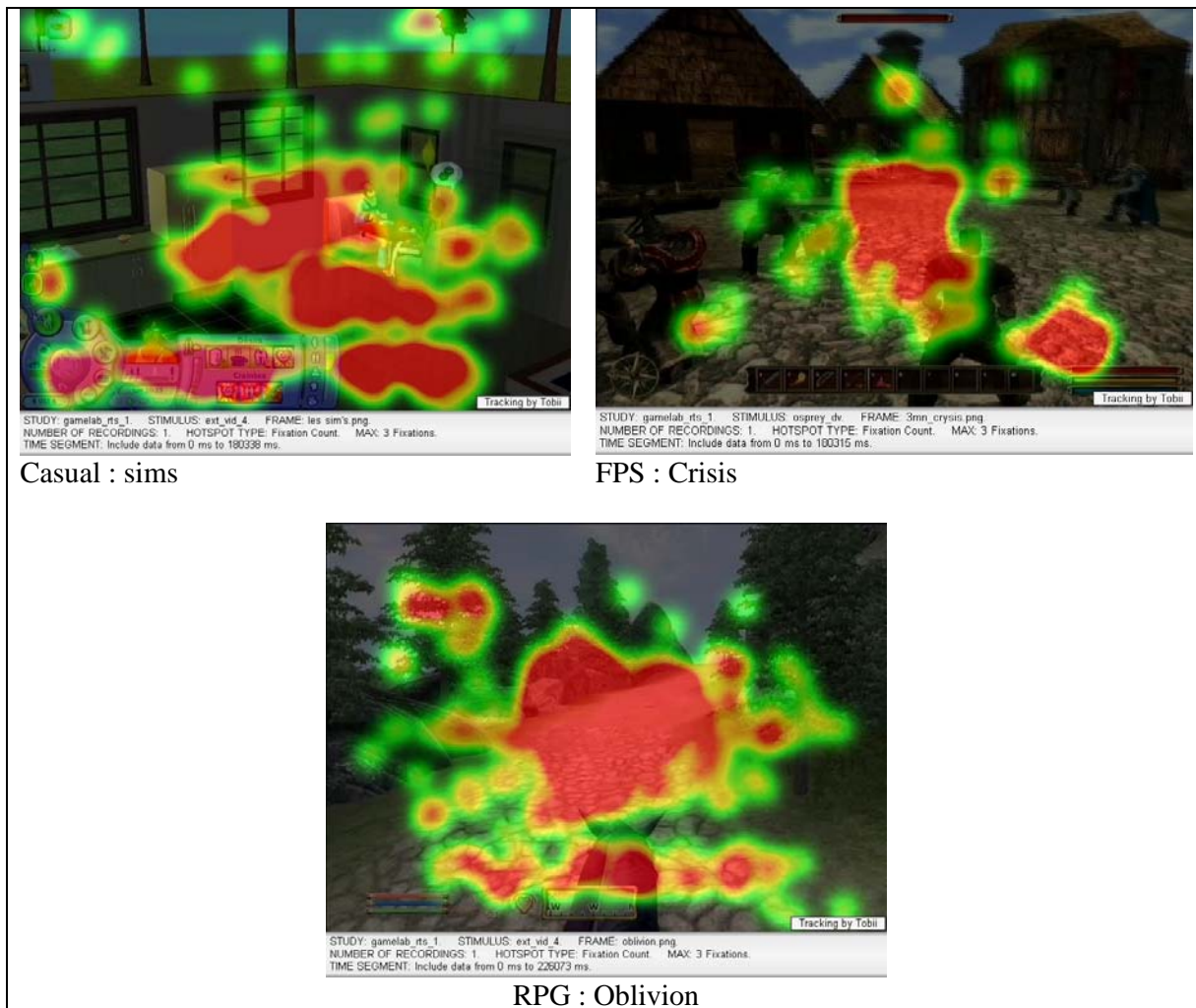


Figure 28. Cartes de Chaleur obtenues à partir des parcours oculaires d'un seul joueur sur chaque type de jeu.

DIFFICULTES RENCONTREES LORS DE LA REALISATION DES TESTS OCULOMETRIQUES.

Lors de la réalisation des tests, nous avons rencontré plusieurs difficultés liés aussi bien au dispositif technologique, qu'aux définitions de certains concepts sur lesquels nous basons notre méthode.

DISPOSITIF TECHNOLOGIQUE

Oculomètre portable

Dans le but d'étudier tous les types de jeux et tous les types de support de ces jeux (ordinateur, console...) nous avons acquis un système oculométrique portable. Le but alors était d'analyser les mouvements oculaires lorsque les participants étaient devant une télévision de salon avec une console. Le souci majoritaire que nous avons rencontré concerne l'analyse des données oculaires et vidéo issues de ces passations. En effet, lorsque le participant a le casque sur la tête et qu'il réalise une tâche, le système enregistre tout ce qui est dans le champ visuel de l'utilisateur. Cela peut être aussi l'écran de télévision sur lequel est retransmis le jeu mais également le mur, la manette ou l'expérimentateur. De ce fait pour isoler ces séquences il faut repasser pour toutes les passations les vidéos des jeux et indexer les temps correspondants pour les retirer pour des fichiers oculaires. Un test réalisé à la fin du projet a été mené en utilisant cet oculomètre sur la console Nintendo DS. Le fait de devoir segmenter les fichiers vidéo puis les fichiers oculaires fait que pour le moment nous n'avons toujours pas pu traiter les données issues de ce test et qu'il n'a pu être intégré dans ce rapport.

Cette difficulté de traitement liée au matériel explique que nous n'avons pu conduire plus de tests sur cet oculomètre. Pour les prochaines passations, nous entrevoyons la possibilité d'enregistrer une clé (un appui touche par exemple) dès que le participant ne regarde plus l'écran et une autre clé (un autre appui touche) dès qu'il regarde de nouveau l'écran afin de pouvoir automatiser à partir de ces clés la segmentation des fichiers oculaires à partir du logiciel « Lutin MOC ».

Oculomètre fixe

Notre laboratoire utilise depuis plusieurs années maintenant le système Tobii 1750 mais nous n'avons jamais été confrontés avant ce projet à la réalité de l'enregistrement des mouvements oculaires lors de jeux vidéo. En démarrant les tests au début du projet nous avons été mis face à des problèmes de paramétrage notamment vidéo qui nous ont pris beaucoup de temps à résoudre. Certains problèmes étaient liés à l'affichage du jeu vidéo sur l'écran du Tobii qui était décalé et de ce fait la position réelle du regard également : quand on demandait au participant de regarder un objet précis de l'interface, la position de son regard qu'on avait en retour était décalée d'autant. D'autres problèmes étaient liés au mode de sauvegarde des vidéos résultant du jeu : selon le type de fichier qu'on enregistrerait les tailles des fichiers oculaires sortant et surtout vidéo pouvaient prendre des tailles très imposantes (jusqu'à plusieurs dizaines de giga-octets pour une passation de 10 minutes pour un seul joueur). Enfin les données oculaires que nous avons en retour et le type d'analyses que nous pouvions faire nous a conduits à privilégier une analyse globale de la passation plutôt que des parties de cette passation, car pour étudier certaines parties il aurait fallu faire une segmentation vidéo manuelle de chaque passation pour repérer les moments qui auraient pu nous intéresser. De ce fait nous avons dû développer nos propres outils d'analyses des mouvements oculaires pour nous permettre d'automatiser très rapidement les traitements liés aux durées de

fixations oculaires, aux amplitudes saccadiques, au nombre de fixations et à leur évolution. Le développement de cet outil a pris du temps. Toutes ces raisons ont fait que nous n'avons pu réaliser plus de tests : au départ il était prévu que nous évaluions notre méthode sur les quatre types de jeu (nous n'avons eu le temps de le faire uniquement pour FPS et RTS) et d'autre part nous n'avons pu tester beaucoup de jeux avec beaucoup de joueurs pour les matrices de typicalité (dernier type de test présenté).

NIVEAU D'EXPERTISE

La définition de novices.

Pour attribuer que les difficultés rencontrées lors de la manipulation du jeu sont bien dues au jeu lui-même et non pas à la manipulation des différents périphériques comme, par exemple, la manipulation de la manette pour se déplacer dans l'espace, il faut s'assurer que ces novices ont déjà utilisé ce type de périphérique pour effectuer les actions requises par le jeu. Ainsi s'ils rencontrent des difficultés nous pourrions être certains qu'elles ne sont pas dues à la manipulation des périphériques du jeu. De ce fait les novices ne seront pas des vrais novices, c'est-à-dire des individus qui n'ont jamais été confrontés au matériel testé.

La définition d'experts.

Lors de la réalisation des tests une question s'est posée à propos des types d'individus que nous pouvons considérer comme expert du jeu : considère-t-on comme un expert, un joueur qui joue très régulièrement, ou alors un joueur qui joue très régulièrement à la catégorie du jeu ciblée par l'évaluation ou alors un joueur qui joue très régulièrement au jeu qu'on souhaite évaluer. Le choix entre ces trois niveaux d'expertise, et plus particulièrement entre les deux derniers niveaux, devra s'effectuer sur la détermination du niveau initial par lequel nous souhaiterons comparer les joueurs non-experts.

Ces deux points permettent de définir trois niveaux d'expertise qui pourraient être ceux qu'il faudrait retenir pour les évaluations, les tests à venir :

- Joueur qui n'est pas expert de la catégorie du jeu à évaluer
- Joueur qui est expert de la catégorie du jeu à évaluer, mais qui n'a jamais joué au jeu en question
- Joueur qui est expert dans le jeu à évaluer. Cette montée en expertise peut être réelle si le jeu à évaluer est un jeu du marché ou alors construite si on demande à ce joueur de jouer pendant plusieurs heures à ce jeu avant de lui faire passer des tests. La qualification de son niveau d'expertise pourra par exemple s'effectuer par le nombre de niveau de jeu auquel le participant aura accédé en l'espace d'un certain nombre d'heures par exemple.

EVOLUTION DES PATTERNS NOVICES – EXPERTS

Pour évaluer certains critères du jeu comme, par exemple, l'apprenabilité ou la compréhensibilité d'une interface nous avons basé nos analyses sur la comparaison entre les experts et les novices avec l'idée que si le jeu est bien évalué sur ces deux critères, les patterns des novices devraient ressembler au fur et à mesure du temps au pattern des experts. Comme nous avons pu le voir dans les analyses précédentes sur les deux catégories de jeux testés et donc sur les quatre jeux évalués les résultats ne présentent pas de tendances très nettes, sans appel. Plusieurs raisons peuvent expliquer ce constat.

D'une part la définition de l'expertise comme nous l'avons décrite plus haut peut en être à l'origine : nos experts n'étant pas forcément de véritables experts du jeu. Les comparaisons entre patterns selon nos méthodes ont pu être quelque peu biaisées par cette difficulté rencontrée dans la définition de l'expertise, qui peut revêtir des formes diverses dans de nombreux domaines, comme

dans la pratique du jeu vidéo. Le mode d'évaluation de l'expertise par recueil de l'expérience dans le domaine pourrait être avantageusement remplacé, comme il a été fait pour la méthode de mesure physiologique, par une évaluation des compétences de joueur lors du test lui-même. Cette éventualité limiterait cependant les possibilités d'évaluation de la typicalité du jeu.

D'autre part il nous apparaît que les durées de passations étaient trop limitées pour pouvoir observer l'aspect évolutif des patterns oculaires, notamment l'évolution de patterns novices vers des patterns experts. Ce biais peut se justifier par deux arguments :

- Premièrement ces expérimentations étant exploratoires, la qualité des enregistrements a été paramétrée à la hausse, pour nous permettre de les exploiter de façon multiple. Cette qualité supérieure d'enregistrement a représenté pour nous une contrainte technique importante, notamment dans le stockage des données ; pour exemple une passation de vingt minutes de jeu occupait seize giga-octets de mémoire.
- Deuxièmement notre méthodologie imposant des mesures appareillées pour comparer les jeux, nous avons demandé aux joueurs de jouer aux deux jeux à la suite, ce qui représentait pour les joueurs novices un exercice assez éprouvant à cause de leur manque de pratique des jeux et la pression liée au sentiment d'être évalué.

Il nous apparaît, à l'issue de cette série d'expérimentations, qu'un renforcement des tendances obtenues serait fortement probable en respectant davantage les deux contraintes décrites précédemment.

CONCLUSION.

Le but de ce sous-projet était de mettre en place une méthode d'évaluation ergonomique des jeux vidéo à l'aide des mouvements oculaires. Lors d'une première étude (livrable de novembre 2007) nous avons tenté de mettre en correspondance les notes données au jeu par les joueurs et les mouvements oculaires. Or les résultats que nous avons obtenus ne sont pas assez nets pour conclure de façon précise sur cette méthode. Par ailleurs, les résultats que nous avons obtenus lors des tests présentés ici au questionnaire nous suggèrent que les participants ne sont pas les mieux placés pour répondre aux questions d'utilisabilité qui leur ont été soumises même si un effort de vulgarisation été effectué.

C'est pour cette raison que dans cette étude, nous n'avons pas essayé de mettre en relation les notes d'évaluation des joueurs et les mouvements oculaires : les deux jeux n'ont pas obtenu de notes assez tranchées pour que l'analyse ait un sens. La méthode que nous présentons ici est plus porteuse à notre sens à condition de revoir le paradigme expert-novice et plus particulièrement la définition d'un novice et d'un expert ainsi que les modes de passations. Même si les résultats ne sont pas très nets, les tendances attendues sont bien là.

Nous obtiendrions certainement des effets plus nets de la qualité du jeu si nous effectuons des tests plus longs, quitte à ne pas enregistrer les mouvements oculaires lors de toute la passation mais à les enregistrer à des moments précis : soit en prenant comme référence le temps ou alors des moments du jeu (début et fin d'un niveau par exemple, ce qui pourrait permettre de mettre experts et novices au même niveau d'analyse). Certes elle n'est pas complètement finalisée, mais nous possédons enfin une méthode qui permet d'analyser les mouvements oculaires des joueurs (mesures objectives de l'utilisabilité du jeu – sa jouabilité) et qu'il faut encore affiner par de nouveaux tests utilisateurs. L'avantage de cette méthode est qu'elle permet d'évaluer très rapidement un ou plusieurs jeux vidéo avec 20-25 utilisateurs et ceci est un argument de poids pour les concepteurs de jeux vidéo. Ainsi pour l'évaluation d'un niveau comme nous l'avons traité ici nous pourrions effectuer l'analyse en 15 jours, passations incluses. Cette durée pourrait être plus courte si les concepteurs se posent des

questions précises comme, par exemple, la visibilité de certaines fonctionnalités ou de certains menus du jeu.

Recueil des données physiologiques

Dans la constitution de méthodes d'analyse ergonomique dans le cadre du projet LUTIN GameLab, le recueil de mesures physiologiques a été réalisé sur des joueurs pendant le cours de leur partie. Cet aspect de l'investigation ergonomique s'est attaché à l'aspect émotionnel de l'activité du joueur et à la définition de phases d'investissement affectif.

Les usages des nouvelles technologies génèrent de nouveaux contextes qu'il est important de cartographier. Il est ainsi crucial dans le développement d'une technologie multimédia de cerner les conditions d'immersion dans un environnement virtuel, les moments d'excitation ou d'ennui qui accompagnent une activité de recherche d'information ou d'amusement.

Dans le cadre particulier du jeu vidéo, il peut ainsi importer de saisir dans quelle mesure une information pourvue par le système de jeu est intégrée du point de vue de son impact sur le niveau de vigilance du joueur. Il est également important de connaître les phases et les variations naturelles de l'investissement du joueur pour apporter au moment opportun les stimulations capables de relancer son intérêt ou au contraire de conduire à un relâchement. Ainsi, une approche du gameplay à l'écoute des aspects viscéraux du comportement doit saisir le rapport entre les conditions d'éveil et de réceptivité affective du joueur et le déroulement de la narration dans un jeu.

De ce point de vue, les mesures physiologiques, qui récupèrent les variations de paramètres corporels associés aux états d'activation du système nerveux autonome, constituent un outil permettant l'acquisition directe des variations implicites de l'état affectif d'un individu. Elles ont l'avantage de ne pas réclamer d'interruption de l'action en cours et de ne pas demander de qualités introspectives particulières de la part des joueurs. Elles sont enfin potentiellement un indicateur beaucoup plus fin de l'évolution des états d'investissement du joueur, possédant un niveau de granularité qui ne saurait être atteint par le recueil du ressenti subjectif du joueur.

Cependant, une difficulté inhérente au recueil de ce type de mesure est leur faible niveau de discrimination à l'égard des qualités phénoménales des émotions éprouvées par un individu.

Il est reconnu que les mesures de l'activation du système nerveux autonome constituent un indicateur fiable du degré de vigilance (Lang, 1995). Elles peuvent être à même de fournir une évaluation de la valeur positive ou négative d'une émotion (Levenson, 1992), les tentatives de catégorisation des signaux physiologiques qui ont été mises en oeuvre (Ekman, Levenson & Friesen, 1983) sont controversées (Cacciopo & Tassinari, 1990) et la littérature offre un panorama broussailleux des capacités de discrimination à associer aux différents capteurs physiologiques.

Toutefois, des études récentes offrent des résultats encourageants pour l'application des méthodes physiologiques aux jeux vidéo. Ainsi Mandryk et al. (2006) ont pu mettre en évidence une indexation possible de l'investissement du joueur, associé à une mesure de challenge perçue, sur des signaux physiologiques recueillies pendant la partie. D'autre part, les travaux de Lisetti & Nasoz (2004) ont accrédité la possibilité d'obtenir une signature physiologique de certaines émotions pendant le visionnage de séquences tirées de films.

Nous avons travaillé sur les conditions de mise en oeuvre d'un recueil de signaux physiologiques fiables et ne perturbant pas l'activité du joueur. Nous avons décidé de concentrer nos efforts sur l'investigation d'un jeu d'action (jeu de tir à la première personne) caractérisé par son fort impact d'immersion. Nous avons ainsi mis l'accent sur la visualisation des variations de l'engagement au cours d'une partie et sur la possibilité de mettre à jour des signatures physiologiques capables de caractériser les attitudes affectives et cognitives dans des moments précis du jeu.

PHASE DE PRISE EN MAIN DU MATERIEL / ELABORATION DES METHODES

RECUEIL DE MESURES PHYSIOLOGIQUES SUR DES IMAGES FIXES

Pour nous familiariser avec le recueil de mesures physiologiques et pour évaluer le potentiel de ces mesures, nous avons procédé à un test des variations physiologiques en relation avec la présentation d'images à caractère émotionnel. Ainsi une sélection de 16 images a été présentée à 12 participants alors qu'étaient recueillies une mesure d'activité électrodermale et une mesure de fréquence cardiaque, sur la base d'une alternance ligne de repos (10 secondes) et présentation d'une image (6 secondes).

Nous ne sommes pas parvenus à mettre en évidence des réactions physiologiques caractéristiques. Ceci peut être expliqué par le choix arbitraire du matériel, les conditions de présentation peu immersives, et le nombre insuffisant de données recueillies pour chaque sujet.

Recueil de mesures physiologiques pendant une résolution de problème

Nous avons procédé à des tests sur le problème des anneaux chinois, dans lequel les participants devaient retirer des cinq pions selon une règle récursive qui stipule qu'on peut enlever ou mettre un pion s'il y a un pion juste à sa droite et rien au-delà. Ce problème est typiquement caractérisé par une phase d'exploration et une phase de planification pour parvenir au but, la première se manifestant par un grand nombre d'erreurs et de retours en arrière dans l'espace problème.

Nous avons testé 10 participants en recueillant l'activité électrodermale et la fréquence pendant qu'ils procédaient à la résolution du problème, en indexant sur les signaux physiologiques chaque coup marqué. Nous avons trouvé une tendance pour l'activité électrodermale à diminuer progressivement d'amplitude au moment de la transition entre la phase exploratoire et la phase de planification dans laquelle on peut considérer que la solution a été trouvée et les étapes nécessaires pour y parvenir sont mises en oeuvre.

MISE EN RELATION DES VARIATIONS PHYSIOLOGIQUES AVEC LE RYTHME NARRATIF D'UN JEU

INTRODUCTION

Un des aspects particuliers du jeu vidéo dans son rapport à la narration concerne la gestion de l'investissement psychologique du joueur. Du fait de la responsabilité particulière de ce dernier dans l'avancement de l'histoire, le game designer doit proposer des modes de narration qui soutiennent la détermination du joueur à progresser dans l'intrigue ou, plus simplement, à actionner les leviers du jeu pour observer leurs conséquences. Le travail de narration dans un jeu vidéo consiste alors à susciter les états émotionnels et cognitifs appropriés à la tâche en cours, à l'action à accomplir pour évoluer dans la partie. Le game designer peut être ainsi désireux de forcer le joueur à augmenter son niveau de vigilance pour le préparer à une épreuve sur le point de survenir, ou au contraire de réduire l'intensité du jeu pour inviter le joueur à une activité d'exploration ou de ravitaillement.

En suivant cette logique particulière de la narration dans un jeu vidéo, nous avons souhaité développer des mesures permettant l'indexation du niveau d'engagement du joueur au cours de la partie. Cette étude est en elle-même intéressante dans la mesure où, à notre connaissance, la description du parcours d'un joueur du point de vue des mesures physiologiques n'a jamais été réalisée. Elle permet d'autre part de valider les conditions de réalisation de tests à destination des entreprises pour l'évaluation de l'efficacité d'une séquence narrative.

Dans cette étude, nous avons utilisé une mesure d'activité électrodermale et une mesure de fréquence cardiaque pour évaluer les caractéristiques de la progression dans une séquence d'introduction du jeu Halo3, en nous basant sur la segmentation naturelle de la séquence permise par l'alternance entre des phases de narration classique dans lesquelles le joueur est spectateur de l'action avec des phases de jeu proprement dites dans lesquelles le joueur est au centre de l'action.

MATERIEL ET METHODE

12 participants ont été recrutés parmi les visiteurs de la Cité des Sciences et de L'Industrie, ainsi que parmi les étudiants de l'université Paris8. Les participants étaient âgés de 18 à 40 ans. La participation au test était récompensée d'une place au cinéma la Géode à la Cité des Sciences.

Bien qu'aucune expertise sur le jeu testé ne fût réclamée, les participants étaient recrutés sur la base de leur capacité à s'orienter dans un environnement numérique et à manipuler une manette de jeu sur console.

Un questionnaire préalable était rempli par les participants, détaillant leur aisance avec les technologies numériques, la fréquence et le temps consacrés au jeu vidéo, le choix de la plate-forme et le type de jeu vidéo le plus familier.

Les participants étaient amenés à jouer individuellement au jeu vidéo Halo 3 de Microsoft, jeu de tir à la première personne (FPS), sur la console de jeu Xbox360. Un bref tutoriel était proposé aux participants pour expliquer les commandes du jeu, dans lequel quelques minutes étaient laissées aux

joueurs pour se familiariser avec la manipulation de l'avatar lors de la séquence introductive du jeu.

Les participants avaient ensuite pour tâche de jouer sans interruption pendant un temps déterminé par la réalisation d'un parcours défini par avance. Tous les participants ont joué au niveau de difficulté Normal, c'est-à-dire le niveau le plus facile parmi trois niveaux de difficulté proposés par le jeu.

Les tests avaient lieu dans une salle du laboratoire LUTIN Userlab à la Cité des Sciences et de l'Industrie. La salle était équipée d'un téléviseur LCD Thomson Scenium (82 cm) ainsi que d'une caméra pointée vers l'écran du téléviseur. Les joueurs étaient assis à environ 1 mètre de l'écran.

Les signaux d'électrocardiographie et d'activité électrodermale (AED) étaient recueillis sur un amplificateur BIOPAC MP35 et visualisés par l'intermédiaire du logiciel BSL Pro.

Pour recueillir l'électrocardiogramme des participants, une Dérivation II de Einthoven a été utilisée : l'électrode négative était fixée sur le poignet droit et l'électrode positive était fixée sur la cheville gauche, l'électrode de référence étant fixée à la cheville droite.

Pour recueillir l'activité électrodermale des participants, nous avons posé les deux électrodes sur l'annulaire et l'auriculaire de la main gauche afin de laisser les trois autres doigts disponibles pour la manipulation de la manette. Les trois mesures étaient échantillonnées à 200 Hz.

La session commençait avec la présentation d'un film de deux minutes permettant la relaxation préalable du joueur et l'obtention de mesures de référence. Ce film représentait des formes géométriques abstraites évoluant lentement en modifiant progressivement leurs couleurs.

Le début et la fin du film étaient indexés dans les courbes physiologiques par l'envoi d'un bref signal sonore sur le troisième canal de l'amplificateur. Il était demandé aux participants d'éviter tout mouvement, en maintenant la manette de jeu calée sur les jambes et en plaçant les doigts équipés d'électrodes repliés sous la manette.

Après la présentation du film, le jeu était lancé représentant la séquence d'introduction et ses développements ultérieurs, jusqu'au moment où le joueur atteignait un point du jeu fixé au préalable, marquant la fin de la session expérimentale.

RESULTATS

SEGMENTATION DE LA SEQUENCE DE JEU

Pour cette expérience, nous avons retenu les valeurs d'activité électrodermale et de fréquence cardiaque associées à la partie introductive de la séquence de jeu testée, représentant environ cinq minutes de jeu, dans laquelle le joueur est amené à accomplir les premiers pas lui assurant une prise en main progressive des commandes et des règles du jeu. Du fait de sa position en ouverture du jeu, cette séquence est caractérisée par une difficulté limitée, avec une montée très progressive de l'intensité du jeu, le joueur étant amené à parcourir une certaine distance sans rencontrer le feu ennemi, jusqu'au premier contact avec une petite troupe d'ennemis, auquel fait suite une succession d'escarmouches permettant toujours un repli du joueur. Cette séquence introductive est de fait caractérisée par l'intention notable des développeurs d'amener progressivement le joueur vers le moment du combat en développant un crescendo matérialisé par l'évolution du thème musical,

initialement dépouillé et en retrait, jusqu'à son complet développement pendant la phase de combat, ainsi que par l'utilisation de dialogues scriptés évoquant l'imminence de la rencontre avec l'ennemi dans la première minute de jeu.

Du fait de la possibilité de segmenter la séquence de jeu en épisodes clairement délimités grâce à l'introduction par les développeurs d'action automatisées, nous avons ainsi la capacité à évaluer l'enchaînement de ces épisodes du point de vue des variations des mesures physiologiques. En particulier, nous étions intéressés par la traduction physiologique de la montée en puissance progressive au début du jeu, par l'évolution des paramètres physiologiques pendant une période de combat, ainsi que par l'impact de la redescende de l'action, immédiatement après le combat quand le joueur peut "souffler" et accéder à un autre niveau.

Nous avons segmenté les données recueillies en trois phases (voir Figure 1) correspondant à des unités d'action dans lesquelles le joueur est poussé à accomplir essentiellement un type de tâche (explorer l'environnement, combattre les ennemis, se ravitailler) et qui peuvent être définies par des attitudes psychologiques spécifiques (préparation à l'action, engagement dans l'action, relâchement). Ces trois épisodes successifs sont également délimités par trois différents contextes géographiques (la jungle, la rivière, un tunnel) ainsi que par l'alternance environnement fermé (couloir) / environnement ouvert (champ de bataille) qui caractérise respectivement le passage de l'épisode "Préparation" à l'épisode "Engagement", et le passage de l'épisode "Engagement" à l'épisode "Relâchement".

Les trois épisodes peuvent être qualifiés selon un certain nombre de paramètres objectifs. La Table 1 distingue la durée moyenne, le nombre moyen d'avertissements (événement dans lequel la barre d'énergie du joueur clignote, indiquant la nécessité de se mettre à couvert) et le nombre moyen de morts du personnage (se traduisant par l'arrêt du jeu).

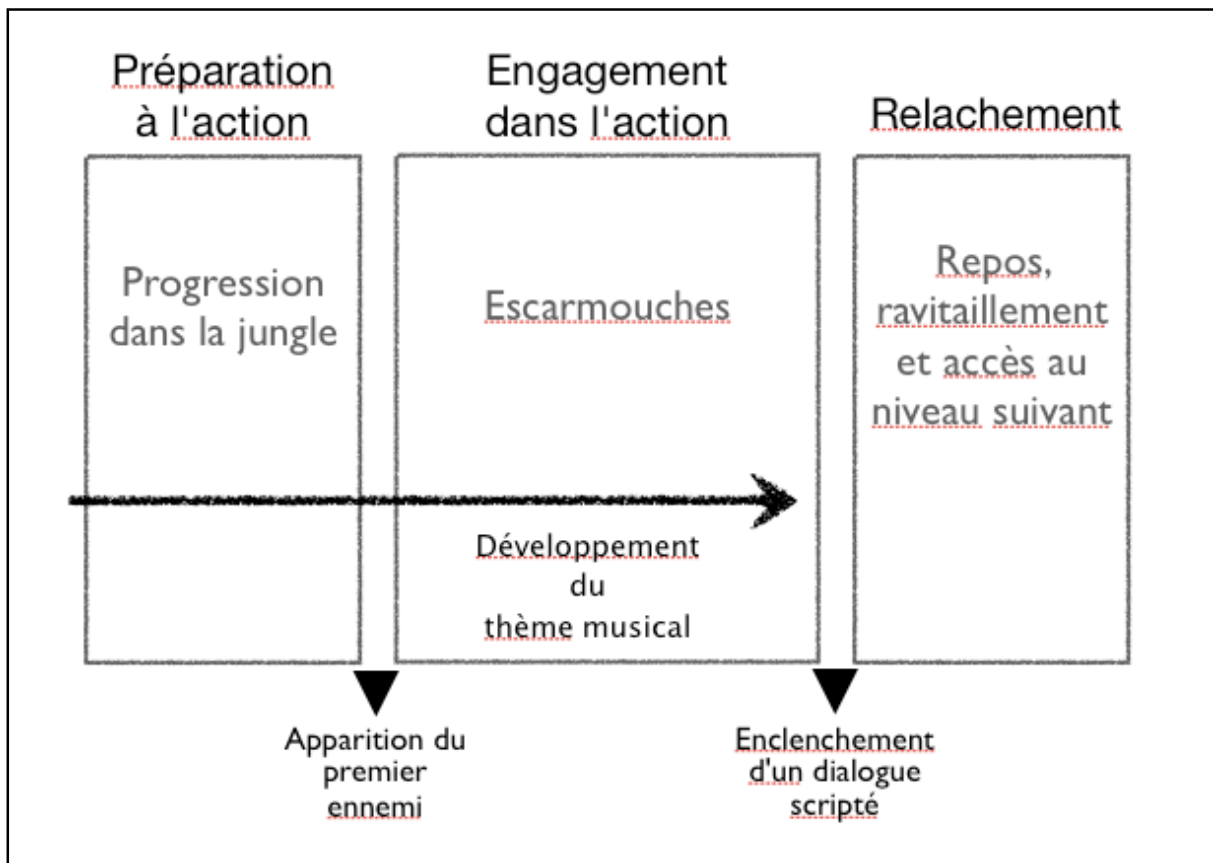


Figure 1. Segmentation de la période de jeu.

Préparation		Engagement		Relachement	
Durée moyenne	59 s	Durée moyenne	3 mn 36	Durée moyenne	54 s
Nbre moyen d'avertissements	0	Nbre moyen d'avertissements	0.5	Nbre moyen d'avertissements	0
Nbre moyen de morts	0	Nbre moyen de morts	0.58	Nbre moyen de morts	0

Table 1. Indicateurs objectifs associés aux trois épisodes.

Nos objectifs dans cette expérience étaient les suivants :

- déterminer la réactivité des mesures électrodermale et cardiaque à un engagement du joueur dans le combat
- établir la possibilité de justifier le découpage de la séquence de jeu à travers la variation d'amplitude des mesures physiologiques

Dans cette expérience, les joueurs pouvaient prendre le temps qui leur était nécessaire pour compléter chaque niveau. En conséquence, la durée des enregistrement varie de manière importante d'un joueur à l'autre, rendant difficile la comparaison directe des courbes et le moyennage des données. Afin d'égaliser le nombre de données pour chaque participant et pour les trois différents épisodes analysés, nous avons rééchantillonné les données selon une valeur déterminée par la référence aux durées de jeu les plus courtes (voir Table 2). Le nombre d'échantillons pour chaque participant et pour chaque période a été transformé selon le pourcentage de différence calculé par rapport aux durées de référence.

<u>Durées moyennes de jeu</u>		<u>Durées de référence</u>		<u>Nouvel échantillonnage</u>	
<u>Préparation</u>	59 s	<u>Préparation</u>	32 s	<u>Préparation</u>	6443
<u>Engagement</u>	3 mn 36	<u>Engagement</u>	2 mn 42	<u>Engagement</u>	32502
<u>Relachement</u>	54 s	<u>Relachement</u>	27 s	<u>Relachement</u>	5354

Table 2. Durées moyennes de jeu, durées de références (correspondant aux temps les plus courts), nouvel échantillonnage (nombre de données retenues), pour les trois épisodes.

Mesures physiologiques

Les données pour chacune des deux mesures ont été normalisées en considérant chaque échantillon comme une fraction de l'étendue maximale des données sur l'ensemble des trois épisodes. Ainsi, pour l'exemple de la mesure électrodermale, pour chaque participant, la valeur maximale d'AED et la valeur minimale d'AED obtenues sur l'ensemble des trois épisodes étaient utilisées.

$$AED \text{ normalisée } (i) = (AED(i) - AED \text{ min} / AED \text{ max} - AED \text{ min})$$

La même méthode a été utilisée pour normaliser les données de FC.

La Figure 2 représente le niveau d'amplitude moyen obtenu pour la mesure d'activité électrodermale et la fréquence cardiaque pour les trois différents épisodes analysés. Il apparaît que l'activité électrodermale subit une augmentation liée à l'engagement dans le combat, comme le confirme une différence significative entre les moyennes des épisodes Préparation et Engagement ($t(20) = 2.68$, $p = 0.014$). Le niveau d'amplitude de l'activité électrodermale tend à rester stable après la cessation des combats, comme le suggère l'absence de différence significative entre les moyennes des épisodes Engagement et Relâchement ($t(20) = 0.26$, $p = 0.798$). Les niveaux d'amplitude de fréquence cardiaque associés aux différents épisodes ne montrent pas de différences marquées. Ainsi la différence entre les épisodes Préparation et Engagement n'est pas significative ($t(20) = 1.02$, $p = 0.319$), de même que la différence entre les épisodes Engagement et Relâchement ($t(20) = -0.57$, $p = 0.57$).

Ces différences peu marquées au niveau des amplitudes moyennes peuvent masquer des évolutions au cours du temps des mesures en relation avec l'agencement des événements de jeu. En conséquence nous considérons les variations globales d'amplitude dans les trois différents épisodes du jeu. Les Figures 3 et 4 montrent le déroulé des mesures d'AED et de FC dans les épisodes Préparation, Engagement et Relâchement.

En ce qui concerne l'activité électrodermale, nous constatons dans l'épisode Préparation une tendance à l'augmentation du niveau de conductance de la peau, manifestée par une régression linéaire positive ($r = 0.64$). Cette tendance est inverse pour la mesure de fréquence cardiaque, avec une tendance au ralentissement des battements du coeur au cours de l'épisode ($r = -0.83$). Les deux mesures présentent de fait une corrélation négative ($r = -0.53$). L'augmentation de l'amplitude de la mesure électrodermale se poursuit dans l'épisode Engagement ($r = 0.65$), alors que l'évolution de la fréquence cardiaque ne montre pas de tendance caractéristique ($r = 0.12$). Après la fin des combats, la mesure électrodermale réagit par une tendance à la diminution d'amplitude ($r = -0.73$), la fréquence cardiaque ayant au contraire tendance à augmenter ($r = 0.64$). Les deux mesures sont en effet inversement corrélées ($r = -0.5$).

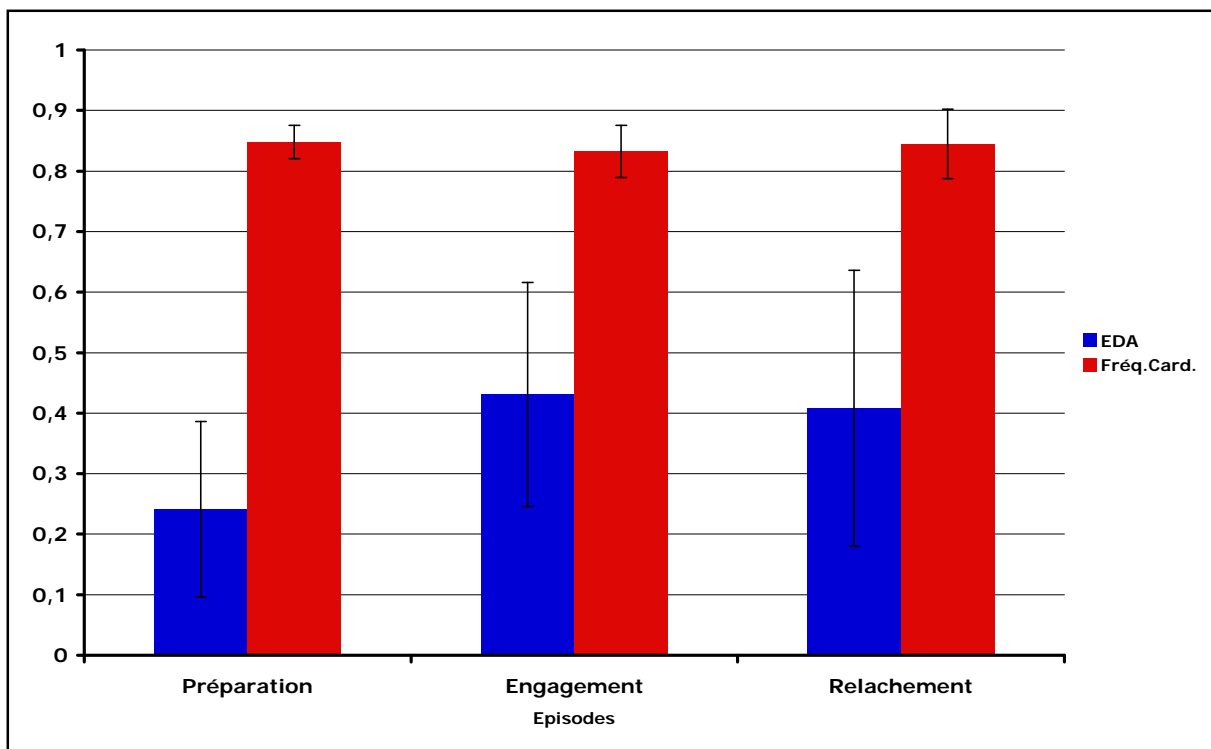


Figure 2. Moyennes et écart-types d'activité électrodermale et de fréquence cardiaque pour les trois épisodes.

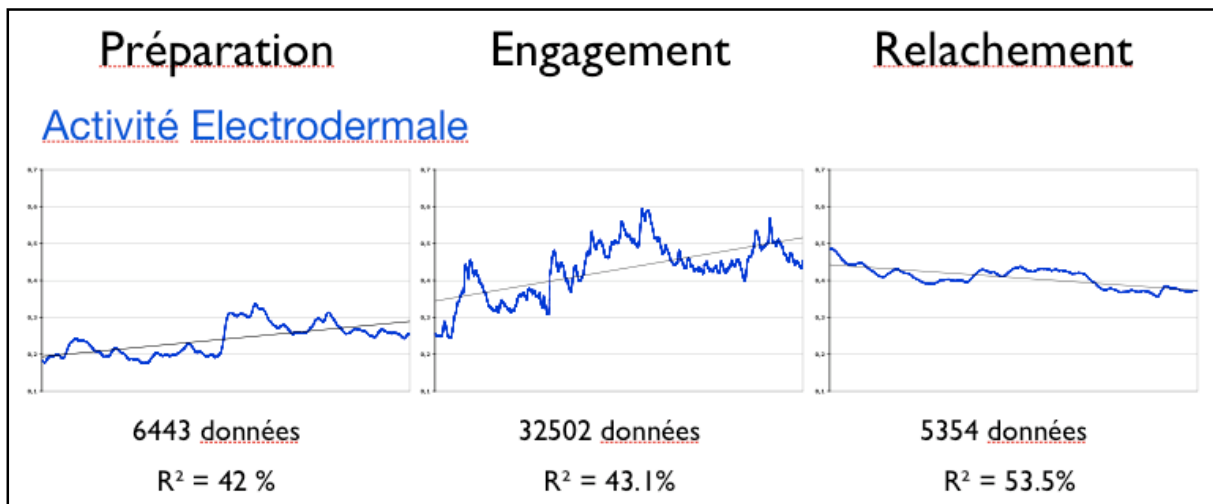


Figure 3. Décours de l'activité électrodermale selon les épisodes de jeu (environ 4 min).

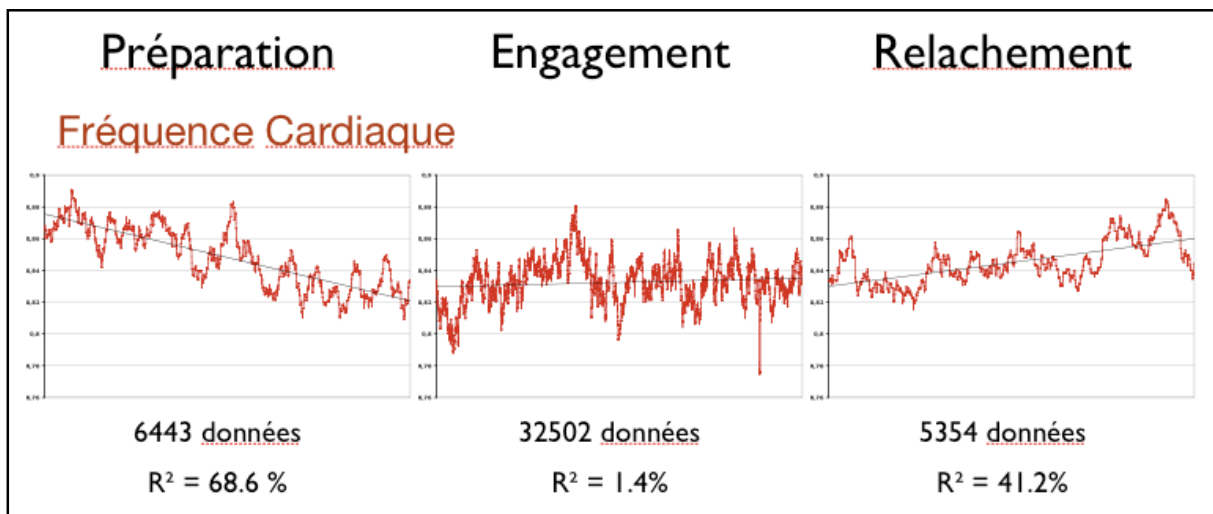


Figure 4. Décours de la fréquence cardiaque selon les épisodes de jeu (environ 4 min).

DISCUSSION

Dans cette expérience, nous voulions pouvoir décrire l'évolution des mesures d'activité électrodermale et de fréquence cardiaque du point de vue de l'organisation narrative d'une séquence de jeu. Nous voulions ainsi pouvoir évaluer l'efficacité de ces mesures pour caractériser les variations de l'engagement du joueur au cours d'une partie.

La première réussite de cette étude réside dans la mise en oeuvre du recueil lui-même. Nous craignons en effet que les signaux soient contaminés par des artefacts liés aux mouvements des joueurs pendant la partie. Dans les conditions de minimisation des mouvements qui étaient celles adoptées, les mesures restent cependant lisibles, et ceci tout en gardant la possibilité d'une immersion du joueur.

Nous obtenons en premier lieu une mesure satisfaisante de l'impact de l'engagement dans le combat par rapport à une séquence d'exploration avec une augmentation consistante de l'amplitude moyenne de l'activité électrodermale. Cette réactivité de l'AED aux changements d'implication du joueur est appuyée par des variations globales au cours du temps qui suivent l'engagement initialement progressif voulu par les développeurs du jeu et le retrait une fois les ennemis éliminés.

L'élévation de l'AED en relation avec l'engagement dans le combat peut-être interprétée comme une hausse de la vigilance du joueur, manifestée par une vasodilatation plus importante suite à l'activation du système orthosympathique. Cette variation traduit ainsi probablement la mobilisation de ressources attentionnelles rendue nécessaire par le suivi visuel des ennemis et de leurs projectiles. L'augmentation de l'amplitude de l'AED au cours de l'épisode Engagement peut de ce point de vue être envisagée comme une anticipation de l'effort à venir, anticipation engendrée par les indices disposés par les développeurs à propos de l'imminence du combat.

Les variations de la fréquence cardiaque restent moins aisément lisibles que celle de l'AED, les changements d'amplitude moyenne entre les épisodes n'étant pas significatifs. Nous observons cependant dans le décours de la séquence de jeu des variations allant dans le sens opposé de celles de l'AED. Il existe ainsi une tendance pour la fréquence cardiaque à diminuer dans l'épisode Préparation. Cette diminution traduit-elle une anticipation des séquences d'action à venir? La baisse de la fréquence cardiaque dans un moment de plus grande intensité peut apparaître paradoxale, mais l'absence d'implication générale du corps dans cette situation rend de fait inutile l'envoi de sang dans les parties périphériques et donc une élévation de la fréquence de contraction des muscles du coeur. Cette baisse peut-être être reliée à la diminution de la fréquence cardiaque observée parfois dans la présentation de stimuli à caractère positif.

RECHERCHE DE PROFILS PHYSIOLOGIQUES ASSOCIES A DES SEQUENCES TYPIQUES D'UN JEU

INTRODUCTION

Les résultats de l'expérience précédente suggèrent un intérêt des mesures physiologiques pour la visualisation des variations de l'engagement du joueur pendant la partie. Cette visualisation reste cependant limitée à une description en termes de positionnement sur un gradient de plus ou moins grande vigilance, ou de plus ou moins grande excitation. Les valeurs physiologiques restent également attachées pour leur interprétation à la description des événements de jeu et ne peuvent être caractérisées de manière indépendante en termes de marqueurs affectifs.

Nous souhaitons dans cette nouvelle expérience obtenir une caractérisation des signaux, indépendante du déroulement des événements représentés pendant le jeu, mais liée directement au ressenti subjectif suscité par ces événements.

En particulier, nous souhaitons obtenir une signature physiologique d'états psychologiques pouvant être considérés comme typiques de l'activité de jeu : un état de frustration lié à une progression difficile dans le jeu, opposé à un sentiment d'aisance engendré par la maîtrise du jeu ; un état de concentration associé à la réalisation d'une tâche difficile, opposé à un état de relâchement dû à l'absence de challenge particulier.

Nous avons cette fois étudié de petites séquences offrant un contraste au niveau de la facilité de progression, du sentiment de sécurité, des tâches à accomplir afin de susciter des réactions variées au niveau physiologiques ainsi que sur le plan du ressenti subjectif.

Nous avons d'autre part procédé au recueil de jugement subjectif en ciblant trois aspects:

- l'aspect cognitif (jugement de la part du joueur sur sa marge de progression dans la séquence),
- l'aspect émotionnel (évaluation du niveau de frustration ressentie pendant la séquence),
et
- l'aspect attentionnel (évaluation du niveau de concentration requis pour compléter la séquence).

Matériel et Méthode

21 participants ont été recrutés parmi les visiteurs de la Cité des Sciences et de L'Industrie, ainsi que parmi les étudiants de l'université Paris8. Les participants étaient âgés de 16 à 36 ans (moyenne d'âge : 24.1). La participation au test était récompensée d'un bon d'achat d'une valeur de 10 euros.

Bien qu'aucune expertise sur le jeu testé n'était réclamée, les participants étaient recrutés sur la base de leur capacité à s'orienter dans un environnement numérique et à manipuler une manette de jeu

sur console.

Un questionnaire préalable était rempli par les participants, détaillant leur aisance avec les technologies numériques, la fréquence et le temps consacrés au jeu vidéo, le choix de la plate-forme et le type de jeu vidéo le plus familier.

Les participants étaient amenés à jouer individuellement au jeu vidéo Halo 3 de Microsoft, jeu de tir à la première personne (FPS), sur la console de jeu Xbox360. Un bref tutoriel était proposé aux participants pour expliquer les commandes du jeu, dans lequel quelques minutes étaient laissées aux joueurs pour se familiariser avec la manipulation de l'avatar lors de la séquence introductive du jeu. Les participants avaient ensuite pour tâche de jouer successivement à quatre séquences à différents endroits du jeu.

Les tests avaient lieu dans une salle du laboratoire LUTIN Userlab à la Cité des Sciences et de l'Industrie. La salle était équipée d'un téléviseur LCD *Thomson Scenium* (82 cm) ainsi que d'une caméra pointée vers l'écran du téléviseur. Les joueurs étaient assis à environ 1 mètre de l'écran.

Les signaux d'électrocardiographie, d'activité électrodermale et de fréquence respiratoire étaient recueillis sur un amplificateur BIOPAC MP35 et visualisés par l'intermédiaire du logiciel BSL Pro. Pour recueillir l'électrocardiogramme des participants, une Dérivation II de Einthoven a été utilisée : l'électrode négative était fixée sur le poignet droit et l'électrode positive était fixée sur la cheville gauche, l'électrode de référence étant fixée à la cheville droite.

Pour recueillir l'activité électrodermale des participants, nous avons posé les deux électrodes sur l'annulaire et l'auriculaire de la main gauche afin de laisser les trois autres doigts disponibles pour la manipulation de la manette. Pour recueillir la mesure de fréquence respiratoire, nous avons utilisé une ceinture respiratoire placée sur l'abdomen des participants. Les trois mesures étaient échantillonnées à 200 Hz.

La session commençait avec la présentation d'un film de deux minutes permettant la relaxation préalable du joueur et l'obtention de mesures de référence. Ce film représentait des formes géométriques abstraites évoluant lentement en modifiant progressivement leurs couleurs. Le début et la fin du film étaient indexés dans les courbes physiologiques par l'envoi d'un bref signal sonore sur le troisième canal de l'amplificateur.

Il était demandé aux participants d'éviter tout mouvement, en maintenant la manette de jeu calée sur les jambes et en plaçant les doigts équipés d'électrodes repliés sous la manette. Après la présentation du film, le jeu était lancé et les participants devaient accomplir la séquence d'introduction du jeu, permettant au joueur une prise en main progressive des commandes et des règles du jeu, jusqu'à un point déterminé à l'avance par l'expérimentateur.

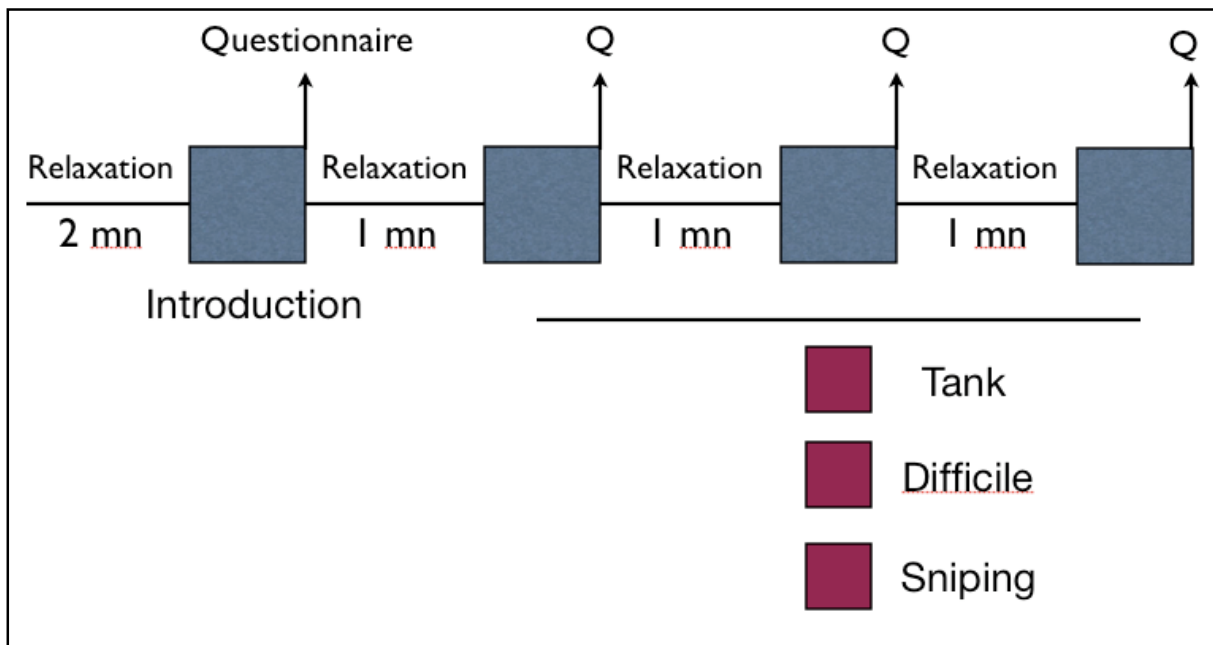


Figure 5. Organisation de la passation.

L'organisation générale des passations était la suivante (voir Figure 5): les participants étaient amenés à accomplir successivement quatre courtes séquences, y compris la séquence d'introduction. Chacune de ces séquences étaient immédiatement suivie d'un questionnaire portant sur l'évaluation de différents aspects ressentis (voir plus bas), puis, une fois le questionnaire complété, d'une nouvelle phase de relaxation d'une minute, similaire à la séquence qui débutait l'expérience.

Introduction		Tank	
Durée moyenne	4 mn 42	Durée moyenne	5 mn 08
Nbre moyen d'avertissements	0.35	Nbre moyen d'avertissements	0.1
Nbre moyen de morts	0.05	Nbre moyen de morts	0.05
Difficile		Sniping	
Durée moyenne	5 mn 16	Durée moyenne	4 mn 17
Nbre moyen d'avertissements	3.25	Nbre moyen d'avertissements	0.15
Nbre moyen de morts	7.1	Nbre moyen de morts	0.05

Table 3. Indicateurs objectifs associés aux différents épisodes.

Les trois séquences de jeu qui suivaient la séquence d'introduction amenaient le joueur à différents endroits du jeu, dans des situations très contrastées du point de vue de la scénarisation de l'action et

de la progression générale (voir Table 3 pour une évaluation de la durée moyenne de jeu et de la difficulté mesurée par les paramètres d'avertissement et de mort du personnage) :

- la séquence “**Tank**” amenait le joueur à conduire un tank. Cette séquence de jeu est linéaire et peu difficile, le joueur étant protégé des tirs ennemis et disposant lui-même d'une puissance de feu considérable.
- la séquence “**Difficile**” présentait une escarmouche dans le niveau de difficulté le plus élevé du jeu. La position défavorable de l'avatar (soumis aux tirs de tireurs embusqués), son faible équipement, la résistance des ennemis et leur férocité rendent la mort du personnage extrêmement probable dès les premières secondes de la séquence.
- la séquence “**Sniping**” dans laquelle le joueur est amené à prendre la position du tireur embusqué, pouvant tirer à distance sur les ennemis qui ne peuvent que riposter difficilement. L'avantage tactique du joueur et sa puissance de feu rendent la séquence aisée bien que requérant un maintien de l'attention pour cibler correctement les ennemis

A la fin de chacune des quatre séquences de jeu, était posée une série de quatre questions portant sur le ressenti du joueur pendant la partie :

- Cette séquence de jeu conduit à des moments d'impasse
- De manière générale, cette séquence de jeu réclame de la concentration
- De manière générale, cette séquence de jeu provoque de la frustration

Les réponses des participants étaient notées de 1 à 5, sur une échelle de Likert : 1. Pas du tout d'accord ; 2. Plutôt pas d'accord ; 3. Ni d'accord, Ni en désaccord ; 4. Plutôt d'accord ; 5. Tout-à-fait d'accord

Une quatrième question évaluait le plaisir de jeu globalement ressenti sur une échelle de 1 (très négative) à 5 (très positive)

Nous espérons offrir aux joueurs des situations suffisamment contrastées pour susciter des impressions différentes du point du ressenti et pouvant déboucher sur des profils physiologiques distincts. En particulier, nous espérons que la séquence “Tank” serait caractérisée par une valeur plus faible sur les échelles d'impasse, frustration et concentration que les autres séquences. A l'opposé, la séquence “Difficile” devrait conduire à des valeurs plus fortes sur ces échelles par rapport aux autres séquences. En position intermédiaire, la séquence “Sniping” devrait être caractérisée par une valeur forte sur l'échelle de concentration, faible sur les échelles d'impasse et de frustration.

RESULTATS

Evaluations subjectives

Pour chaque séquence de jeu, les participants procédaient à une évaluation des niveaux d'impasse, de frustration, de concentration et du plaisir de jeu sur une échelle de 1 à 5. Un test de chi carré appliqué aux distribution des réponses révèle que les séquences de jeu sont évaluées différemment, mais seulement pour les échelles d'impasse et de frustration (voir Table 4).

	<i>Introduction</i>	<i>Difficile</i>	<i>Sniping</i>	<i>Tank</i>	χ^2	p
Impasse	1.9	3.4	1.8	1.8	35.14	0.0004
Frustration	2.1	3.9	2	1.8	42.73	<.0001
Concentration	3.6	4.3	4.3	3.3	19.16	0.08
Plaisir	3.9	4	4	4	5.58	0.78

Table 4. Notes moyennes obtenues sur les différentes échelles et chi2 correspondant.

La Figure 6 représente la distribution des réponses sur l'échelle d'évaluation du sentiment d'impasse pour les quatre séquences de jeu. Conformément à nos attentes, la séquence "Difficile" est celle dans laquelle les participants choisissent majoritairement la valeur la plus élevée, tandis que la séquence "Tank" est caractérisée par une forte proportion de réponses inférieures à 3. La séquence "Sniping" est celle qui est considérée majoritairement comme la moins susceptible de conduire à des moments d'impasse.

La Figure 7 représente la distribution des réponses sur l'échelle d'évaluation du sentiment de frustration pour les quatre séquences de jeu. De manière similaire à l'échelle d'impasse, la majorité des réponses supérieures à 3 sont attribuées pour la séquence "Difficile", alors que les séquences "Tank" et "Sniping" sont caractérisées par des valeurs faibles sur l'échelle de frustration.

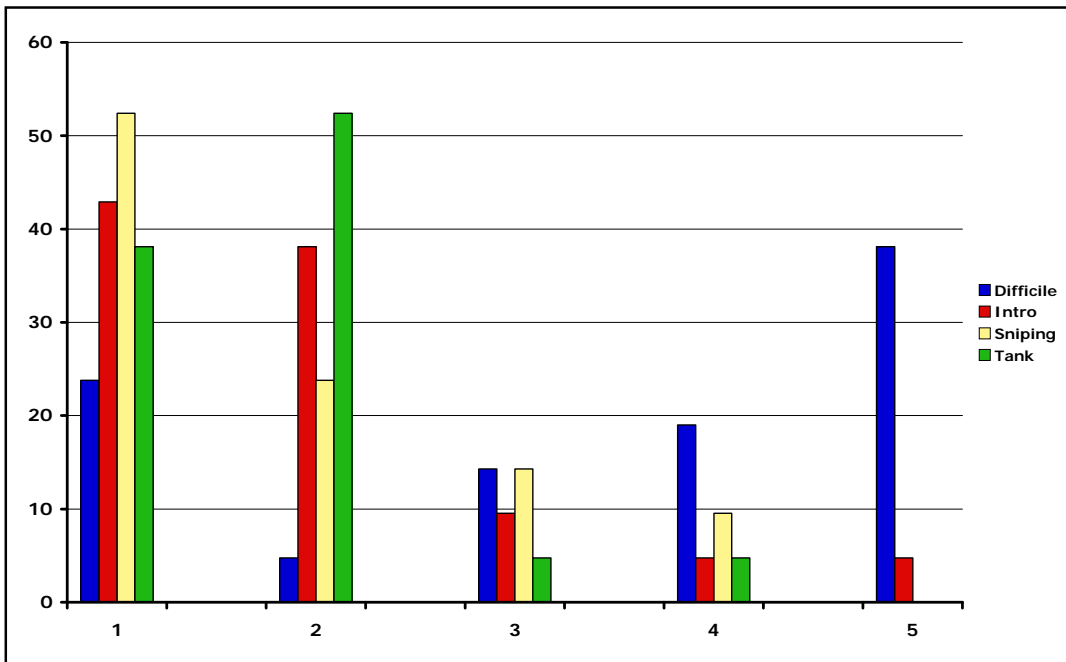


Figure 6. Distribution des réponses (en pourcentages) sur l'échelle d'impatse

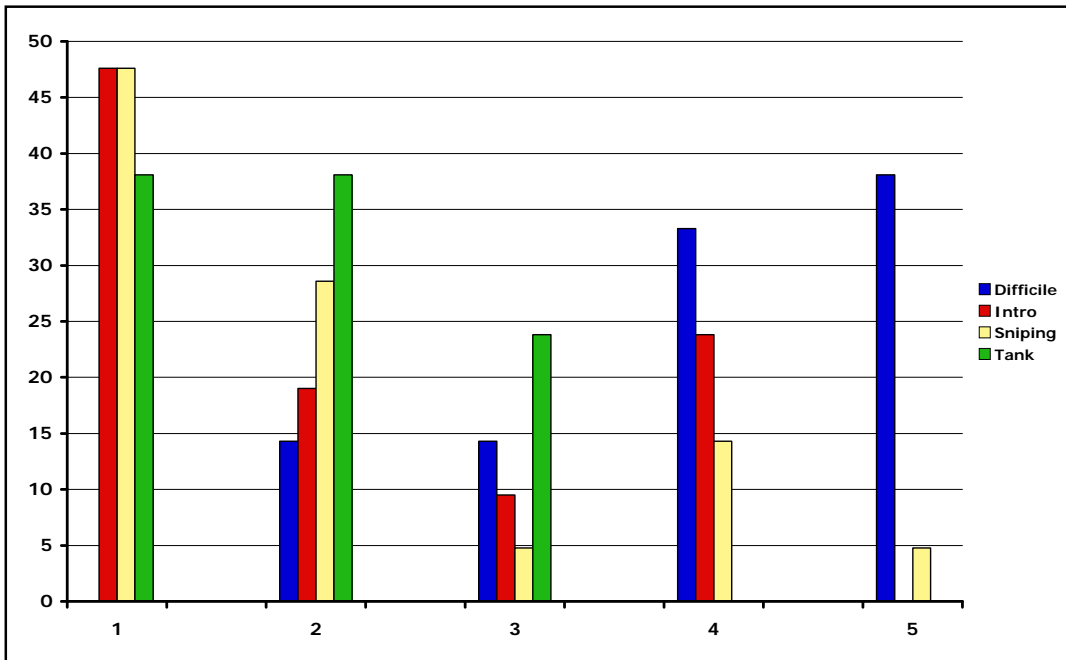


Figure 7. Distribution des réponses (en pourcentages) sur l'échelle de frustration

Mesures Physiologiques

Pour faciliter la manipulation des données, nous avons rééchantillonné les signaux physiologiques à 50 Hz. Nous avons retiré des analyses les données situées au-delà ou en-deça de 2 écart-types par rapport à la moyenne calculée.

Au niveau des amplitudes moyennes d'activité électrodermale, la condition Difficile est caractérisée par la plus forte amplitude, néanmoins les différences entre les trois moyennes sont extrêmement faibles. Une ANOVA prenant pour VI le type de séquence confirme une absence d'effet significatif de ce facteur ($F_{2, 17} = 0.99$, $p = 0.38$). Les valeurs de fréquence respiratoire pour chaque séquence sont également très similaires et ne sont pas caractérisées par un effet significatif ($F_{2, 16} = 1.35$, $p = 0.27$). De même, les moyennes de fréquence cardiaque pour chaque séquence ne sont pas significativement différentes ($F_{2, 17} = 1.05$, $p = 0.36$).

Corrélations entre mesures physiologiques et valeurs subjectives

Afin de corrélérer les trois mesures physiologiques aux quatre mesures subjectives nous avons centré-réduit chacun des échantillons à partir des moyennes et écart-types de chaque participant. Les résultats de la corrélation entre les valeurs physiologiques et les valeurs subjectives sont exposés dans la Table 5.

Valeur Subjective	Mesure Physiologique	Direction	Pearson	p
Frustration	AED	+	0.236	0.05
Impasse	AED	+	0.152	0.23
Concentration	AED	+	0.039	0.78
Plaisir	AED	+	0.097	0.49
Frustration	Fréq.Card.	-	0.133	0.29
Impasse	Fréq.Card.	-	0.268	0.04
Concentration	Fréq.Card.	-	0.284	0.05
Plaisir	Fréq.Card.	-	0.014	0.92
Frustration	Fréq.Resp.	-	0.217	0.07
Impasse	Fréq.Resp.	-	0.186	0.14
Concentration	Fréq.Resp.	-	0.221	0.11
Plaisir	Fréq.Resp.	+	0.024	0.86

Table 5. Mesures de corrélation entre les valeurs subjectives et les mesures physiologiques.

Nous pouvons constater l'absence de forte corrélation entre les différentes mesures. Une tendance se manifeste toutefois dans l'association entre des valeurs fortes d'AED et des valeurs fortes sur

l'échelle de frustration, entre des valeurs faibles de fréquence cardiaque et des valeurs fortes sur l'échelle d'impasse, et entre des valeurs faibles de fréquence respiratoire et des valeurs fortes sur l'échelle de frustration.

Signature physiologique de l'immersion dans le jeu

Nous avons échoué à mettre en évidence une signature physiologiques particulière pour chaque séquence de jeu. Il est néanmoins possible de montrer une signature globale marquant la transition entre un état neutre de repos et un état d'engagement.

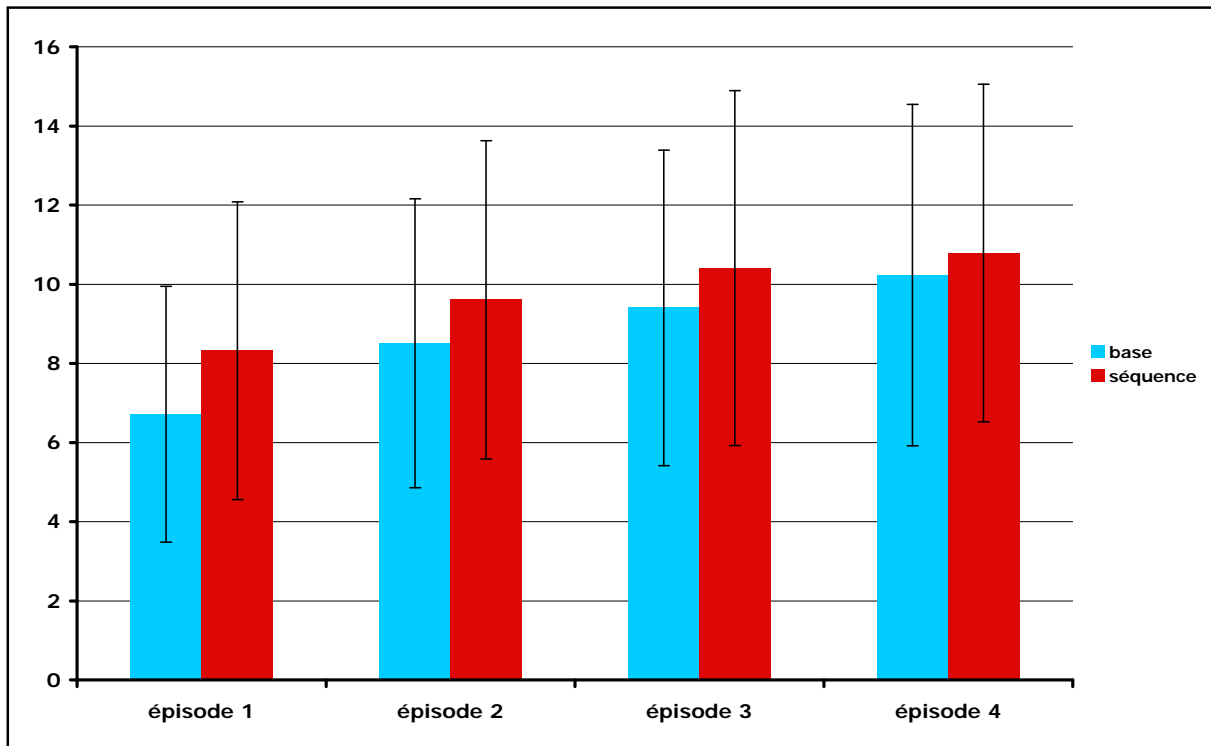


Figure 8. Moyennes et écart-types d'activité électrodermale associées aux lignes de base et aux séquences de jeu pour chaque épisode testé.

Lorsque nous examinons les résultats du point de vue du contraste entre les épisodes de jeu et la ligne de base qui les précédait immédiatement, et ceci indépendamment du type d'épisode concerné, nous observons pour la mesure d'AED (*Figure 8*) une augmentation générale de l'amplitude moyenne. Cette augmentation est significative pour tous les épisodes (*voir Table 6*). La mesure de fréquence respiratoire connaît également une élévation de son amplitude moyenne dans les phases de jeu par rapport à un état de repos (*Figure 9*). Cette élévation est significative pour tous les épisodes (*voir Table 6*).

La mesure de fréquence cardiaque n'est pas caractérisée par un type de contraste particulier entre état de repos et période de jeu (*Figure 10*), nous observons cependant pour les troisième et

quatrième épisodes une tendance pour la fréquence cardiaque à diminuer pendant les phases de jeu, une diminution qui n'est toutefois pas significative (*Table 6*).

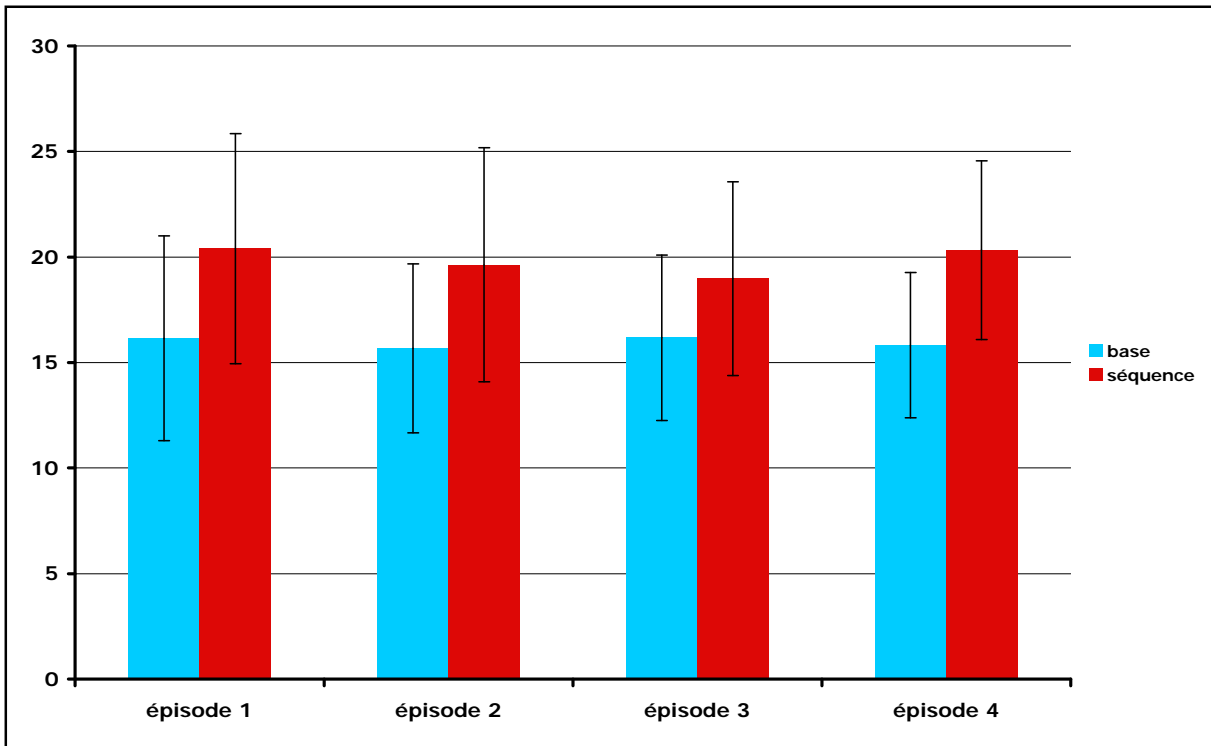


Figure 9. Moyennes et écart-types de fréquence respiratoire associées aux lignes de base et aux séquences de jeu pour chaque épisode testé.

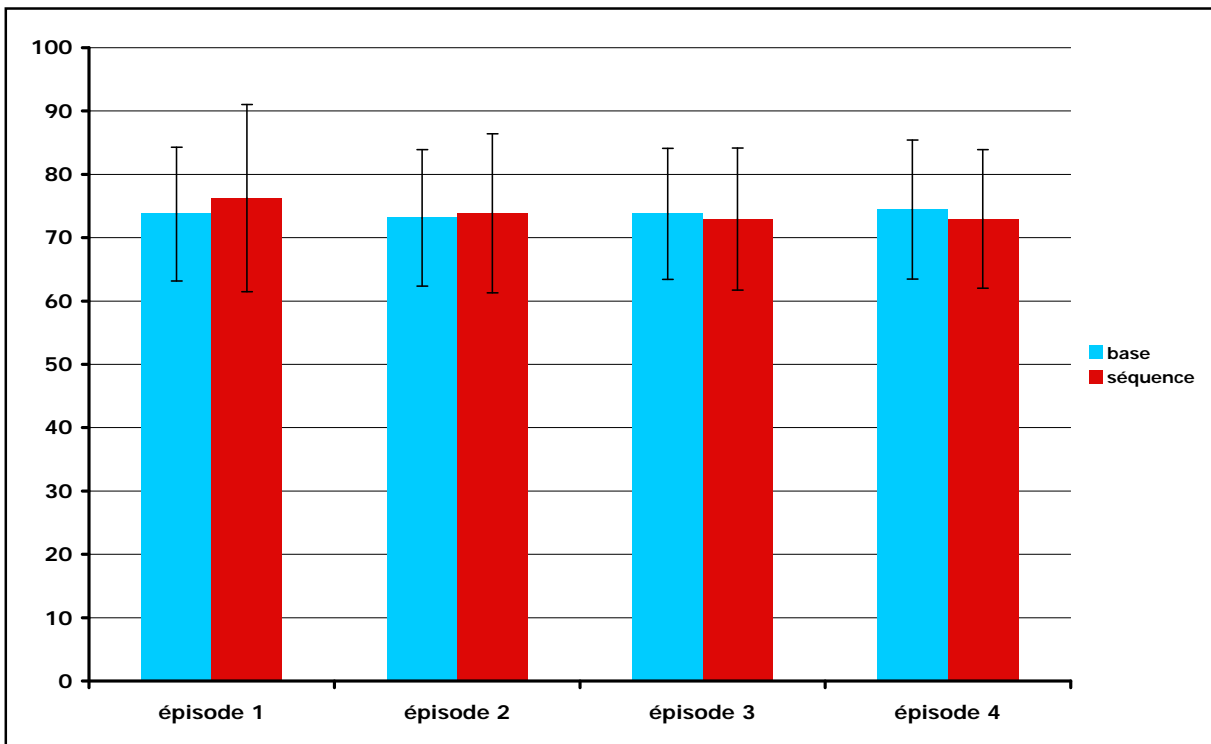


Figure 10. Moyennes et écart-types de fréquence cardiaque associées aux lignes de base et aux séquences de jeu pour chaque épisode testé.

	AED		Fréq.Card.		Fréq.Resp.	
	t	p	t	p	t	p
Episode 1	-6.36	<.0001	-1.5	0.14	-4.42	<.001
Episode 2	-3.5	<.01	-0.56	0.57	-3.43	<.01
Episode 3	-3.2	<.01	0.85	0.41	-2.9	0.01
Episode 4	-3.4	<.01	1.57	0.13	-5.98	<.0001

Table 6 . Différences entre ligne de base et séquence de jeu pour chacun des quatre épisode testé.

Discussion

Dans cette expérience, nous voulions valider la possibilité de distinguer des signatures physiologiques spécifiques attachées à des états affectifs ou cognitifs dans des situations de jeu typiques. Nous ne sommes pas parvenus à trouver des contrastes marquants au niveau des données physiologiques et au niveau de la relation entre les valeurs subjectives et les valeurs physiologiques.

Plusieurs problèmes peuvent être évoqués :

- les échelles utilisées pour les mesures subjectives de concentration et de plaisir de jeu ne sont pas utilisées pleinement et ne sont pas discriminantes pour les différents épisodes testés. Il est possible que ces mesures ne soient pas appropriées, leur concept restant flou, ou bien la manière dont est posée la question n'incitant pas à graduer finement l'impression ressenti.
- les épisodes de jeu ne sont peut-être pas adéquats pour susciter les variations attendues. En particulier Halo3 étant un jeu considéré comme réussi par la critique de jeu vidéo, il n'est peut-être pas à même de susciter un véritable état de frustration et conduit à un plaisir de jeu indépendant de la difficulté proposée, en raison de ses qualités intrinsèques de jouabilité.

Bien que nous n'ayons pas mis en évidence de signatures associées à l'état d'aisance ou à l'état de frustration, à l'état de concentration par rapport à l'état de relâchement, nous montrons toutefois une signature globale de l'investissement du joueur, que l'on peut supposer être un marqueur de l'état d'immersion dans un environnement virtuel, avec une augmentation franche et globale de l'amplitude d'AED et de fréquence respiratoire. A ceci peut-être ajoutée la tendance pour la fréquence cardiaque à diminuer dans certains épisodes par rapport à l'état de repos.

D'autre part, certaines tendances intéressantes peuvent être remarquées dans les données : une tendance pour l'activité électrodermale à augmenter avec la mesure de frustration (le niveau d'AED est le plus élevé dans la condition Difficile), une tendance pour la fréquence cardiaque à diminuer avec l'augmentation du sentiment d'impasse (le niveau de fréquence cardiaque est le plus faible dans la condition Difficile). Enfin, l'amplitude de la fréquence respiratoire est la plus faible dans la condition Sniping qui est elle-même celle évaluée comme nécessitant le niveau le plus élevé de concentration. Ces tendances dans nos résultats peuvent être résumées de cette façon : l'investissement rendu nécessaire par l'augmentation de la difficulté du jeu se traduit par une augmentation globale de l'activité électrodermale et par une diminution de la fréquence cardiaque (peut-être liée au sentiment d'excitation), la fréquence respiratoire, qui tend globalement à augmenter pendant les phases de jeu, connaît une diminution relative quand la tâche à accomplir réclame une certaine concentration.

Conclusion

Dans ce projet, nous avons réussi à mettre en place une méthodologie permettant l'acquisition de mesures physiologiques pendant une partie de jeu vidéo. L'utilisation de trois mesures a permis de d'envisager une indexation des moments d'investissement du joueur au cours de la partie. Cette indexation peut être recommandée à des développeurs lors de la phase de test du rythme narratif d'une séquence particulière, ou de la mise en valeur de l'impact de certains événements de relance émotionnelle. Nous disposons donc d'une mesure objective qui peut, avec certaines précautions, remplacer les tests habituels fondés sur le recueil du ressenti subjectif ou la visualisation de vidéo de jeu.

Nous avons toutefois échoué à mettre en évidence des profils clairs de mesures physiologiques permettant d'étiqueter certaines phases du jeu comme générant des émotions spécifiques, tels que la frustration ou l'ennui. La forte variabilité des mesures physiologiques pour chaque participant semble rendre difficile la mise en évidence de ce type de contrastes. Une analyse plus détaillée des protocoles individuels s'avère nécessaire pour obtenir des informations pertinentes. Dès lors, un travail de recherche sur les conditions d'établissement de signatures physiologiques et sur les mesures subjectives les plus appropriées reste à accomplir.

Références bibliographiques

- Albrecht, J. E. & et O'Brien, E. J. (1993). Updating a mental model: Maintaining both local and global coherence. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 19, 1061-1070.
- Alkan, S. & Cagiltay, K. (2007). Studying computer game learning experience through eye tracking. *British Journal of Educational Technology*, 38, 538-542.
- Baccino, T. (2004). *La lecture électronique*. Presse Universitaire de Grenoble
- Bagot, J.-D. (1996). *Information, sensation et perception*. Paris : Armand Colin.
- Bonnardel, N. (2000). Towards understanding and supporting creativity in design: Analogies in a constrained cognitive environment. *Knowledge-Based Systems*, 13, 505-513
- Cacioppo, J. & Tassinary, L. (1990). Inferring psychological significance from physiological signals. *American Psychologist*, 45, 16-28.
- Ekman, P., Levenson, R.W., & Friesen, W.V. (1983). Autonomic nervous system activity distinguishes among emotions. *Science*, 221, 1208-1210.
- El-Nasr, M.S. and Su Yan, S (2006). Visual Attention in 3D Video Games. ACE (ACM SIGCHI International Conference on Advances in Computer Entertainment Technology). Hollywood, June 14-16, 2006.
- Everatt, J., Bradshaw, M. E., & Hibbard, P. B. (1998). Individual differences in reading and eye movement control. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 223-242). Oxford, England: Elsevier.
- Goldberg, J. H., and Kotval, X. P. (1999). Computer Interface Evaluation Using Eye Movements: Methods and Constructs. *International Journal of Industrial Ergonomics*, 24, 631-645.
- Holyoak K., Gentner D., Kokinov B. (1998), *Advances in Analogy Research: Integration of Theory and Data from the Cognitive, Computational, and Neural Sciences*, *NBU Series in Cognitive Science*, New Bulgarian University, Sofia, 1998, (pp. 336-344).
- Jacob, R.J.K. & Karn, K.S. (2003). Eye Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises (Section Commentary). In J. Hyona, R. Radach, and H. Deubel (Eds): *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, (pp. 573-605), Amsterdam, Elsevier Science.
- Jacobs, A.M., & Lévy-Schoen, A. (1987). Le contrôle des mouvements des yeux dans la lecture : Questions actuelles. *L'Année Psychologique*, 87, 55-72.
- Johansen, S.A., Noergaard, M., & Rau, J. (2008). Can eye tracking boost usability evaluation of computer games?. Paper presented at: CHI 2008, Evaluating User Experiences in Games.
- Johnson-Laird, P. N. (1995). Mental models, deductive reasoning, and the brain. In M.-S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 999-1008). Cambridge (M.A.) : M.I.T. Press.
- Jönsson, E. (2005). If Looks could Kill – An evaluation of eye tracking in computer games. Master Thesis
- Just, M.A., & Carpenter, P.A. (1980). A theory of reading : From eye fixations to comprehension. *Psychological Review*, 87, (pp.329-354).
- Kenny, A. , Koesling, H., Delaney, , D., McLoone, S. & Ward, T. (2005). A Preliminary Investigation into Eye Gaze Data in a First Person Shooter Game. *Proceedings 19th European Conference on Modelling and Simulation*.
- Lang, P.J. (1995). The emotion probe. *American Psychologist*, 50, 372-385.
- Levenson, R. (1992). Autonomic nervous system differences among emotions. *Psychological Science*, 3, 23-27.
- Lisetti, C. & Nasoz, F. (2004). Using Noninvasive Wearable Computers to Recognize Human Emotions from Physiological Signals. *Eurasip Journal on Applied Signal Processing*, 11, 1672, 1687.
- Mandryk, R.L., Inkpen, K.M., & Calvert, T.W. (2006). Using psychophysiological techniques to measure user experience with entertainment technologies. *Behaviour & Information Technology*, 25, 141-158.
- Rayner, 1998 K. Rayner, Eye movements in reading and information processing: 20 years of research, *Psychological Bulletin* 124 (1998), pp. 372–422.
- Sanders, A. F. (1993). Processing information in the functional visual field. In G. d'Ydewalle & J. Van Rensbergen (Eds.), *Perception and cognition: Advances in eye movement research* (pp.3-22). Amsterdam: North Holland.
- Sennersten, C. & Lindley, C. (2008). Evaluation of real-time gaze logging by a 3D game engine. 12th IMEKO TC1 & TC7 Joint Symposium on Man Science Measurement, September 3 – 5, 2008, Annecy, France.
- Simon H. A. (1995). Problem forming, problem finding and problem solving in design. In A. Collen, & W. Gasparski (Eds.), *Design & Systems*. 245-257, New Brunswick : Transaction Publishers.

- Simon, H.A. (1973). The structure of ill structure of ill structured problems. *Artificial Intelligence*, 4, 181-201.
- Weil-Barais, A. et coll. (1993), L'homme cognitif, Paris, P.U.F, collection Quadrige.

ANNEXES

ANNEXE 1. ETUDE EXPLORATOIRE SUR L'ANALYSE DES MOUVEMENTS OCULAIRES DANS LES JEUX VIDEO.

2 – ADAPTATION DES METHODES ERGONOMIQUES AU JEU VIDEO

2.1 - Les méthodes ergonomiques dédiées au jeu vidéo

La recherche bibliographique a donné lieu à une mise en ligne des articles recensés sur un site Internet de type Spip (www.lutin-userlab.fr/gamelab). L'objectif était de communiquer l'ensemble des informations récoltées à tous les acteurs du projet GameLab. Le site comprend

- Une rubrique sur les heuristiques et les tests utilisateurs regroupe les articles portant principalement sur des analyses ergonomiques réalisées sur des jeux vidéo.
- Une rubrique sur le gameplay évoque les grands principes du game design.
- Une rubrique sur les mesures physiologiques répertorie des informations sur des mesures pouvant être effectuées pendant des tests sur les jeux vidéo (réponse électro-dermale, potentiels évoqués...?).
- Une rubrique intitulée "fiches méthodes" propose un inventaire des méthodes de test en ergonomie (protocoles verbaux, cheminement cognitif, entretiens...).
- La rubrique "synthèse des heuristiques" propose les résultats d'un travail collaboratif entre le Lutin et le Lip6 (laboratoire rattaché à Paris VI) concernant la synthèse et la mise en place d'heuristiques adaptables à l'analyse de jeux vidéo.

2.2 - Les méthodes ergonomiques adaptées pour le jeu vidéo

Dans la littérature anglosaxonne, les heuristiques d'analyse ergonomique sont celles de Nielsen. En France, ce sont celles de Bastien et Scapin.

2.2.1 - Adaptation des heuristiques de Nielsen

Melissa Federoff (*Federoff M. (2002). Heuristics and usability guidelines for the creation and evaluation of fun in video games, Ph D. Indiana University ; Sykes, J, & Melissa Federoff, M. (2006). Player-centred game design. CHI, 1731-1734*) a repris notamment les heuristiques de Nielsen afin de les adapter à l'analyse des jeux. Son étude constitue un point de départ dans le domaine de la recherche. Son objectif est de permettre ensuite la construction d'une liste standard d'heuristiques à l'usage de la communauté de développement des jeux vidéo. Elle a examiné les processus implicites et explicites d'une évaluation heuristique et les tests d'utilisabilité qui peuvent être réalisés par les concepteurs de jeux vidéo. Elle a observé cinq personnes faisant partie d'une équipe de conception durant une journée puis leur a fait passer des entretiens. Ses résultats montrent qu'instituer des processus plus formels d'évaluation des jeux s'est avérée très utile pour les concepteurs.

2.2.2 - Adaptation des heuristiques de Bastien et Scapin

La même démarche a été entreprise avec les heuristiques de Bastien et Scapin. Une étude similaire à celle de Melissa Federoff a été menée pour évaluer si les critères ergonomiques définis par Bastien et Scapin pouvaient également s'adapter aux jeux vidéo. Cette étude a permis de révéler que tous les critères s'appliquaient effectivement à l'analyse heuristique des jeux vidéo.

Nous disposons ainsi de deux listes d'heuristiques. Celle qui est adaptée de Bastien et Scapin a été testée et il reste à les intégrer.

3 – TEST D'HEURISTIQUES : LE COUPLAGE REUSSITE, OCULOMETRIE et HEURISTIQUES

Pour tester les heuristiques de Bastien et Scapin deux jeux vidéo ont été évalués : *Nancy Drew : la malédiction du manoir de Blackmoor* et *Nancy Drew : Danger au cœur de la mode*. Ces deux jeux vidéos sont des jeux d'enquête où le joueur doit résoudre plusieurs énigmes pour résoudre le problème posé au début du jeu.

3.1 - L'analyse d'heuristique des deux jeux vidéos

Une étude similaire à celle de Melissa Federoff a été menée pour évaluer si les critères ergonomiques définis par Bastien et Scapin pouvaient également s'adapter aux jeux vidéo. Cette étude a permis de révéler que tous les critères s'appliquaient effectivement à l'analyse heuristique des jeux vidéo (cf. Annexe 1 : Adaptation des critères de Bastien et Scapin).

L'analyse heuristique du jeu (annexe 2) « *Nancy Drew et la malédiction du manoir Blackmoor* » a montré tout d'abord un manque de contrôle utilisateur. Il n'est pas possible notamment d'interrompre les dialogues alors qu'ils sont parfois très longs, voire inutiles. D'autre part, on a relevé des difficultés liées à la gestion des erreurs. Typiquement, le jeu ne permet pas au joueur d'éviter, de comprendre ni de corriger ses erreurs. Par conséquent, il peut réitérer plusieurs fois les mêmes erreurs sans les comprendre, puis se retrouver bloqué dans le jeu, sans indice pour sortir de l'impasse. Cela peut amener le joueur à abandonner la partie.

L'analyse heuristique du jeu « *Nancy Drew : Danger au cœur de la mode* » (annexe 2) est tout à fait similaire à celle de « *Nancy Drew, la malédiction du manoir* », alors que « *danger au cœur de la mode* » est la dernière version du jeu sorti en France. On aurait pu s'attendre à des améliorations concernant l'interface, mais les difficultés pour les joueurs sont semblables. Principalement, on a observé des problèmes majeurs concernant la gestion des erreurs. A de nombreuses occasions dans le jeu, notamment lors de la résolution de certaines énigmes, le joueur peut faire des erreurs, mais le logiciel ne lui propose pas de message pour les comprendre ni les corriger. Le sentiment d'agacement susceptible de s'instaurer est alors inévitable. Toutefois, lorsque le joueur "meurt", la rubrique "seconde chance" du menu principal lui permet de revenir au dernier endroit où il se trouvait et de recommencer l'énigme en question. Dans ce jeu, la vie est infinie, mais si le joueur se sent frustré ou découragé car il ne réussit pas la tâche, il risque d'abandonner rapidement le jeu.

3.2 - Test utilisateurs des jeux vidéo

La méthode utilisée pour évaluer ces jeux vidéo se compose de deux tests utilisateurs : un test de jeu libre, appelé "phase exploratoire", et un test oculométrique.

Dans un premier temps, une phase exploratoire a été mise au point. L'objectif de ce test est de repérer quelles difficultés rencontrent les joueurs lorsqu'on leur propose de découvrir le jeu et d'y jouer librement, sans tâche imposée.

Ensuite, en fonction des problèmes soulevés par les testeurs, le test oculométrique a permis d'analyser les mouvements oculaires. L'objectif de ce test est d'observer quels éléments de l'interface sont susceptibles de poser problème chez le joueur, puis de mettre en corrélation les données obtenues avec les résultats des questionnaires d'utilisabilité.

3.2.1 – Phase exploratoire

4 joueurs ont été recrutés pour ce test. Ils sont tous étudiants de sexe masculin, âgés entre 24 et 27 ans. Ce sont des joueurs occasionnels plutôt attirés par les jeux de stratégie (RTS, Real Time Strategy) ou les jeux de tirs à la première personne (FPS, First Person Shooter).

Les séquences sélectionnées pour ce test sont les trente premières minutes du jeu. C'est ce qu'on appelle la phase d'accostage d'un jeu vidéo. L'accostage, proche du guidage assisté, permet au joueur de se familiariser avec l'interface, les périphériques, il découvre l'histoire et les personnages. L'utilisateur était invité à jouer librement à un premier jeu durant 25 minutes.

Fourni après les phases de jeu libre, le questionnaire de satisfaction permet de définir quantitativement l'appréciation des joueurs. Il s'agit principalement de découvrir les difficultés rencontrées pendant le jeu, de savoir ce qui leur a plu ou, au contraire, les a déçu, et quelles suggestions ils pourraient apporter si on leur demandait d'améliorer le jeu. Enfin, concernant les deux jeux, on leur demande lequel ils ont préféré et pourquoi. Puis la personne jouait au deuxième jeu. Notons que l'expérimentateur devait effectuer un contre-balancement en ce qui concerne l'ordre de présentation des jeux. De la même manière que précédemment, la personne jouait pendant 25 minutes puis répondait au questionnaire. A la fin de la passation, on demandait au joueur quel jeu il avait préféré, lequel il aimerait poursuivre, et éventuellement lequel il serait prêt à acheter.

D'après les résultats obtenus au questionnaire de satisfaction, on a pu détecter les difficultés principales rencontrées par les joueurs. Les types de problèmes remarqués sont les modes de déplacement, les graphismes et le manque de contrôle utilisateur. Les joueurs estiment que le mode de déplacement est difficile car il faut pointer le curseur à l'endroit où l'on souhaite se rendre puis cliquer. Ils pensent que ce

système n'est pas intuitif et qu'il serait préférable de concevoir un jeu complètement en 3D et pouvoir contrôler soi-même le personnage que l'on incarne. Ce problème relève plus du gameplay que de l'ergonomie, mais l'ensemble des joueurs l'a soulevé. Concernant les graphismes, les joueurs ont noté que les textures, les dessins des objets ainsi que des personnages, et plus globalement l'ensemble des décors, sont plutôt pixellisés, géométriques et statiques. Bien que le jeu soit homogène à ce niveau-là, cela constitue une gêne importante pour tous les joueurs ayant déroulé le test. Enfin, le dernier point essentiel concerne le manque de contrôle utilisateur. Effectivement, il est impossible dans le jeu de stopper ou de passer les cinématiques, souvent très longues, et les joueurs n'ont pas la possibilité d'interrompre les dialogues qui sont parfois trop longs voire inutiles.

3.2.2 – Tests oculométriques

Il existe peu d'études réalisées sur les jeux vidéo, particulièrement avec la méthode de l'oculométrie. Dans un premier temps, l'objectif de cette étude est d'observer quel type de données il est possible de recueillir, en quoi l'analyse des résultats peut donner lieu à des recommandations ergonomiques et vérifier la pertinence de la méthode de l'oculométrie combinée à la diffusion d'un questionnaire d'utilisabilité pour analyser les jeux vidéo dans le but de les améliorer. Ensuite, grâce au dépouillement des données oculométriques dans une situation de jeux vidéo, que peut-on déduire des mouvements oculaires ? Comment et en quoi peuvent-ils nous aider dans l'évaluation d'un jeu ? Ainsi, on suppose que si les mouvements oculaires peuvent nous permettre d'évaluer les jeux, alors on devrait observer des corrélations entre les mouvements oculaires et les notations aux critères ergonomiques (questionnaire d'utilisabilité inspiré des critères de Bastien & Scapin, annexe 3).

Participants - 10 testeurs ont été recrutés (4 femmes et 6 hommes) tous âgés entre 22 et 37 ans. Parmi ces 10 personnes, 5 sont étudiants, les autres exercent divers métiers (technicien CNRS, pharmacienne, infographiste...). Tous les utilisateurs sont à l'aise avec les technologies numériques et 8 testeurs jouent aux jeux vidéo soit quelques fois par mois soit moins d'une fois par mois, la plupart sur des PC. Ils jouent notamment à des jeux d'arcade, d'aventure, des RTS, des FPS ou des jeux de rôle. Les 2 autres personnes ne jouent pas du tout.

Dispositif technique – L'oculomètre utilisé (Tobii) est un système vidéo qui détermine les coordonnées (x,y) du regard à partir des reflets pupillaires des deux yeux. Le test est dirigé par deux ordinateurs reliés entre eux et à l'oculomètre : un des deux ordinateurs effectuent la capture vidéo de ce que voit l'utilisateur et l'enregistrement des mouvements oculaires, le second ordinateur gère le déroulement du jeu. Enregistrer les mouvements oculaires lors du déroulement d'un jeu vidéo sur PC nécessite deux ordinateurs pour éviter des ralentissements aussi bien au niveau du jeu que de l'enregistrement des mouvements oculaires dues à une demande de ressources informatiques très importantes par les différents systèmes (jeux et tobii).

Conception des scénarii - Concernant le choix des séquences et la conception des scénarios, il était nécessaire d'identifier deux scènes (une dans chaque jeu) possédant le plus de points communs pour qu'une comparaison des résultats soit possible. Les deux séquences se déroulent dans une seule pièce, il s'agit dans chacune de résoudre une énigme nécessitant un minimum de réflexion, et si jamais le joueur commet une erreur, la conséquence est irrémédiable car soit il est renvoyé dans l'une, ou bien il meurt dans l'autre. Chaque séquence est jouée pendant un maximum de 20 minutes.

Séquence de « Nancy Drew, la malédiction du manoir Blackmoor » - Le joueur commence une nouvelle partie. Il découvre l'introduction du jeu : présentation de l'histoire, du mystère à résoudre, des personnages. Ensuite, l'expérimentateur charge une partie pour amener le joueur à la séquence à analyser avec l'oculomètre. Le joueur se trouve dans un couloir du manoir. L'expérimentateur va lui présenter Loulou le perroquet. Il lui explique que pour obtenir un indice de la part de Loulou, il va devoir lui concocter un gâteau. L'expérimentateur montre la pièce où se trouvent les ingrédients pour le réaliser. Mais pour ne pas intoxiquer le perroquet, le joueur doit impérativement trouver la recette lui expliquant quels ingrédients il ne faut surtout pas donner à Loulou. Car la conséquence serait fatale : le perroquet intoxiqué rendrait l'âme aussitôt et le joueur serait inexorablement remercié du manoir.

Séquence de « Nancy Drew : danger au cœur de la mode » - Le joueur commence une nouvelle partie. L'introduction lui présente le contexte, l'histoire, les personnages. Etant donné que

dans ce jeu il n'est pas possible de récupérer le fichier de sauvegarde, l'expérimentateur prend le jeu en main et amène le joueur jusqu'à la scène où doit se dérouler l'analyse des mouvements oculaires. Le joueur découvre le studio d'un photographe avec qui il doit discuter. Sa mission est de faire des tirages de photographies. Dans un premier temps, il doit chercher le manuel explicatif caché quelque part dans une étagère du studio. Une fois le manuel trouvé et les informations mémorisées, le joueur doit se rendre dans la chambre noire. C'est ici qu'il doit réaliser les tirages. La grande difficulté est de se retrouver dans le noir total car il est impératif d'éteindre complètement les lumières. Le testeur doit alors se souvenir de l'emplacement des trois bacs contenant les produits chimiques. Et si par malheur il clique à côté de l'un des bacs, le résultat est immédiat : tout le studio explose !

Conception du questionnaire - Le questionnaire proposé aux joueurs porte sur le respect des critères de Bastien et Scapin. Chaque critère a été reformulé simplement pour être accessible à des personnes ne connaissant pas l'ergonomie (voir annexe 3 pour l'intégralité du questionnaire). Les joueurs devaient répondre sur une échelle de Likert en 7 points (1 : pas du tout d'accord et 7 : tout à fait d'accord) afin d'exprimer leur opinion quant au respect ou non de chaque critère.

Déroulement du test - Les 10 utilisateurs ont été recrutés dans les couloirs de la Cité des Sciences. Lorsque le joueur arrivait dans la salle de test de l'oculomètre, on lui demandait de s'installer confortablement devant le Tobii. Pendant 5 minutes, on lui présentait le test. On lui expliquait qu'il allait jouer à deux jeux vidéo. Après chaque jeu, il devait répondre à un questionnaire. Ensuite, on lui proposait de remplir le pré-questionnaire servant à définir le profil des joueurs, ainsi qu'un formulaire lui demandant son accord pour la passation et le fait d'être filmé. Une fois ces papiers remplis, on demande à l'utilisateur de se positionner confortablement devant le Tobii, de préférence en mettant les coudes sur la table afin de stabiliser au maximum la position des yeux. On lui expliquait qu'il devait éviter de bouger la tête, sinon le Tobii risquait de perdre son regard. Dès que le joueur était prêt, l'expérimentateur lui expliquait qu'avant de jouer, il passerait par une phase de calibration, c'est-à-dire qu'il devrait suivre des yeux un point bleu apparaissant sur l'écran. Ce point prenait successivement 9 positions différentes à l'écran. On démarrait alors la calibration. Après la calibration, on lançait le premier jeu. Notons que l'ordre de présentation des jeux a été contre-balancé afin d'éviter des biais d'expérimentation. Le joueur découvrait l'introduction et le didacticiel. Après l'introduction, l'expérimentateur proposait une consigne que la personne devait suivre. A ce moment, on démarrait l'enregistrement des mouvements oculaires. Au bout de 20 minutes, le jeu était interrompu. Le testeur devait répondre au questionnaire d'utilisabilité. Ensuite, on démarrait le deuxième jeu, qui se déroulait exactement de la même manière que pour le premier. Après avoir rempli le dernier questionnaire d'utilisabilité, l'expérimentateur offrait une place pour voir un film à la Géode en remerciement du temps consacré par la personne à l'expérimentation.

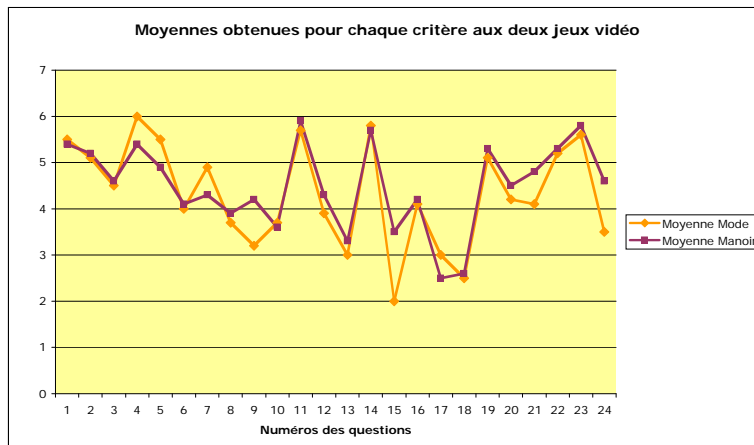


Un joueur devant le Tobii qui présente le jeu vidéo

Résultats - Nous nous attendons à observer des corrélations entre les notations au questionnaire d'utilisabilité et les mouvements oculaires. En effet, si les mouvements oculaires peuvent nous permettre d'évaluer les jeux, alors on devrait observer des corrélations entre les mouvements oculaires et les notations aux critères ergonomiques (questionnaire d'utilisabilité selon les critères de

Scapin et Bastien).

Questionnaire d'utilisabilité – le graphique 1 présentent les moyennes obtenues pour chaque critère par les deux jeux. Les notes vont de 1 (le critère n'est pas bien respecté) à 7 (le critère est bien respecté).



Graphique 1. Moyenne des notes obtenues par les deux jeux aux critères d'utilisabilité.

On s'est posé la question de savoir si les deux jeux avaient été évalués de la même manière par les joueurs. On a pu mettre en évidence une différence significative dans la notation attribuée par les joueurs uniquement à la question n° 15 : "Grâce au jeu, le joueur ne peut pas se tromper ni faire d'erreurs" [$t(9)=2,29$; $p<.05$].

Le tableau 1 récapitule les questions les mieux et les moins bien notées dans chaque jeu.

	Mode	Manoir
Questions les mieux notées	<p>N° 4 : Dès que le joueur effectue une action, le jeu lui répond et réagit</p> <p>N° 11 : Le jeu réagit uniquement lorsque le joueur a effectué une action</p> <p>N° 14 : Dans le jeu, il existe différents niveaux de difficulté en fonction de l'expérience du joueur</p>	<p>N° 11 : Le jeu réagit uniquement lorsque le joueur a effectué une action</p> <p>N° 14 : Dans le jeu, il existe différents niveaux de difficulté en fonction de l'expérience du joueur</p> <p>N° 23 : Le vocabulaire utilisé dans le jeu est simple et clair</p>
Questions les moins bien notées	<p>N° 13 : Dans le jeu, il existe toujours plusieurs manières pour atteindre un objectif ou réaliser une tâche</p> <p>N° 15 : Grâce au jeu, le joueur ne peut pas se tromper ni faire d'erreurs</p> <p>N° 17 : Au cas où le joueur fait une erreur, le jeu lui explique clairement pourquoi il s'est trompé</p> <p>N° 18 : Si le joueur fait une erreur, le jeu lui offre un</p>	<p>N° 13 : Dans le jeu, il existe toujours plusieurs manières pour atteindre un objectif ou réaliser une tâche</p> <p>N° 17 : Au cas où le joueur fait une erreur, le jeu lui explique clairement pourquoi il s'est trompé</p> <p>N° 18 : Si le joueur fait une erreur, le jeu lui offre un message clair et précis pour la corriger</p>

	message clair et précis pour la corriger	
--	--	--

Tableau 1. Critères les mieux et les moins bien notés dans les deux jeux

Globalement, les joueurs estiment que les jeux proposent un bon feedback immédiat, prennent bien en compte l'expérience de l'utilisateur et le critère concernant la signifiante des codes et dénominations est bien respecté. En revanche, les joueurs considèrent que les jeux manquent de flexibilité, dans le sens où il n'existe pas plusieurs manières d'effectuer la même tâche ou atteindre un même objectif. De plus, ils ont émis des critiques en ce qui concerne le critère de gestion des erreurs : pas de protection, pas de correction et pas de message d'erreur.

Au vu de ces résultats, on note la similitude d'analyse avec celle qui est effectuée par l'ergonome avant la passation des tests. Il en ressort une complexité particulière liée aux erreurs, qu'il serait judicieux d'améliorer dans les prochains jeux de la série *Nancy Drew*.

Performances - 6 joueurs sur 10 ont échoué à la tâche demandée dans Mode tandis que 6 joueurs sur 10 ont réussi dans Manoir. Les abandons sont équivalents pour chaque jeu à un joueur près. (*Manoir* : 3 et *Mode* : 2). On ne peut pas expliquer pourquoi on observe trois fois plus d'échecs dans Mode que dans Manoir. Plusieurs explications sont possibles. Soit la tâche demandée était trop complexe par rapport à l'expérience du joueur sur ce jeu (Mode), soit elle était trop simple dans Manoir.

Mouvements oculaires –

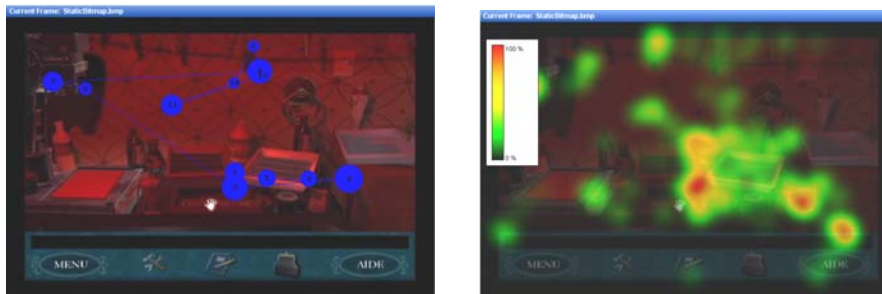


Figure 2. Scanpath (à gauche) et carte de chaleur (à droite)

Les données oculométriques analysées sont le nombre et les durées de fixations. Selon les résultats, on n'observe pas de différence significative des moyennes du nombre de fixations par minute entre les jeux Mode (103,80) et Manoir (119,92) [$t(9)=1,68$; $p>.05$]. Nous avons choisi d'observer le nombre de fixations par minute car on n'observe pas les mêmes durées de jeux entre ceux qui ont joué pendant 20 minutes, ceux qui ont trouvé la solution plus rapidement et ceux qui ont abandonné. Ce moyennage permet de comparer les jeux sur un même niveau. En revanche, on observe une différence significative des durées moyennes de fixation entre Mode (435ms) et Manoir (385ms) [$t(9)=3,50$; $p<.01$].

Lien entre les mouvements oculaires et questionnaire d'utilisabilité -

Dans « *Mode* », on observe une seule corrélation entre les résultats au questionnaire et les mouvements oculaires. Tandis que dans « *Manoir* » on peut en observer plusieurs. Globalement, on observe des liens entre les mouvements oculaires et les critères "guidage", "charge de travail", "homogénéité" et "signifiante des codes et dénominations".

Dans « *Mode* », on observe une corrélation négative (figure 3, $r=-0,46$) entre les résultats au critère "charge de travail" et le nombre de fixations. Autrement dit, plus la charge de travail a été mal notée par les joueurs, donc lorsqu'elle est trop importante au niveau de l'interface, plus le nombre de fixations oculaires augmente.

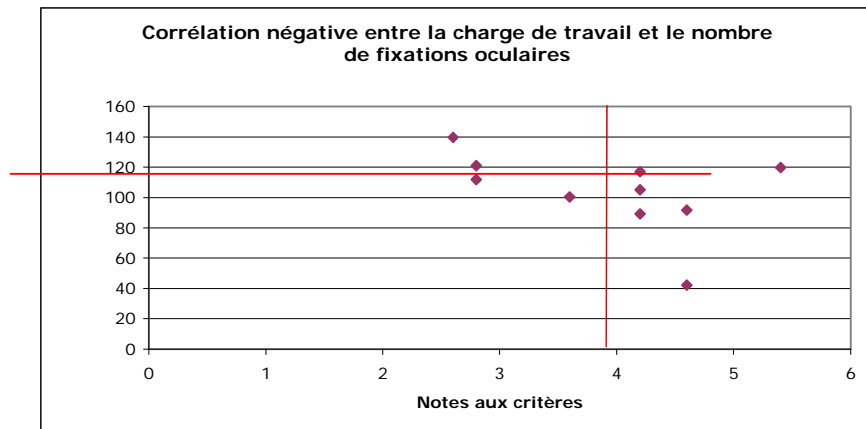


Figure 3

Dans « Manoir », on trouve également cette corrélation négative entre la charge de travail et le nombre de fixations oculaires (figure 4, $r=-0,49$). Mais on trouve aussi une corrélation positive entre la notation au critère "charge de travail" et la durée de fixation (figure 5, $r=0,66$). C'est-à-dire que plus la charge de travail a été jugée importante par les joueurs, plus la durée de fixation oculaire augmente.

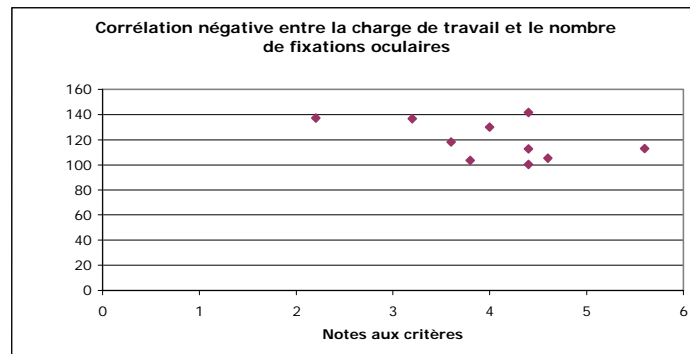


Figure 4

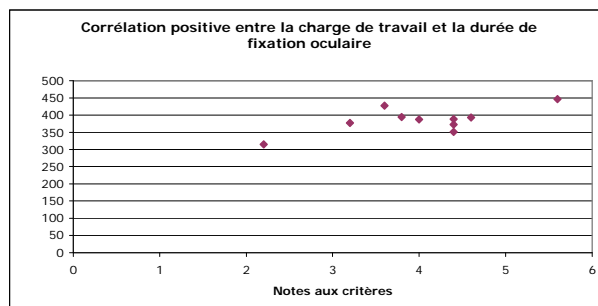


Figure 5

D'autre part, dans « Manoir », on observe une corrélation positive entre les notations au critère de "guidage" et le nombre de fixations (figure 6, $r=0,67$). Autrement dit, plus le guidage a été bien noté, plus le nombre de fixations oculaires augmente. Mais on observe également une corrélation négative entre la notation au critère

"guidage" et la durée de fixation (figure 7, $r=-0,54$). C'est-à-dire que plus le guidage est jugé positivement par les joueurs, moins la durée de fixation oculaire est importante.

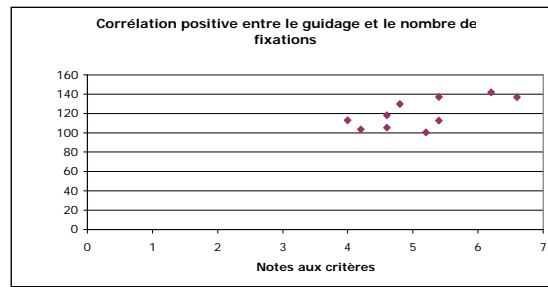


Figure 6

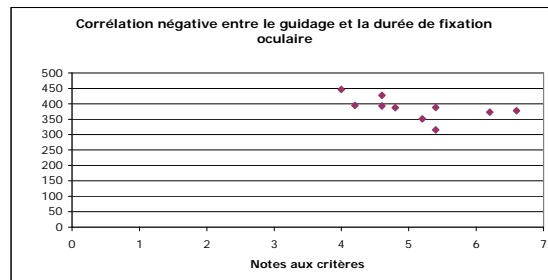


Figure 7

Dans « *Manoir* », on observe également une corrélation positive (figure 8, $r=0,63$) entre la notation au critère "homogénéité" et le nombre de fixations oculaires : plus le jeu est jugé homogène et cohérent, plus le nombre de fixations augmente.

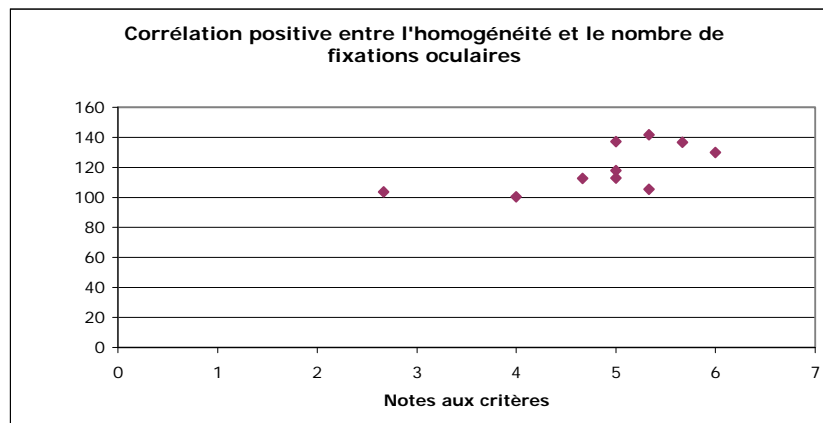


Figure 8

Enfin, toujours dans « *Manoir* », on peut observer une corrélation positive (figure 9, $R=0,62$) entre la notation au critère "signifiante des codes et dénominations" et le nombre de fixations oculaires. C'est-à-dire que plus la signifiante des codes est jugée simple et facile à comprendre, plus le nombre de fixations oculaires augmente.

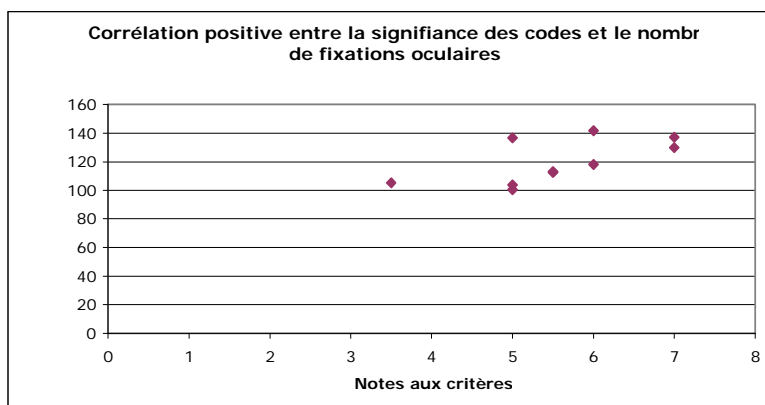


Figure 9

En résumé, on remarque une similitude entre les deux jeux en ce qui concerne la corrélation négative entre la charge de travail et le nombre de fixations oculaires. On peut dire que sur ce critère, les deux jeux ont été évalués de la même manière.

En conclusion – Cette étude qui visait à corréler les notes obtenues par les jeux aux critères d'utilisabilité et les mouvements oculaires, a montré qu'il est possible de mettre en évidence une relation entre ces deux types variables. Ces résultats suggèrent donc de continuer à creuser cette relation pour déterminer plus exactement l'interprétation de ces corrélations en terme de recommandations ergonomique, ainsi que de les étudier pour certaines parties du jeu joué et non pour la totalité comme cela a été le cas ici. En effet il est probable que certains résultats ou absence de résultats proviennent de notre analyse qui est globale, que ce soit au niveau des mouvements oculaires qu'au niveau du questionnaire. Grâce à l'acquisition d'un nouveau logiciel il va être maintenant possible d'analyser uniquement certaines séquences du jeu et de relier chacune de ces séquences aux critères d'utilisabilité.

ANNEXE2. ADAPTATION DES CRITERES DE BASTIEN ET SCAPIN AUX JEUX VIDEO.

Critères ergonomiques	Application aux jeux vidéo	Explication
1. Guidage	Oui	<p>Le guidage est important dans un jeu vidéo car il permet au joueur de savoir où il se situe dans le jeu, de connaître ce qu'il peut faire ou non. Par exemple, un tutorial de départ permet la prise en main du jeu.</p> <p>Par exemple, afficher le score, un chrono, les points de vie restants, ceux des ennemis peut contribuer à guider le joueur tout au long du jeu.</p> <p>Par exemple, le guidage lui permet de savoir où il en est dans la tâche qu'il doit accomplir, en affichant notamment le nombre de monstres qu'il lui reste à tuer, ou combien de tours il lui reste pour achever sa course de voitures.</p> <p>Le guidage permet également au joueur d'obtenir des informations supplémentaires, éventuellement à sa demande. On peut prendre l'exemple de Final Fantasy où certains personnages non joueurs (PNJ) peuvent servir d'aide à la continuation du jeu. Egalement, dans le jeu Midnight Club, en appuyant sur la touche "select" le joueur peut avoir accès au parcours total qu'il doit suivre pour réaliser sa course de voitures.</p>

1.1. Incitation	Oui	L'incitation est indispensable dans un jeu vidéo dans le sens où il s'agit de mettre en place des outils d'aide, au cas où le joueur serait bloqué. D'autre part, des chronos, ou des plans permettent au joueur de savoir où il en est exactement dans le déroulement du jeu.
1.2. Groupement / Distinction entre items	Oui	Dans un jeu vidéo, il est nécessaire de bien organiser les informations. Par exemple, si dans un jeu de course de voitures le chrono est situé en bas à gauche, il ne doit pas changer de place lors de la course suivante. De la même manière, si le score est situé en haut à gauche dans <i>Crash Bandicoot</i> , si le joueur augmente d'un niveau le score doit rester au même endroit.
1.3. Feedback immédiat	Oui	Dans un jeu vidéo, il est indispensable que le système propose un feedback immédiat au joueur dès que celui-ci effectue une action. Par exemple, s'il s'agit d'un jeu où des dialogues sont présents (<i>Nancy Drew, Les Experts, Sherlock Holmes, Final Fantasy...</i>), lorsque le joueur pose une question à un personnage non joueur, ce dernier doit lui répondre immédiatement. D'autre part, en début de jeu, si le joueur sélectionne "charger une partie", une fois le chargement terminé il est important de proposer un feedback en précisant que sa partie a effectivement été chargée.
1.4. Lisibilité	Oui	Dans les menus, les dialogues, les bulles d'information, ou toute autre information dans le jeu, il est indispensable de tenir compte de la taille des caractères, leur typographie, leur couleur, de l'espacement des lettres, des mots, des phrases et des lignes. D'autre part, il faut tenir compte des caractéristiques cognitives et perceptives des joueurs. Il faut surtout prendre en compte le fait que certains joueurs peuvent passer plusieurs heures devant l'écran de télévision. C'est d'autant plus important qu'un écran de télévision provoque une plus grande fatigue visuelle que l'écran d'un ordinateur.
2. Charge de travail	Oui et non	Ce critère concerne la charge de travail au niveau perceptif et cognitif.
2.1. Brièveté	Oui et non	Ce critère peut s'appliquer mais cela dépend des jeux vidéo. En effet, parfois le but du jeu peut être de retenir un maximum d'éléments ou bien de chercher des indices dans une interface surchargée d'éléments graphiques ou d'informations. Cependant, si la charge de travail est vraiment trop importante, le joueur risque de se sentir frustré s'il ne réussit pas les tâches qui lui sont demandées. Il faut alors trouver un juste milieu, en fonction de chaque jeu.
2.2. Densité informationnelle	Oui et non	Afficher uniquement les informations nécessaires est important notamment dans le menu principal du jeu ou les sous-menus, lorsqu'il s'agit pour le joueur de gérer des paramètres ou des options. Toutefois, dans le déroulement du jeu, présenter uniquement des informations pertinentes risque de faciliter énormément les

		tâches que les joueurs doivent accomplir. Ainsi, il ne faut pas que le jeu devienne trop simple.
3. Contrôle explicite	Oui et non	Ce critère correspond à la relation entre le fonctionnement du jeu et les actions des joueurs.
3.1. Actions explicites	Oui et non	Dans un jeu vidéo, le joueur ne doit pas tout contrôler. Par exemple, dans une course de voitures, il ne doit pas pouvoir contrôler ses adversaires, dans un jeu de tennis, il ne doit pas pouvoir contrôler le joueur adverse. Ou bien dans Tomb Raider, il n'y a aucun intérêt à ce que le joueur contrôle les ennemis qu'il doit abattre. Donc, le système ne doit pas exécuter uniquement les actions des joueurs. Au contraire, il doit provoquer des surprises, susciter l'intérêt grâce à des imprévus. Mais, lorsque le joueur fait une action, le système doit exécuter ce que le joueur demande au moment où il le demande. Par exemple dans Crash Bandicoot, si le joueur utilise la fonction "sauter", les personnages du jeu doit effectivement sauter.
3.2. Contrôle utilisateur	Oui	Le joueur doit avoir la possibilité de mettre pause quand il le souhaite et reprendre sa partie quand il veut. Il doit pouvoir charger une partie à n'importe quel moment et avoir accès aux options (sonores, visuelles) n'importe quand. En revanche, en ce qui concerne la sauvegarde, il existe trois types de cas : - Soit le joueur peut sauvegarder n'importe où et à n'importe quel moment du jeu (ex : Tomb Raider). - Soit il n'a pas besoin de sauvegarder car il existe un système de sauvegarde automatique (ex : Midnight Club Dub Edition). - Soit il doit trouver ce que l'on appelle des points de sauvegarde ou des checkpoints afin de pouvoir enregistrer sa partie (ex : Final Fantasy).
4. Adaptabilité	Oui	Ce critère concerne la capacité du système à réagir selon le contexte et les besoins ou préférences des utilisateurs.
4.1. Flexibilité	Oui	Il est important de réduire le nombre d'actions nécessaires afin que le joueur accomplisse une action (par exemple sauter, courir, taper...). Ce type d'action doit être réalisé avec un seul clic sur un bouton de la manette.
4.2. Prise en compte de l'expérience utilisateur	Oui	Dans un jeu, il est nécessaire de proposer plusieurs manières d'atteindre un but ou d'accomplir une tâche. Il s'agit de s'adapter à la fois aux novices et aux experts. Ainsi, il convient de proposer des niveaux de difficulté différents (facile, normal, difficile).

5. Gestion des erreurs	Oui	Ce critère correspond aux moyens mis en oeuvre afin d'éviter ou de réduire les erreurs, ou bien de les corriger.
5.1. Protection contre les erreurs	Oui	Par exemple, si le joueur doit saisir un nom ou un pseudo pour un personnage (Final Fantasy, Dark Cloud), on doit lui proposer un moyen d'annuler ou de corriger sa saisie. D'autre part, s'il appuie malencontreusement sur un mauvais bouton, cela ne doit pas entraîner de conséquences irréversibles.
5.2. Qualité des messages d'erreur	Oui	Si le joueur appuie sur un mauvais bouton, il convient de lui indiquer clairement sur quel bouton il doit appuyer. Les messages doivent être clairs, précis, concis et courts.
5.3. Correction des erreurs	Oui	Dans un jeu, il faut permettre au joueur de corriger son erreur, en lui expliquant pourquoi il s'est trompé et en lui proposant la bonne procédure à suivre afin qu'il ne réitère pas son erreur.
6. Homogénéité / Cohérence	Oui	Dans un jeu, les commandes, les informations présentes en permanence, la localisation des items (points de vie, chrono, score...) ne doivent pas changer durant le jeu. L'homogénéité de ces éléments permet au joueur de prendre en main le jeu plus facilement, cela diminue le temps de recherche des informations et permet un apprentissage plus rapide du jeu.
7. Signifiante des codes et dénominations	Oui	Dans les jeux, les icônes doivent être compréhensibles. On trouve parfois de nombreuses icônes (Les Experts, Sherlock Holmes, L'île mystérieuse...), celles-ci nécessitent parfois un apprentissage. Une des solutions est de proposer une aide accessible à n'importe quel moment du jeu. Par exemple, proposer une description de l'item, de l'outil ou de l'icône afin de connaître son utilité.
8. Compatibilité	Oui	Au sens large, un jeu vidéo doit pouvoir être joué dans n'importe quel pays et traduit dans la plupart des langues (anglais, espagnol, français, italien, allemand...). De nombreux jeux vidéo proposent un choix de la langue en tout début de jeu (Baldur's gate, Harry Potter...). Le langage utilisé dans le jeu doit être compréhensible par la plupart des personnes, quelque soit leur âge, catégorie socio-professionnelle, connaissance en jeu vidéo... Mais parfois, un apprentissage préliminaire est essentiel en début de partie (par exemple les tutoriaux).

Annexe 2 - Analyse heuristiques des deux jeux vidéo testés.

Critères ergonomiques	Nancy Drew, la malédiction du manoir de Blackmoor	Nancy Drew, danger au coeur de la mode
Guidage	<p>Oui : le calepin avec tous les objectifs accomplis et ceux qu'il reste à faire ainsi que le téléphone portable qui permet d'avoir accès à Internet. Le joueur sait où il en est et ce qui lui reste à faire.</p>	<p>Oui : le calepin avec tous les objectifs accomplis et ceux qu'il reste à faire ainsi que le téléphone portable qui permet d'avoir accès à Internet. Le joueur sait où il en est et ce qui lui reste à faire.</p>
Incitation	<p>Manque d'incitation.</p> <ul style="list-style-type: none"> - Parfois, le joueur est bloqué car les items qu'il doit collecter, les indices qu'il doit trouver puis mettre bout à bout afin de trouver la solution ne sont pas très explicites. Par exemple, les indices à collecter pour ouvrir la forge souterraine sont éparpillés dans différents endroits du manoir (hall, galerie souterraine, chambre de Jane...). De plus, pour les acquérir il faut résoudre des énigmes assez complexes. Et lorsque le joueur s'aperçoit que l'élément qu'il vient de gagner est une galette ronde (qui est en réalité une lune !), il ne comprend ni l'utilité de cet item, ni ce que c'est exactement, et rien dans le jeu ne lui permet de comprendre. - Il n'y a pas d'outils d'aide (hormis le calepin qui récapitule les objectifs). - Il n'y a pas de plans, de cartes pour savoir où il est dans le manoir. - En dernier recours, le joueur peut aller consulter 	<ul style="list-style-type: none"> - Il n'y a pas d'outils d'aide (hormis le calepin qui récapitule les objectifs). - Il n'y a pas de plans, de cartes pour savoir où il est dans le manoir. - En dernier recours, le joueur peut aller consulter la notice sur Internet.

	la solution sur Internet.	
Groupement / Distinction entre items	Oui : le menu en bas est composé d'éléments similaires.	Oui : le menu en bas est composé d'éléments similaires.
Feedback immédiat	Oui : Réponse immédiate des personnages non joueurs lors des dialogues. Oui : dès que l'histoire évolue, les nouveaux éléments et nouveaux objectifs sont inscrits dans le carnet. Le joueur le repère car dans le menu, l'icône du carnet clignote en vert avec un petit point d'exclamation. Oui : lorsqu'on charge, sauvegarde ou quitte une partie, un message de demande de confirmation s'affiche.	Oui : Réponse immédiate des personnages non joueurs lors des dialogues.
Lisibilité	Oui : lors des dialogues, les critères de lisibilité sont bien respectés. Mais dans le carnet, une typographie de type manuscrit est utilisée mais elle ne nuit pas à la lisibilité.	Oui : lors des dialogues, les critères de lisibilité sont bien respectés. Mais dans le carnet, une typographie de type manuscrit est utilisée mais elle ne nuit pas à la lisibilité.
Charge de travail		
Brièveté	Non : trop de lecture (dialogues, livres, carnet...) Oui : les informations importantes concernant l'histoire et les tâches à accomplir sont rappelées dans le carnet. Non : parfois le nombre d'étapes pour réaliser une tâche est excessif (par exemple pour ouvrir la forge, il faut dérouler le jeu quasiment jusqu'à la	Non : parfois le nombre d'étapes pour réaliser une tâche est excessif (par exemple pour ouvrir la forge, il faut dérouler le jeu quasiment jusqu'à la fin pour passer cette étape). Non : trop de lecture (dialogues, livres, carnet...) Oui : les informations importantes concernant l'histoire et les tâches à accomplir sont rappelées dans le carnet.

	fin pour passer cette étape).	
Densité informationnelle	Oui : le menu ne contient que trois items, ce critère est respecté. Non : certains objets dans le jeu, notamment des livres à lire, propose énormément d'informations inutiles. Elles ne servent qu'à induire le joueur en erreur et à surcharger sa mémoire de travail.	Oui : le menu ne contient que trois items, ce critère est respecté.
Contrôle explicite		
Actions explicites	Oui : critère respecté	Oui : critère respecté
Contrôle utilisateur	Non : le joueur ne peut pas passer les dialogues ni les cinématiques. Il est obligé de tout lire, il n'a pas le contrôle. Oui : le joueur peut mettre pause, sauvegarder, charger ou quitter à tout moment du jeu.	Non : le joueur ne peut pas passer les dialogues ni les cinématiques. Il est obligé de tout lire, il n'a pas le contrôle. Oui : le joueur peut mettre pause, sauvegarder, charger ou quitter à tout moment du jeu.
Adaptabilité		
Flexibilité	Oui : le système de "point & clic" permet de s'adapter à des joueurs n'ayant pas l'habitude de se servir d'un ordinateur. Oui : même si en fin de compte, toutes les étapes doivent être réalisées pour terminer le jeu, les joueurs peuvent les réaliser dans un ordre différent. Ce jeu n'est pas linéaire. Ils ont donc la possibilité d'emprunter plusieurs chemins pour arriver à la fin.	Oui : le système de "point & clic" permet de s'adapter à des joueurs n'ayant pas l'habitude de se servir d'un ordinateur.
Prise en compte de l'expérience utilisateur	Oui : au début, le joueur a le choix de jouer en mode "détective débutant" ou "détective confirmé".	Oui : au début, le joueur a le choix de jouer en mode "détective débutant" ou "détective confirmé". Mais ce

	Mais ce choix est irrémédiable, il n'est pas possible de changer cette option en cours de partie.	choix est irrémédiable, il n'est pas possible de changer cette option en cours de partie.
Gestion des erreurs		
Protection contre les erreurs	Non : les petites énigmes présentes dans le jeu ne sont pas toujours clairement expliquées. Le joueur peut se sentir perdu, à ne pas savoir quoi ni comment faire (jeu dans le hall, boîte des animaux dans la chambre, nourrir le perroquet...).	
Qualité des messages d'erreur	Non : il n'y a jamais de message d'erreur.	Non : il n'y a jamais de message d'erreur.
Correction des erreurs	Non : il n'y a jamais de message d'erreur.	Non : il n'y a jamais de message d'erreur.
Homogénéité	Oui : critère respecté.	Oui : critère respecté.
Signifiante des codes et dénominations	Oui : la signification des icônes est suffisamment explicite et il y en a peu.	Oui : la signification des icônes est suffisamment explicite et il y en a peu.
Compatibilité	Oui : le langage utilisé est simple et clair, accessible à tous. Le fait de jouer avec la souris simplifie la prise en main du jeu.	Oui : le langage utilisé est simple et clair, accessible à tous. Le fait de jouer avec la souris simplifie la prise en main du jeu.

Annexe 3 - Questionnaire d'utilisabilité rempli par les utilisateurs
après chaque jeu

Critères ergonomiques	Description du critère, que l'utilisateur doit noter de 1 (Pas du tout d'accord) à 7 (Tout à fait d'accord)
1. Guidage	Le guidage concerne les moyens mis en oeuvre pour conseiller, orienter, informer et conduire le joueur.
1.1. Incitation	Le jeu incite le joueur à réaliser des actions, des tâches ou des missions particulières.
1.2. Groupement / distinction entre items	Dans le jeu, les informations de même type sont regroupées.
	Tandis que celles qui sont différentes se distinguent facilement.
1.3. Feedback immédiat	Dès que le joueur effectue une action, le jeu lui répond et réagit.
1.4. Lisibilité	La lecture est facile dans le jeu (au niveau des polices, couleurs, styles de caractères...).
2. Charge de travail	Ce critère concerne la charge de travail imposée par le jeu, à la fois au niveau de la perception et de la mémoire.
2.1. Brièveté	Le travail de lecture (en ce qui concerne les textes du jeu) est rapide.
	Le travail de lecture (en ce qui concerne les textes du jeu) est facile.
	Le nombre d'étapes pour réaliser une tâche est réduit.
	Les étapes pour atteindre l'objectif qui vous est demandé sont faciles.
2.2. Densité informationnelle	On trouve uniquement des informations utiles et pertinentes dans le jeu (dans les dialogues, l'histoire...).
3. Contrôle explicite	Ce critère correspond à la relation qui existe entre le fonctionnement du jeu et les actions des joueurs.
3.1. Actions explicites	Le jeu réagit uniquement lorsque le joueur a effectué une action.
3.2. Contrôle utilisateur	Le joueur garde toujours la main sur le jeu, il contrôle tout.
4. Adaptabilité	Ce critère correspond à la capacité du jeu à réagir en fonction des besoins et préférences des joueurs.
4.1. Flexibilité	Dans le jeu, il existe toujours plusieurs manières pour atteindre un objectif ou réaliser une tâche.
4.2. Prise en compte de l'expérience utilisateur	Dans le jeu, il existe différents niveaux de difficulté en fonction de l'expérience du joueur
5. Gestion des erreurs	Ce critère correspond au fait que le jeu permet d'éviter de faire des erreurs ou de les corriger.
5.1. Protection contre les erreurs	Grâce au jeu, le joueur ne peut pas se tromper ni faire d'erreurs.
	Le joueur sait toujours ce qu'il peut faire ou non.
5.2. Qualité des	Au cas où le joueur fait une erreur, le jeu lui explique clairement pourquoi il

messages d'erreur	s'est trompé.
5.3. Correction des erreurs	Si le joueur fait une erreur, le jeu lui offre un message clair et précis pour la corriger.
6. Homogénéité	Dans le jeu, le graphisme est homogène.
	Dans le jeu, toutes les fonctionnalités offertes sont logiques, pertinentes et cohérentes.
	Dans le jeu, tous les éléments qui se ressemblent sont toujours placés au même endroit.
7. Signifiante des codes et dénominations	Les icônes sont facilement compréhensibles.
	Le vocabulaire utilisé dans le jeu est simple et clair.
8. Compatibilité	Le jeu s'adapte à n'importe qui

Annexe 3 Questionnaires d'expertise papier

QUEL JOUEUR ETES-VOUS ?

Age _____

Sexe _____

Profession _____

1. Etes-vous à l'aise avec les technologies numériques ?

- Oui
- Non

2. Jouez-vous aux jeux vidéo ?

- Oui
- Non

3. A quelle fréquence jouez-vous ?

- Plusieurs heures par jour
- Quelques fois par semaine
- Quelques fois par mois
- Moins d'une fois par mois

4. Sur quel type de plate-forme jouez-vous ?

- PC
- Console
- Autre (précisez : _____)

5. A quel(s) type(s) de jeu(x) jouez-vous ?


- Jeux de tir à la première personne (ex. : Counter Strike)
- Jeux de stratégie en temps réel (ex. : Age of Empire)
- Jeux de simulation (ex. : Flight Simulator)
- Jeux de rôle (ex. : Oblivion)
- Jeux de plate-forme (ex. : Super Mario Bros)
- Jeux de combat (ex. : Street Fighter)
- Autre (précisez : _____)

6. Avez-vous déjà joué à ces jeux PC (indiquez une estimation du nombre d'heures) ?

- Warhammer 40000 (ou extension) : _____ heure(s)
- Perimeter : _____ heure(s)

Annexe 4. Evaluation ergonomique des RTS

Dark Crusade



- Affichage des ressources groupés et explicite (1+)
- Pictogrammes de ressources peu explicite (3-)
- Minicarte avec positionnement du joueur explicite (3+)
- Orientation de la vision du joueur représentée sur la mini carte (3+)
- positionnement par couleur pour chaque joueur (7+)
- Menus secondaires de gestion de la partie discrets (2+)
- principe de ressources relativement simple (1+)
- information de l'aide hiérarchisée (2+)
- textes concis (2+)
- aide graduée (4+)
- graphismes assez peu gourmands (8+)
- unique
- Un bouton de retour arrière unique et permanent (3+)
- Les éléments non accessibles sont présents mais grisés (3+)
- Commandes en deux niveaux (2-)
- Cohérence dans l'emplacement des commandes (6+)
- Pictogrammes explicites (imagés et taille importante) (7+)
- Visualisation de l'avancement de la construction des unités (3+)
- visualisation de l'états des unités (3+)
- Commande de gestions des unités (équipes, armes, etc.)
- Raccourcis d'interface permettant de sélectionner un type d'unité (pour les experts) (2+,4+)
- Possibilité de masquer l'interface (pour les experts) (2+,4+)
- icône de gestion des paramètres réseau peu saillante (7-,2+)

Perimeter



- Menus secondaire de gestion du jeu peu saillant (2+)
- La gestion de l'énergie est complexe (générateurs, récolte par terraformation, etc.) et peu explicite (beaucoup de tatonnement) (1-,4-)
- Aide contextuelle dense (2-)
- graphismes gourmands (8-)
- Barre horizontale de commande hétérogène (des boutons de même forme n'ont pas les mêmes fonctions) (6-)
- Menu de construction 1/2 avec deux niveaux de cliques et trois menus différents (2-)
- onglets de constructions peu explicites (7-)
- Aucune possibilité de modification de l'interface (4-)
- visibilité du mécanisme « terraformation » (important dans le jeu) inexistante (1-)
- Icones de terraformation peu significatives (7-)
- Gestion de la base contre intuitive (3-)
- Mini carte avec localisation du joueur peu saillante (3-)
- La distinction entre les joueurs est difficile à établir (mêmes unités) (7-)
- visualisation des ressources difficile à comprendre, en deux parties peu saillantes (3-, 2-)
- boutons de commandes multifonctions (bouclier à coté d'upgrade bâtiment) (6-)
- pictogrammes peu explicites (3-)
- Menu de construction 2/2 avec une interaction complexe entre les possibilités de construction (4-)
- niveau de puissance des unités réparti en lignes et en colonnes (1-)
- troisième groupements d'unités distinctes peu explicite (7-)
- gestion des unités construites alambiquée (1-)
- chaque groupe d'unité nécessite la construction d'un bâtiment (information difficile à obtenir) (2-, 1-)

Annexe 5. Les oculomètres Tobii & SMI

Oculomètre Tobii 1750



IView X HED (SMI)



Annexe 6. Consignes des passations RTS

Consigne de Dark Crusade

Vous allez jouer à un jeu vidéo de type « Stratégie en Temps Réel ».

Dans ce type de jeu, le joueur incarne le commandant d'une base armée devant conquérir un territoire où se trouvent d'autres joueurs adverses.

Dans le jeu « Dark Crusade », le principe est de développer sa base le plus rapidement possible en construisant des bâtiments et en recherchant des technologies. Ces développements nécessitent des ressources que le joueur pourra accroître en capturant des points stratégiques. Un point stratégique se capture en envoyant des soldats y déposer un drapeau.

Lorsque sa base sera suffisamment évoluée, le joueur aura la possibilité de former des soldats et de construire des véhicules assez puissants pour prendre l'avantage sur ses adversaires.

Vous allez maintenant prendre le commandement des opérations ; pour mener à bien votre mission vous devrez faire évoluer votre base suffisamment pour pouvoir produire des véhicules blindés lourds (cf. figure 1).



Figure 1 : Véhicules blindés lourds

Ces véhicules vous seront indispensables pour installer votre suprématie sur le territoire.

Consigne de perimeter

Vous allez jouer à un jeu vidéo de type « Stratégie en Temps Réel ».

Dans ce type de jeu, le joueur incarne le commandant d'une base armée devant conquérir un territoire où se trouvent d'autres joueurs adverses.

Dans le jeu « Perimeter », le principe est de développer sa base le plus rapidement possible en construisant des bâtiments et en recherchant des technologies. Ces développements nécessitent des ressources que le joueur pourra accroître en étendant son territoire par terraformation. La terraformation consiste à étendre la zone de sa base à l'aide d'outils pour rendre le terrain exploitable.

Lorsque sa base sera suffisamment évoluée, le joueur aura la possibilité de former des soldats et de construire des véhicules assez puissants pour prendre l'avantage sur ses adversaires.

Vous allez maintenant prendre le commandement des opérations ; pour mener à bien votre mission

vous devrez faire évoluer votre base suffisamment pour pouvoir produire des véhicules avancés de deuxième catégorie (cf. figure 1).



Figure 1 : Véhicules de 2^{ème} catégorie

Ces véhicules vous seront indispensables pour installer votre suprématie sur le territoire.

Annexe 7. Questions posées en cours de jeu RTS

Question 1 : Localiser le plus rapidement possible votre jauge de ressources.

Question 2 : Localiser le plus rapidement possible votre position sur la carte.

Question 3 : Localiser le plus rapidement possible le menu de construction de bâtiments.

Annexe 8. Questionnaire d'évaluation subjective

Jeu :

Participant : S...



Nous vous proposons maintenant de répondre à un questionnaire visant à décrire votre appréciation subjective du jeu à travers différents critères.

Pour chaque critère, il vous sera demandé de répondre sur une échelle en sept points allant de « je ne suis pas du tout d'accord » (1) à « je suis tout à fait d'accord » (7), vous devrez entourer le chiffre correspondant à votre choix, tel que vous pouvez le voir dans l'exemple suivant :

Exemple : J'ai bien compris le fonctionnement de ce questionnaire :

1 2 3 4 5 6 7

Pas du tout d'accord Tout à fait d'accord

→ Ce qui veut dire : « oui, j'ai tout à fait bien compris le fonctionnement du questionnaire ».

Evaluation du jeu (Vous pouvez revenir sur l'écran de jeu si besoin) :

1

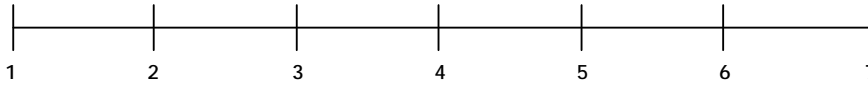
Vous n'avez pas éprouvé de difficulté pour connaître les différentes actions à entreprendre dans la partie.

1 2 3 4 5 6 7

Pas du tout d'accord Tout à fait d'accord

2

Vous avez trouvé que les commandes du jeu étaient plutôt bien groupées : celles appartenant à la même catégorie ensemble, les autres séparées.

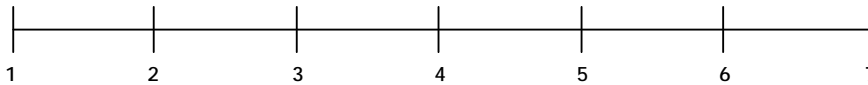


Pas du tout
d'accord

Tout à fait
d'accord

3

Vous avez toujours su si le jeu avait pris en compte l'action que vous veniez d'entreprendre.

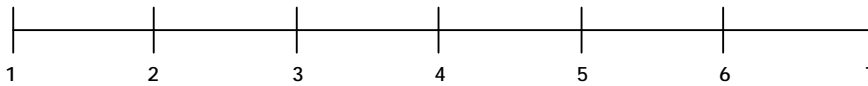


Pas du tout
d'accord

Tout à fait
d'accord

4

Vous n'avez pas éprouvé de difficulté à déchiffrer les textes présentés dans le jeu.



Pas du tout
d'accord

Tout à fait
d'accord

5

Dans le jeu, le nombre d'étapes pour réaliser une tâche est généralement peu important.



Pas du tout
d'accord

Tout à fait
d'accord

6

Les informations données par le jeu vous ont semblé inutiles (dans les dialogues, l'histoire...).

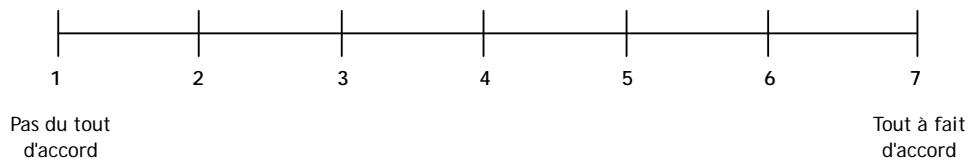


Pas du tout
d'accord

Tout à fait
d'accord

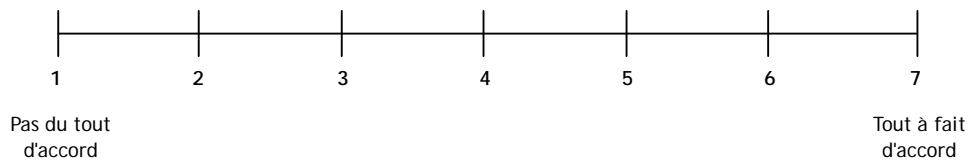
7

Le jeu pouvait réagir alors que vous n'aviez entrepris aucune action.



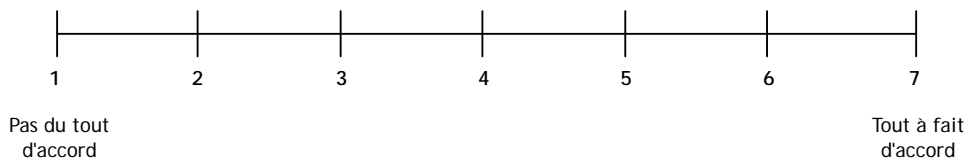
8

Dans le jeu, il vous a semblé qu'il existait toujours plusieurs alternatives pour atteindre un objectif.



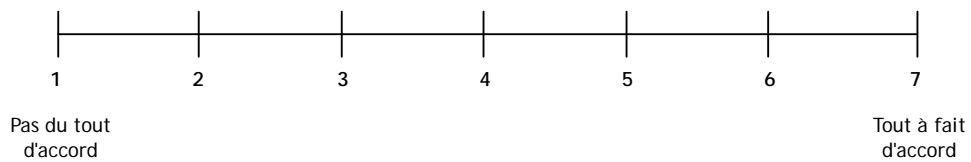
9

La difficulté du jeu n'était pas adaptée à votre niveau de joueur.



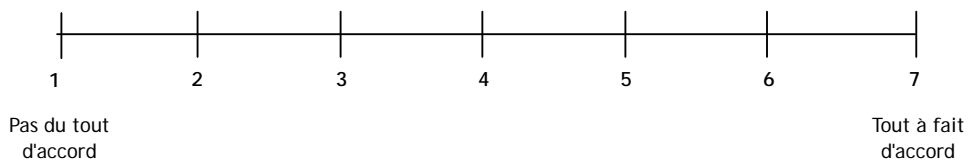
10

Les fonctions du jeu vous ont évité de faire des erreurs (aides contextuelles, présentation des fonctions, etc.) .



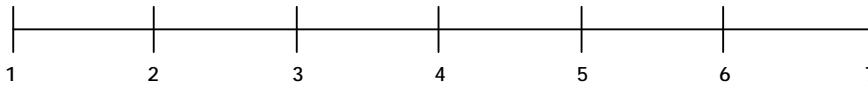
11

Vous connaissiez à tout moment votre marge de manœuvre dans le jeu.



12

Les textes du jeu ne vous ont pas paru trop longs à lire.

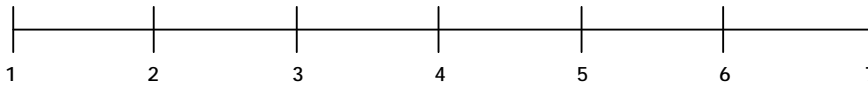


Pas du tout
d'accord

Tout à fait
d'accord

13

Vous avez trouvé l'interface du jeu globalement homogène (dans ses contrôles, couleurs, typographies...).

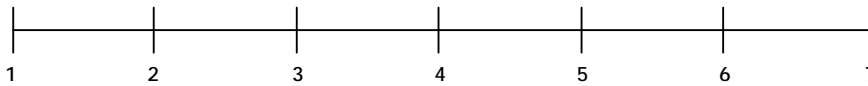


Pas du tout
d'accord

Tout à fait
d'accord

14

Dans le jeu, les fonctionnalités offertes ne vous ont pas semblé logiques.



Pas du tout
d'accord

Tout à fait
d'accord

15

Le vocabulaire du jeu vous a paru compliqué (trop technique, trop soutenu, etc.).



Pas du tout
d'accord

Tout à fait
d'accord

16

Le graphisme de l'interface vous a aidé à comprendre les différentes fonctions.

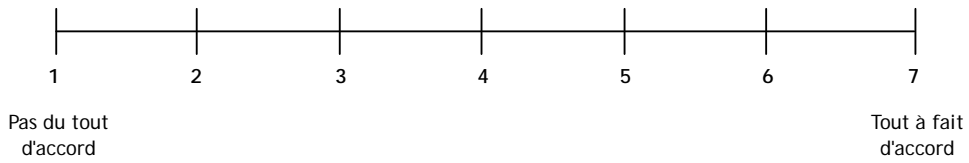


Pas du tout
d'accord

Tout à fait
d'accord

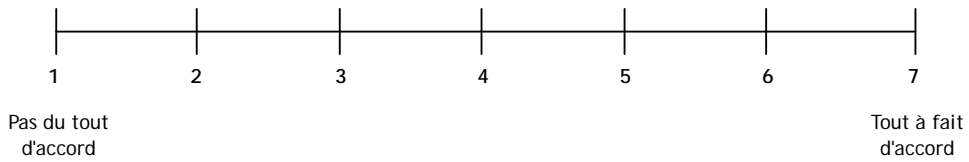
17

Ce jeu s'adresse à un large public (sexe, âge, expérience, etc.)



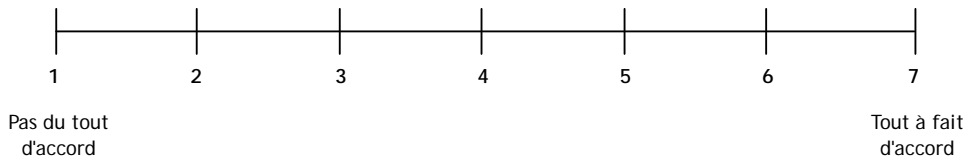
18

L'interface (zone où se situent les commandes) du jeu vous a paru trop présente à l'écran.



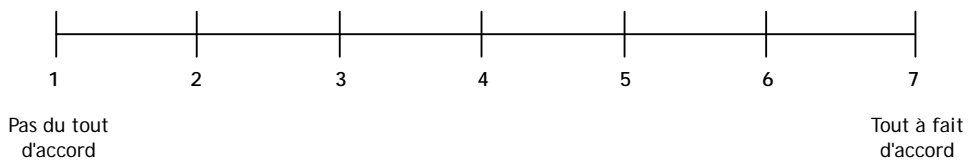
19

Le jeu vous offrait la possibilité de visualiser en permanence les informations concernant l'état de votre partie (ex. : jauges d'énergie, score, niveau, etc.)



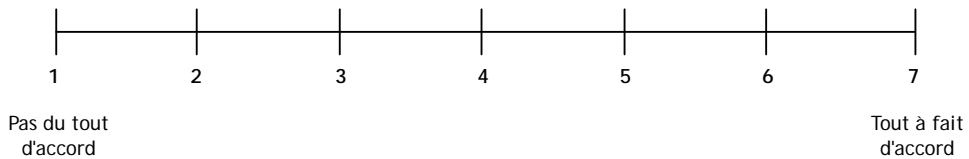
20

Le jeu ne reprend pas suffisamment les standards présents dans les autres jeux de stratégie en temps réel.



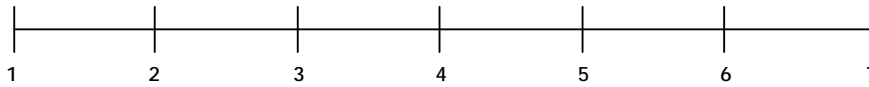
21

Vous vous êtes senti impliqué très tôt dans la partie.



22

Les textes du jeu ne vous ont pas posé de difficulté de compréhension.



Pas du tout
d'accord

Tout à fait
d'accord

23

Vous n'avez jamais perdu le contrôle des commandes du jeu (même momentanément).

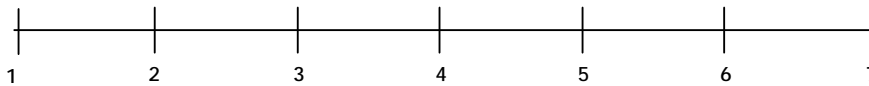


Pas du tout
d'accord

Tout à fait
d'accord

24

Si vous avez fait des erreurs, le jeu vous l'a toujours clairement indiqué.



Pas du tout
d'accord

Tout à fait
d'accord

25

Dans ce jeu, il vous fallait accomplir plusieurs tâches pour atteindre un objectif.



Pas du tout
d'accord

Tout à fait
d'accord

26

Vous avez trouvé les fonctions du jeu difficiles à retenir.

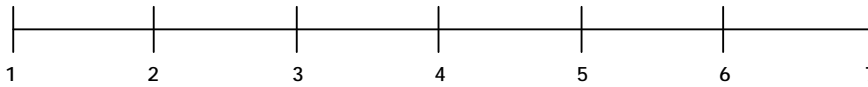


Pas du tout
d'accord

Tout à fait
d'accord

27

Le déroulement du jeu vous a étonné par ses rebondissements.

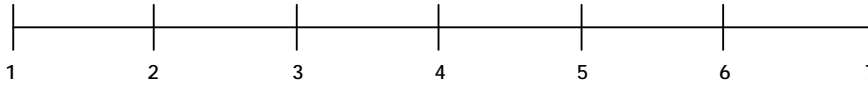


Pas du tout
d'accord

Tout à fait
d'accord

28

Les adversaires vous ont semblé d'un bon niveau.

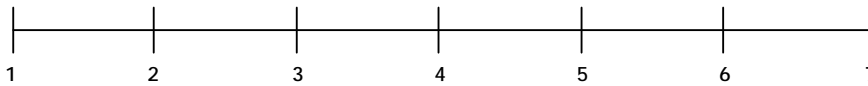


Pas du tout
d'accord

Tout à fait
d'accord

29

Vous avez eu l'impression d'avoir une grande liberté d'action pendant le jeu.



Pas du tout
d'accord

Tout à fait
d'accord

30

En cas d'erreur, le jeu vous a correctement orienté pour résoudre votre erreur.

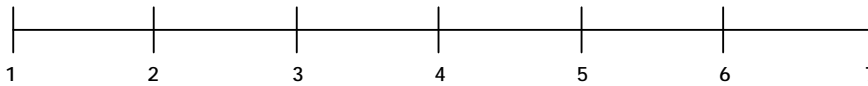


Pas du tout
d'accord

Tout à fait
d'accord

31

Vous avez toujours cru pouvoir gagner la partie.

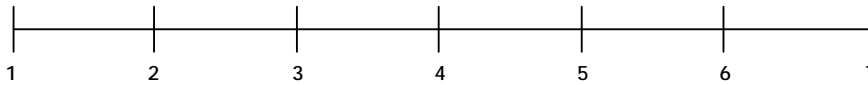


Pas du tout
d'accord

Tout à fait
d'accord

32

Les étapes pour atteindre l'objectif qui vous est demandé vous ont paru faciles.

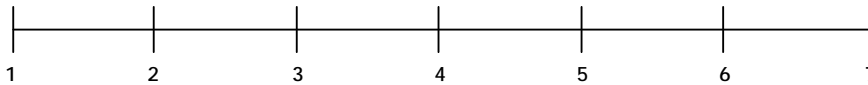


Pas du tout
d'accord

Tout à fait
d'accord

33

Le jeu vous a fourni les indications dont vous aviez besoin pour avancer tout au long de la partie.

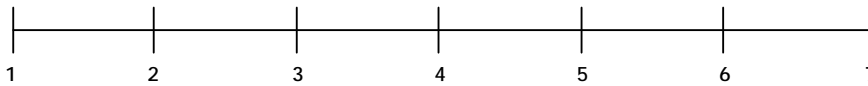


Pas du tout
d'accord

Tout à fait
d'accord

34

Le jeu vous a fait remarquer vos avancées/victoires lorsqu'elles se produisaient.



Pas du tout
d'accord

Tout à fait
d'accord

35

Le rythme du jeu vous a semblé approprié : assez intense mais sans excès.

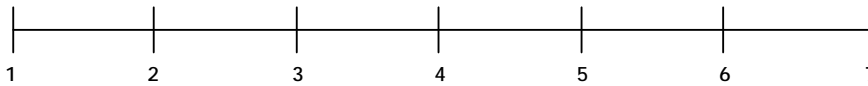


Pas du tout
d'accord

Tout à fait
d'accord

36

Le jeu vous a donné l'impression d'acquérir de nouvelles compétences/connaissances.

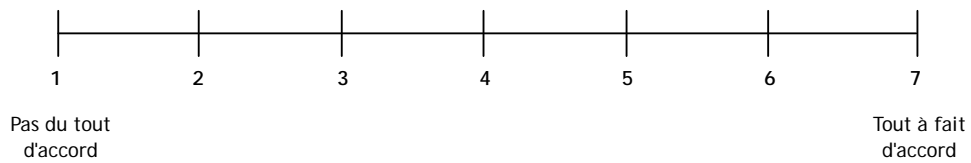


Pas du tout
d'accord

Tout à fait
d'accord

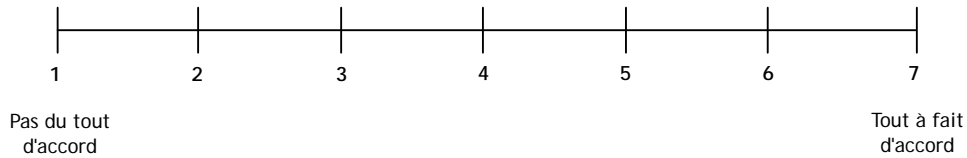
37

Vous pensez que la prochaine partie sera attrayante.



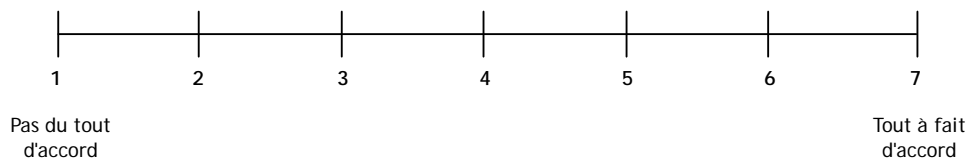
38

Vous avez eu l'impression qu'il n'existait qu'une façon de gagner.



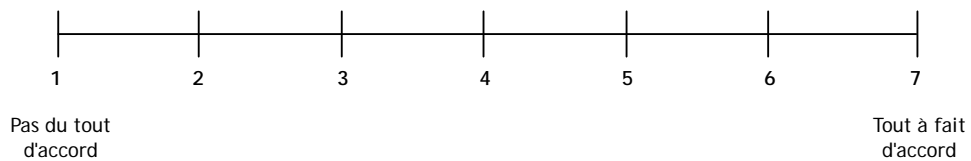
39

Selon vous, les effets visuels et sonores du jeu étaient bons.



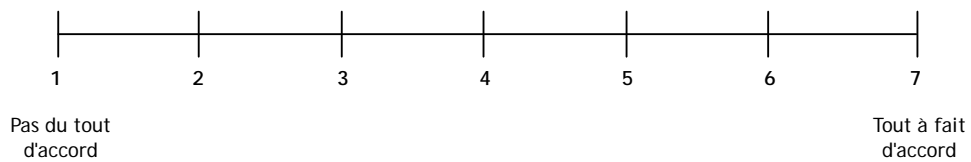
40

L'agencement des contrôles du jeu ne vous a pas semblé naturel.



41

Le jeu vous a donné accès à des possibilités supplémentaires à mesure que vous avanciez.



Annexe 9. Questionnaire d'expertise numérique

questionnaire_expertise_fps1_s14.xls

Age : 23
Sexe : m
Profession : étudiant
Prénom : Damien
Numéro de participant : et14

Donnez-vous une note de 1 à 5 pour situer votre aisé face aux nouvelles technologies
Pas à l'aise - 1 2 3 4 5 - Très à l'aise

Cochez les appareils que vous utilisez (ou avez utilisé) en indiquant une estimation du nombre d'heures/semaine dans la case du dessous

<input checked="" type="checkbox"/> Télévision	<input checked="" type="checkbox"/> Ordinateur	<input checked="" type="checkbox"/> Radio	<input type="checkbox"/> GPS	<input checked="" type="checkbox"/> Wi	<input checked="" type="checkbox"/> Playstation	<input checked="" type="checkbox"/> Xbox	<input checked="" type="checkbox"/> Téléphone Portable / PDA	<input checked="" type="checkbox"/> Consolles portables	Autre: Préciser
7	20	8		2	7	7	20		

Jouez-vous (ou avez-vous joué) aux jeux vidéo ?
 Non Oui

Si oui, indiquez la fréquence

Ordinateur	<input type="radio"/> moins de 1/semaine	<input type="radio"/> Environ 1/semaine	<input checked="" type="radio"/> De 1 à 2/jours	<input type="radio"/> plus de 2/jours
Console fixe	<input type="radio"/> moins de 1/semaine	<input checked="" type="radio"/> Environ 1/semaine	<input type="radio"/> De 1 à 2/jours	<input type="radio"/> plus de 2/jours
Console portable	<input type="radio"/> moins de 1/semaine	<input checked="" type="radio"/> Environ 1/semaine	<input type="radio"/> De 1 à 2/jours	<input type="radio"/> plus de 2/jours

A quel(s) jeu(x) avez-vous déjà joué ?

Donnez un exemple	Vous y avez joué	Non	1	2	3	4	5 Beaucoup
<input checked="" type="checkbox"/> Jeux de tra à la première personne (FPS) - Gears Of War	Vous y avez joué	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
<input checked="" type="checkbox"/> Jeux de stratégie en temps réel (RTS) - Dawn Of War	Vous y avez joué	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input checked="" type="checkbox"/> Jeux de gestion - Civ4	Vous y avez joué	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input checked="" type="checkbox"/> Jeux de course/sport/simulation - Burnout	Vous y avez joué	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input checked="" type="checkbox"/> Jeux de rôle/RPG - Neverwinter Nights	Vous y avez joué	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input checked="" type="checkbox"/> Jeux de plate-forme/aventure	Vous y avez joué	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input checked="" type="checkbox"/> Jeux de combat/battle real - Dead Or Alive	Vous y avez joué	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="checkbox"/> Arcade	Vous y avez joué	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input checked="" type="checkbox"/> Jeux multijoueurs (MMO) - WOW	Vous y avez joué	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Annexe 10. Consigne des passations FPS

Consigne « Halo 3 »

Vous incarnez un soldat se battant pendant le 3^{ème} opus d'une guerre futuriste lors de la première bataille.

Dans cette phase de jeu vous devez accompagner votre coéquipier extra-terrestre dans une mission de conquête de territoire avec passage de points stratégiques. L'action du jeu consiste à avancer en suivant votre équipier (triangle bleu) tout en vous défendant des ennemis qui tentent de vous abattre. Vous disposez pour cela d'armes à feu et de grenades.

Pour vous déplacer, vous devez utiliser le pad gauche. Pour vous orienter vous utiliserez le pad droit. Le tir se fait avec la grande gâchette droite, le jet de grenade se fait avec la grande gâchette gauche. Vous pouvez changer d'arme avec le bouton jaune (Y).

Pendant le jeu, je vous poserai trois courtes questions portant sur le jeu, il faudra y répondre rapidement.

Vous pouvez commencer

Consigne « The history channel : Battle for the Pacific »

Vous incarnez un soldat se battant pendant la seconde guerre mondiale lors de La Bataille du Pacifique.

Dans cette phase de jeu vous devez accompagner vos coéquipiers dans une mission de conquête de territoire avec destruction de pièces d'artillerie. L'action du jeu consiste à avancer en suivant votre équipier (rond rouge) tout en vous défendant des ennemis qui tentent de vous abattre. Vous disposez pour cela d'armes à feu et de grenades.

Pour vous déplacer, vous devez utiliser le pad gauche. Pour vous orienter vous utiliserez le pad droit. Le tir se fait avec la grande gâchette droite, le jet de grenade se fait avec la petite gâchette droite. Vous pouvez changer d'arme avec la petite gâchette gauche.

Pendant le jeu, je vous poserai trois courtes questions portant sur le jeu, il faudra y répondre rapidement.

Vous pouvez commencer

Annexe 11. Questions posées en cours de jeu FPS

Questions à poser en cours de jeu, toutes les 2 mn à partir de la deuxième minute de jeu :

Question 1 : Pouvez-vous m'indiquer le nombre de munitions dont vous disposez ?

Question 2 : Pouvez-vous m'indiquer le nom de l'arme que vous utilisez ?

Question 3 : Pouvez-vous m'indiquer de combien de grenades vous disposez ?

SP4 Méthodes d'évaluation en sémiotique du *gameplay*

Sébastien Genvo, Serge Bourchardon

Objectif

Les présentes méthodes avaient pour objectif général de diagnostiquer les éventuels problèmes de *gameplay* d'une production vidéoludique, à partir d'une analyse de contenus du système de jeu.

Attendus scientifiques initiaux

Pour formuler ces méthodes, il était en premier lieu nécessaire de partir d'une définition du *gameplay*¹⁴. Cette notion, qui ne connaît pas d'équivalence exacte en français, est employée de façon instinctive par les professionnels (concepteurs, journalistes, etc.) et les amateurs de jeux vidéo. Pour les amateurs, le *gameplay* constitue l'un des critères fondamentaux de qualité d'un bon jeu vidéo. Il est alors fréquemment employé pour qualifier ce qui fait qu'un jeu est bon, indépendamment de ses seuls exploits techniques (beauté des graphismes, etc.). Il renvoie plus particulièrement aux modalités d'action du joueur dans l'univers fictionnel. C'est à travers son *gameplay* qu'un jeu vidéo pourra donc inciter l'utilisateur à adopter une posture d'attitude ludique, par la façon dont les mécanismes du système se délivrent dans leur actualisation. Lorsqu'il actualise un jeu vidéo, le joueur va prendre connaissance **du fonctionnement du système, de ses mécanismes (*game*)** et va en éprouver le potentiel de « jeu » (la jouabilité) en adoptant une attitude ludique (*play*), ce que traduit le terme de *gameplay* en regroupant ces deux aspects dans une même notion de forme nominale. S'il n'est pas possible de jouer aux échecs sans connaître préalablement les règles précédant l'action, un logiciel informatique d'échecs peut permettre à l'utilisateur de jouer sans en connaître par avance les règles. Les règles ne se donnent pas par avance mais se découvrent à travers la **progression du joueur dans le jeu**. Mais pour que le joueur puisse faire l'expérience du jeu, il est nécessaire de trouver un équilibre suffisant entre les contraintes mises en place par la structure et la liberté d'action du joueur (comme le souligne Roger Caillois, le jeu, pour exister, nécessite une certaine « latitude » d'invention laissée au joueur). En somme, s'il est trop contraint, le joueur ne pourra pas être « créatif ». En revanche, une absence de contraintes cohérentes au sein du système de jeu ne permettra pas au joueur de prendre connaissance des buts à poursuivre et des moyens possibles de leurs réalisations. L'attention se porte de la sorte

¹⁴ Les éléments de définitions du *gameplay* sont issus de la thèse de Sébastien Genvo sur le *game design* de jeux vidéo (2006 : 204-237). Le lecteur pourra s'y reporter pour avoir davantage de précisions sur ce point.

tout autant sur la structure que sur l'action elle-même, nécessitant un équilibre constant entre engagement et distanciation afin que l'action puisse être à la fois maintenue et évaluée (le joueur devant pouvoir prendre en compte les règles qui sont mises à sa disposition pour progresser et de quelle façon celles-ci lui permettent de s'exprimer). C'est cette dynamique même qui va constituer le *gameplay* de l'œuvre et susciter le (dé)plaisir du joueur, **selon ses propres représentations et connaissances relatives à l'activité ludique**. Ces quelques réflexions permettent de formuler trois grandes catégories de méthodes visant à mener une évaluation de la qualité d'un *gameplay* à partir d'une analyse du système de jeu.

Tout d'abord, la progression supposée du joueur dans l'œuvre peut être analysée à partir d'**éléments de sémiotique narrative (narratologie) appliqués** au médium vidéoludique. De nombreux auteurs considèrent en effet que la structure d'un jeu vidéo doit être analysée selon un modèle narratif pour permettre la prise en compte des choix mis à disposition du joueur. Stephen Kline, Nick Dyer-Witheford et Greig De Peuter (*Digital Play*, 2003) avancent notamment que l'interactivité est comprise comme l'allocation d'ouverture ou de fermeture à différents degrés à l'intérieur d'un « scénario » donné. Dans le même sens, Pierre Bruno remarque que « les mêmes structures du récit, les mêmes symboles se retrouvent dans les contes et légendes, dans un film d'Arnold Schwarzenegger ou les jeux vidéo » (*Les jeux vidéo*, 1993 : 38). Il relève notamment comment certains des jeux qu'il analyse sont structurés comme une « succession d'épreuves » (*ibid.* : 59).

L'interactivité incite bien entendu à adapter les modèles de narratologie aux spécificités du médium, et à ce titre les méthodes devaient se concentrer sur le lien existant dans les jeux vidéo entre récit et « *level design* » (soit l'architecture des niveaux de jeu), car la construction spatiale de l'univers ludique est un élément crucial dans la progression du joueur et dans sa découverte de la structure du jeu (pour Henry Jenkins par exemple, le *game designer* est à considérer comme un architecte de la narration). L'attendu consistait donc ici à éprouver ce cadre théorique pour l'évaluation qualité d'un jeu vidéo.

Outre la progression du joueur, la pertinence des mécanismes de jeu mis en scène selon les connaissances et/ou émotions que l'on souhaite transmettre devaient aussi faire l'objet d'une attention toute particulière. A cette fin, **les apports de la ludologie devaient être mobilisés et éprouvés**, cette discipline ayant pour vocation d'analyser, selon Gonzalo Frasca, « la structure et des éléments [d'un jeu] – particulièrement ses règles – de même que sur la création de typologies et modèles pour expliquer les mécanismes de jeux » (*Video game theory reader*, 2003 : 222). Il s'agit par exemple de mobiliser les catégories du jeu formulées par Roger Caillois (hasard, compétition, imitation, vertige) pour vérifier si l'activité demandée au joueur

correspond bien aux nécessités du « scénario » ludique mis en place. Enfin, la jouabilité se fonde essentiellement sur le stock de connaissances du joueur, sur ses compétences et sur la façon dont il va les acquérir et les mettre en œuvre (voir notamment l'ouvrage *Theory of fun* de Raph Koster). Un dernier ensemble de méthodes devait donc être dédié à la **description du joueur modèle postulé par le système de jeu**. Il s'agissait de permettre au concepteur d'évaluer si les phases d'apprentissage et de transmission de connaissances correspondaient bien aux exigences requises par ce modèle.

Le rôle du travail de recherche était donc d'identifier dans un premier temps des méthodes d'évaluation de la qualité du *gameplay* établies dans la littérature pour en envisager le portage dans le cadre du projet Lutin Game Lab. Ces méthodes devaient être expérimentables sur différents genres de jeux pour ne pas exclure *a priori* un projet de développement du protocole d'évaluation.

Paramètres analysés et conditions de mise en œuvre

À partir des réflexions précédentes, trois catégories de méthodes ont été identifiées et détaillées, chaque méthode ayant un intérêt pour l'évaluation du *gameplay*. Ces grandes catégories, regroupant chacune des sous-catégories, permettaient de balayer tous les points importants relatifs au *level design* et au *gameplay*. Ainsi, la base de travail pour guider les recherches était définie comme suit:

- La tension dramatique
 1. « Vie » du niveau et progressivité de la difficulté
 2. Cohérence entre les constituants de l'univers (personnages, etc.), objets d'ambiance (météo, etc.), objets de mise en scène (caméra, ce qui détermine la perception du joueur, etc.).
 3. Circulation dans un niveau, cohérence entre les épreuves à affronter et les passages essentiels à mettre en valeur (cette méthode est intéressante dans les niveaux avec notion de parcours, il s'agira conjointement de prendre en compte la répartition efficiente des objets dans le niveau).
 4. Gestion du temps et des conflits qu'il génère (par exemple, adéquation entre le niveau défini pour le joueur et le rythme des actions demandées).

- Mécanismes ludiques
 1. Adéquation entre la nature du mécanisme (hasard, conflit, imitation, réflexe, réflexion, etc.) et la situation de jeu proposée (accostage, résolution finale, etc.).

2. Opportunité de déclenchement des mécanismes (les actions disponibles pour le joueur sont adaptées ou non à la situation).

- Construction et mise en pratique des connaissances du joueur

1. Saturation cognitive, adéquation entre les informations transmises (savoir) et les actions demandées.

2. Répartition opportune des compétences (savoir-faire) pour la résolution des épreuves et des énigmes.

La première phase de formalisation du projet a donc consisté à effectuer un état de l'art au sein de la littérature sur le *level design* et le *gameplay* (voir bibliographie), afin d'identifier les principes fondamentaux pouvant permettre l'évaluation d'une œuvre vidéoludique et correspondant aux catégories énumérées précédemment. Par la suite, les informations commençant à s'accumuler, les catégories ont été simplifiées et certaines fusionnées. C'est en essayant de répartir dans les différentes parties toutes les informations regroupées jusque là, que nous avons identifié un problème : certaines parties semblaient redondantes, elles n'étaient pas toutes aussi consistantes ou certaines catégories semblaient difficiles à renseigner. Une catégorisation définitive a donc été établie comme suit :

- Cohérence de l'univers.
- Circulation dans le niveau.
- Gestion du temps.
- Adéquation entre mécanisme et situation.
- Saturation cognitive, adéquation entre les informations transmises (savoir) et les actions demandées.

Mise en forme de l'outil d'évaluation

Parce qu'il était nécessaire de prendre en considération les contraintes et souhaits des développeurs, tant du côté créatif que marketing, l'outil d'évaluation ne devait pas être trop dirigiste dans ses recommandations d'amélioration. Il était aussi nécessaire de proposer un outil suffisamment souple pour prendre en compte les évolutions des attentes des joueurs, qui sont centrales dans l'appréciation d'un produit culturel (voir Hennion, 2005). Dans ce cadre, si les principes à respecter pour assurer la qualité d'un *gameplay* sont stables, les façons dont

ils prennent forme évoluent en revanche dans le temps, selon l'évolution du marché, les innovations technologiques, etc. Il était nécessaire de proposer un outil qui permette de tenir compte de ces changements. Dans ce cadre, il a été décidé de mettre en place une plateforme collaborative de type wiki.

Un wiki, s'il est consulté régulièrement par la communauté des développeurs, peut en effet facilement s'adapter aux évolutions et ainsi rester une référence dans le domaine. Cet outil s'adresse aux développeurs de jeu, et les personnes les plus expertes dans le domaine du jeu vidéo ce sont elles. **Nous avons donc décidé de les laisser participer au projet en leur permettant d'ajouter les références qui leur semblent importantes.** Ce point permettait également d'**attirer l'attention de la communauté sur l'ensemble du projet à partir d'une entrée collaborative gratuite.** De même l'évaluation de la qualité est effectuée par le concepteur lui-même, qui peut se rapporter à une catégorie précise selon ses besoins et trouver des principes à respecter pour s'assurer de la qualité du *gameplay*, de même que des exemples de jeux mettant en œuvre ces principes. La modération des modifications se fera a posteriori par un chercheur associé au Lutin.

Détails sur le fonctionnement de la plateforme wiki Lutin game lab

Chaque catégorie de méthodes devait comporter au moins 4 principes généraux à respecter permettant d'assurer la qualité du produit. Par exemple, il a été établi par de nombreux auteurs qu'il faut alterner les situations où le joueur est dominant (moyens > obstacles) et les situations de grande tension (obstacles > moyens). Chacun de ces principes connaît une série de traductions possibles qui se retrouvent de façon récurrente dans plusieurs jeux, mais qui découlent avant tout de la créativité de chaque concepteur et de ses impératifs de développement. Pour le principe précédent, certains jeux accompagnent notamment le joueur avec un personnage non joueur (PNJ) pour lui faciliter l'apprentissage des premiers moments de jeu, ce PNJ étant mis au second plan par la suite. Chaque traduction de principe peut être exemplifiée à partir de situations issues de jeux édités. À titre indicatif, *Gears of war* sur Xbox 360 propose d'accompagner le joueur lors du didacticiel et du premier niveau par un équipier qui lui sert de tuteur, pour laisser par la suite le joueur être le guide de son équipe. Ces multiples indications permettent donc de servir de référencement pour les concepteurs. **Le *game designer* peut notamment estimer si ses choix de conception sont en adéquation avec les normes établies précédemment, mais il peut également décider de s'en écarter**

dans une optique d'innovation. En effet, la plateforme wiki Lutin game lab permet aussi d'établir un ensemble de fiches d'identité de jeux, présentant la façon dont chacun décline les principes de qualité du *gameplay*. Pour faciliter ce travail d'évaluation, le site fonctionne donc selon deux types d'entrées, l'un correspondant aux catégories de principes à respecter (progression dans le niveau, cohérence de l'univers, adéquation entre mécanismes ludiques et situation, etc.), l'autre à la liste des jeux référencés. Le projet a donc une structure matricielle qui peut être utile pour un développeur. Il peut regarder dans les fiches de chaque catégorie les principes à respecter et voir des exemples significatifs de plusieurs jeux ou, s'il apprécie particulièrement un jeu, il peut le prendre pour modèle en regardant ses différentes caractéristiques sur la fiche du jeu. Et comme nous l'avons souligné, **la plateforme se veut ouverte à la communauté des concepteurs**, qui peuvent augmenter ou créer de nouvelles fiches d'exemples selon la façon dont leur jeu répond à certains principes de qualité. La communauté peut apporter de nouvelles traductions de principes, ou des principes généraux si ceux-ci sont fondés sur le recoupement de plusieurs sources d'analyses scientifiques et professionnelles. Conjointement, **la recension des différents jeux permet d'établir un réservoir d'idées de mécanismes en *game design*, qui sera à la fois profitable à la formation des futurs *game designers* et aux concepteurs souhaitant faire un état de l'art sur l'existant en fonction d'un besoin particulier.** L'exploration de la base de données peut donc se faire à partir d'une double entrée, soit par nom de jeux, soit par catégorie de méthodes.



Illustration 1. Page d'accueil du wiki (capture d'écran du prototype initial)

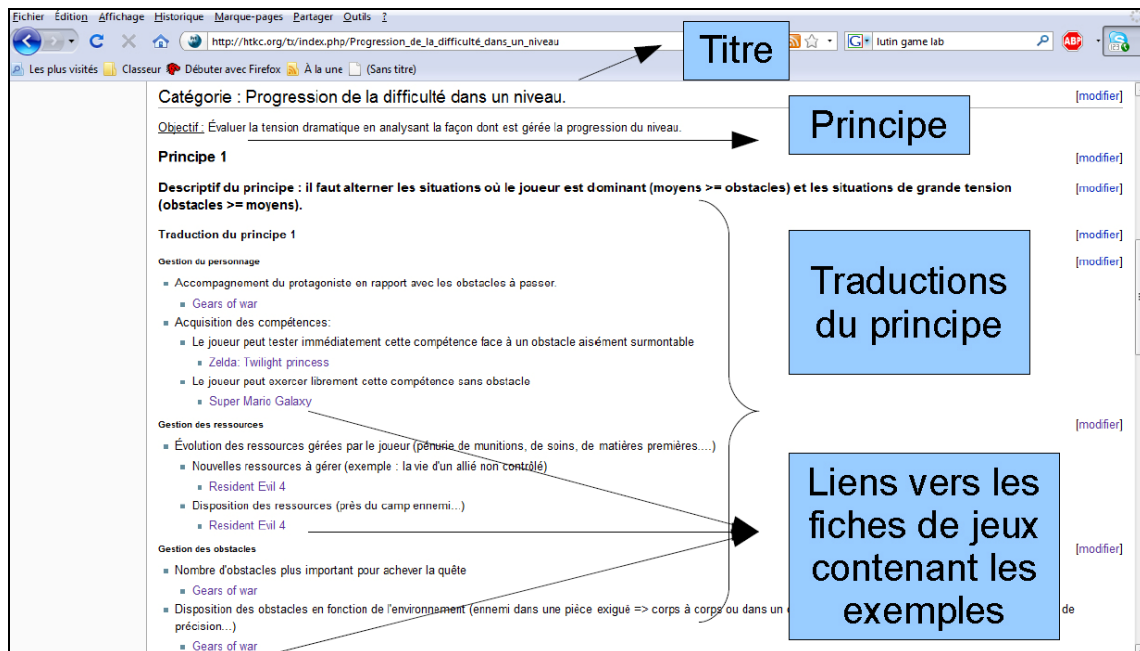


Illustration 2. Exemple de catégorie de méthodes (capture d'écran du prototype initial)



Illustration 3. Exemple de fiche de jeu (capture d'écran du prototype initial)

Mise en forme de l'interface

Un dernier point d'importance dans la finalisation de la plateforme était la mise au point d'une interface intuitive et conviviale permettant de retenir l'attention de la communauté, tout

en facilitant la participation. Tout comme sur le prototype de plateforme (voir capture d'écrans précédentes), il a été décidé de favoriser une entrée par catégorie de principes mais aussi par fiches de jeux. Il était également important que la plateforme soit intégrée au site du Lutin, de sorte à permettre sa vocation de communication sur le projet, ce qui imposait une certaine charte graphique. Il a également été décidé de développer une plateforme wiki propriétaire, afin de permettre une mise à jour dynamique des données, ce que ne permettait le prototype, conçu à partir de la solution logicielle mediawiki. Par exemple, une modification à un certain degré d'arborescence de la dénomination d'un principe ne répercutait pas ce changement à l'ensemble des niveaux où le principe apparaissait, obligeant en report manuel de la correction dans l'ensemble des pages. De plus, le moteur MediaWiki utilisé n'avait qu'une gestion très limitée des liens entre parties. Ce mode de fonctionnement ne permettait donc pas de répondre aux impératifs d'utilisabilité de la plateforme et un développement *ad hoc* a donc été initié.

Lutin Gamelab
observatoire des jeux video, **ergonomie** • **gameplay**

Méthodes d'évaluation du gameplay par l'intermédiaire du level design

Les présentes méthodes ont pour objectif général d'aider le concepteur à diagnostiquer la qualité de son gameplay, à partir d'une réflexion se fondant sur le level design. Des méthodes de validation sont établies à partir de la littérature. Chacune des méthodes est détaillée en principes à respecter pour assurer la qualité du produit, ces principes étant déclinés sous forme de mises en application concrète (traductions de principe), avec des études de cas précis pour la plupart des traductions. Le concepteur peut de la sorte estimer si ses choix de conception sont en adéquation avec les normes établies précédemment (ce qui peut également le décider à s'en écarter dans une optique d'innovation). Ce site propose en définitive un ensemble de fiches d'identité de jeux. Il se veut collaboratif et ouvert à la communauté des concepteurs, qui peuvent augmenter ou créer de nouvelles fiches d'exemples selon la façon dont leur jeu répond à certains principes de qualité.

Authentification

Nom d'utilisateur

Mot de passe

Connexion

Navigation

Par catégorie
Choisis une catégorie ▼

Par jeu
Choisis un jeu ▼

Recherche

Rechercher

Divers

© Lutin Gamelab 2008

Illustration 4. Capture d'écran de la page d'accueil avec l'interface définitive

Lutin Gamelab
 observatoire des jeux video, *ergonomie* • *gameplay*

Progression dans le niveau
 Objectif : Evaluer la tension dramatique en analysant la façon dont est gérée la progression du niveau.

Catégorie
 Progression de la difficulté dans un niveau.

Principe 1
 Descriptif : Il faut alterner les situations où le joueur est dominant (moyens >= obstacles) et les situations de grande tension (obstacles >= moyens).

Gestion du personnage
 Accompagnement du protagoniste en rapport avec les obstacles à passer.

- [Gears of war](#)

Acquisition de nouvelles compétences:

- Le joueur peut tester immédiatement une nouvelle compétence face à un obstacle aisément surmontable
 - [Zelda: Twilight princess](#)
- Le joueur peut exercer librement une nouvelle compétence sans obstacle
 - [Super Mario Galaxy](#)

Gestion des ressources
 Les ressources gérées par le joueur évoluent afin de moduler la difficulté (munitions, soins, matières premières....)

- Nouvelles ressources à gérer (exemple : la vie d'un allié non contrôlé)
 - [Resident Evil 4](#)

Utilisateur
 jeremy est connecté(e)
 Mon profil utilisateur
 Editer l'article
 Déconnexion

Navigation
 Par catégorie
 Choisis une catégorie
 Par jeu
 Choisis un jeu

Recherche
 Recherche

Divers

Illustration 5. Capture d'écran d'une page de catégorie de principe avec l'interface définitive

Difficultés rencontrées

Outre la nécessité de développer une solution logicielle wiki propriétaire (qui n'était pas envisagée à l'origine), le principal problème a été de ne pouvoir établir une phase de retour des entreprises partenaires compte tenu de leur faible disponibilité et de leur implication dans d'autres aspects du projet, où leurs apports comportaient un caractère impératif pour la réussite des sous-projets annexes. Dans ce cadre, une expérimentation du wiki sur des développements en cours n'a pas pu être menée non plus. Ces problèmes ne constituent pas néanmoins un handicap à la réussite de la plateforme wiki, car il s'agit justement d'un outil ouvert à l'ensemble de la communauté francophone des développeurs, cette phase de *feedback* sera donc perpétuelle à partir du moment où l'existence du site sera publicisée.

Déroulement effectif des tâches en comparaison du calendrier prévisionnel

- **Prévisionnel initial**

SP4	Méthodes <i>gameplay</i>			T.+3		T.+6		T.+9		T.+12		T.+15		T.+18		T.+21		T.+24
T4.1	Formalisation																	
T4.2	Expérimentation sur des jeux existants																	
T4.3	Expérimentation sur des jeux en développement																	
T4.4	Feedback																	

- **Déroulement effectif**

SP4	Méthodes <i>gameplay</i>			T.+3		T.+6		T.+9		T.+12		T.+15		T.+18		T.+21		T.+24
T4.1	Formalisation et état de l'art																	
T4.2	Expérimentation sur des jeux existants																	
T4.3	Développement du wiki prototype																	
T4.4	Développement de l'interface finale																	

Bibliographie

- Bachimont B., *Ingénierie des connaissances et des contenus : le numérique entre ontologies et documents*, Paris, Hermès – Lavoisier, 2007.
- Bruno Pierre, *Les jeux vidéo*, SYROS, 1993.
- Caillois Roger, 1967, *Les jeux et les hommes*, Paris, Gallimard, 1958.
- Crawford Chris, *The art of computer game design*, McGraw-Hill Osborne Media, 1984.
- Crawford Chris, *Chris Crawford on Game Design*, 2003, Indianapolis, New Riders.
- Desurvire, H., Caplan, M., AND Toth, J. A. 2004. Using heuristics to evaluate the playability of games. In *Extended Abstracts of the 2004 Conference on Human Factors in Computing Systems*. ACM Press, New York, 1509-1512.
- Falstein, N., Barwood, H. The 400 Project, Available at http://theinspiracy.com/400_project.htm
- Federoff, M. 2002. Heuristics and usability guidelines for the creation and evaluation of fun in video games. Unpublished thesis, Indiana Univ., Bloomington.
- Fortin, Tony Philippe Mora et Laurent Trémel, *Les jeux vidéo : pratiques, contenus et enjeux sociaux*, Paris, L'Harmattan, coll. Champs visuels, 2005.
- Frasca Gonzalo, *Videogames of the oppressed : videogames as a mean for critical thinking and debate*, Master Thesis, Georgia Institute of technology, 2001.
- Frasca Gonzalo, « Simulation versus narrative », pp. 221-235, in : Wolf Mark J.P., Perron Bernard, dirs., *The video game theory reader*, Routledge, 2003.
- Fullerton et Al., *Game design workshop*, CMP Books, 2004.
- Genvo Sébastien, *Introduction aux enjeux artistiques et culturels des jeux vidéo*, L'Harmattan, 2003,.
- Genvo Sébastien, Dir., *Le game design de jeux vidéo : approches de l'expression vidéoludique*, L'Harmattan, 2006.
- Greenfield Patricia, Retschizki Jean, *L'enfant et les médias : les effets de la télévision, des jeux vidéo et des ordinateurs*, Editions Universitaires de Fribourg, 2002.
- Guardiola Emmanuel, *Ecrire pour le jeu*, 2000, Dixit.
- Hennion A., 2005, « Pour une pragmatique du goût », *Papiers de Recherche du CSI*, 1, en ligne, <www.csi.ensmp.fr/Items/>
- Henno Jacques, *Les jeux vidéo*, Le cavalier Bleu, 2002.
- Ichbiah Daniel, *La saga des jeux vidéo*, Vuibert, 2004.
- Jenkins Henry, 2002, « Game design as narrative architecture », in : Pat Harrington and Noah Frup-Waldrop, Eds., *First Person*, Cambridge, MIT Press, <http://web.mit.edu/cms/People/henry3/games&narrative.html>

- Jollivalt Bernard, *Les jeux vidéo*, Presses universitaires de France, 1994.
- Kelman Nic, *Jeux vidéo - L'art du XXIème siècle*, Assouline, 2005.
- Kent Steve L., *The ultimate history of video games*, New York, Three River Press, 2001.
- Kerbrat Jean-Yves, *Manuel d'écriture de jeux vidéo*, 2006, L'Harmattan.
- Kerdellant Christine, Grésillon Gabriel, *Les Enfants puce : Comment Internet et les jeux vidéo fabriquent les adultes de demain*, Denoël, 2003.
- Kline S., Dyer-Witheford N., De Peuter G., *Digital Play*, Montreal & Kingston, McGill-Queen's University Press, 2003.
- Koster R., 2005, *A theory of fun*, Scottsdale, Paraglyph Press.
- Le Diberder Alain & Frédéric, *L'univers des jeux vidéo*, La découverte, 1998.
- Le Diberder Alain & Frédéric, Dirs., « La création de jeux vidéo en France en 2001 », *bulletin du département des études et de la prospective*, 139, 2002.
- Lefebvre Jérémie, *La société de consolation*, Sens et Tonka, 2000.
- Lenhard Gentianne, *Faut-il avoir peur des jeux vidéo ?*, ESF Editeur, 1999.
- Montagnana Vincent, *L'empire des jeux*, Timée-Editions, 2005.
- Rolling Andrew & Ernest Adams, *on game design*, New Riders, 2003.
- Rolling Andrew & Morris Dave, *Conception et Architecture des jeux vidéo*, Vuibert, 2005.
- Natkin Stéphane, *Jeux vidéo et médias du XXIe siècle*, Vuibert, 2004.
- Rasmussen J., 1983, *Skills-Rules-Knowledge : signals, signs, and symbols*, IEEE Transactions on Systems: Man and cybernetics, 13.
- Réseaux*, « Les jeux vidéo », CNET, 67, septembre/octobre 1994.
- Roustan Mélanie, dir., *La pratique du jeu vidéo : réalité ou virtualité ?*, L'Harmattan, 2003.
- Salen Katie, Zimmerman Eric, *Rules of Play*, MIT Press, 2004.
- Trémel Laurent, *Jeux de rôles, jeux vidéo, multimédia, les faiseurs de mondes*, Presses Universitaires de France, 2001.
- Virole Benoît, *Du bon usage des jeux vidéo*, Hachette littératures, 2003
- Winnicott D. W., *Jeu et réalité, l'espace potentiel*, Paris, Gallimard, 1971.
- Wolf Mark J.P., Perron Bernard, dirs., *The video game theory reader*, Routledge, 2003.

Contributeurs au sous-projet 4

Coordination et état de l'art :

Sébastien Genvo, Université de Limoges

Serge Bouchardon, UTC de Compiègne

Formalisation des méthodes et conception du prototype de la plateforme :

Jennifer Dumont, étudiante en Génie biologique (5^{ème} année d'école d'ingénieurs), UTC de Compiègne

Guilhem Lettron, étudiant en Génie informatique (4^{ème} année d'école d'ingénieurs), UTC de Compiègne

Développement de l'interface définitive :

Jérémy Cerri, étudiant en Génie informatique (5^{ème} année d'école d'ingénieurs), UTC de Compiègne

Antoine Vincent, étudiant en Génie informatique (5^{ème} année d'école d'ingénieurs), UTC de Compiègne



SP5 Rapport final du projet Lutin Gamelab

Expert artificiel pour le diagnostic

**Marc Damez
Nicolas Labroche
Marie-Jeanne Lesot**

Laboratoire d'Informatique de Paris 6

Février 2009

Table des matières

Objectifs	148
Description initiale	148
Axes de travail développés dans le cadre du projet :	149
Structure du rapport	149
1. Etats de l'art	151
1.1 Les heuristiques d'évaluation des jeux vidéo	151
1.2 Emotions et signaux physiologiques	152
1.2.1. Les traces oculaires	152
1.2.2. Analyse des autres signaux physiologiques	154
2. Modèle proposé	154
2.1 Structuration hiérarchique des heuristiques de jeux vidéo	154
2.2 Spécification du modèle pour chaque type de jeu	155
2.2.1 Les jeux de tir, First Person Shooter (FPS)	155
2.2.2 Les jeux de sport	158
2.2.3 Les jeux de stratégie, Real Time Strategy (RTS)	158
2.2.4 Les Role Playing Game (RPG)	159
2.3 Modèle d'évaluation automatique : formalisme des gabarits	160
2.3.1 Présentation des gabarits	161
2.3.2 Avantages du formalisme des gabarits pour l'évaluation des jeux vidéo	161
2.3.4 Extension aux treillis	162
3. Mise en œuvre du modèle : expérimentations	163
3.1 Données objectives	163
3.2 Données subjectives	164
3.3 Fouille de données dans les signaux physiologiques	165
3.3.1. Formulation du problème et solutions proposées	166
3.3.2. Construction du jeu de données	166
3.3.3. Expérimentations et résultats	167
3.3.4. Conclusion	167
4. Réalisations logicielles	167
4.1 GamelabQuest	167
4.1.1 Structure des questionnaires	168
4.1.2 Les QCM évidentiels	168
4.1.3 Utilisateurs de GamelabQuest	169
4.1.4 Modules fonctionnels de l'application GamelabQuest	169
4.2 EvalGen	171
4.3 Plugin GamEval	174
4.4 GamExpé	174
4.3.1. Conception d'un modèle spécifique de gabarit	175
4.3.2. Diagnostic d'un jeu	176
5. Conclusions et perspectives	176
Références	178
Annexe	180

Objectifs

L'objectif du projet Lutin Game Lab a été de fédérer les forces de recherche académiques et de développement des entreprises pour faire faire un pas significatif aux méthodologies d'évaluation de la qualité d'un jeu vidéo fini ou en cours de développement.

Le Laboratoire d'Informatique de Paris 6 (LIP6) a notamment été sollicité pour :

- ses compétences en analyse de données et apprentissage, permettant d'extraire des connaissances automatiquement en utilisant des algorithmes intelligents,
- ses connaissances en informatique, implémentation de système permettant la mise en place d'une plateforme globale de traitement.

Description initiale

Plus précisément, les objectifs du LIP6 dans le projet, tels que décrits dans le sous-projet 5, sont rappelés ci-dessous :

« Il s'agit de construire un expert intelligent d'aide au diagnostic. Le rôle d'un tel système est de permettre l'identification de problèmes contraires aux objectifs initiaux, pendant toutes les différentes phases de développement du jeu. Les données à traiter peuvent être objectives (obtenues par capteurs oculaires, capteurs haptiques, traces d'interaction, etc.) ou subjectives (questionnaires aux différents acteurs du monde du jeu vidéo, (concepteurs, éditeurs, joueurs, presse spécialisée...), remarques ou impressions notées par des joueurs ou des experts, etc.) L'ensemble de ces informations sera représenté par des descripteurs flous, plus à même de prendre en compte les variabilités individuelles lors d'évaluation subjectives ou l'imprécision relative aux instruments de mesures pour les données objectives. L'analyse de ces données permettra la construction d'un diagnostic facilement interprétable par un expert humain du domaine. »

Afin de réaliser ces travaux de recherche et de développement, le LIP6 a proposé une décomposition en 4 sous-tâches, validées par l'ensemble des partenaires :

- Tâche 1 : définition des objectifs ludiques

Il s'agit de constituer une typologie des objectifs ludiques pour les différents types de jeux. Un recensement des objectifs d'un jeu devra comporter l'ensemble des avis des concepteurs du jeu tout au long de sa phase de développement. Les objectifs dont le système vérifiera qu'ils sont satisfaits seront choisis en concertation avec les entreprises partenaires.

- Tâche 2 : récolte de données pour l'évaluation qualitative des objectifs

Les objectifs énoncés précédemment (tâche 1) devront être identifiés automatiquement à partir d'un ensemble de données objectives et subjectives. Il s'agit donc d'établir une corrélation entre ces objectifs ludiques et l'ensemble des données récoltables pour l'évaluation automatique de l'accomplissement de ces objectifs.

Cette tâche doit être effectuée en collaboration avec le SP2 chargée de l'extraction d'expertise auprès des professionnels, ainsi qu'auprès des joueurs. Cette étape est cruciale pour le reste de l'étude car elle permet la transcription des connaissances « entreprise » du domaine en information exploitable par les méthodes d'analyse de données.

- Tâche 3 : Analyse des données et formulation du diagnostic

L'analyse des données précédemment récoltées (tâche 2) devra permettre la conception de recommandation sous forme de diagnostic. Il s'agit de mettre en évidence le succès ou l'échec des différents objectifs ludiques pour un jeu. Les analyses devront montrer les caractéristiques produisant les effets voulus ainsi que les effets non souhaités, nuisibles à l'accomplissement des objectifs.

- Tâche 4 : validation par les entreprises

Enfin, la dernière tâche consiste en un feedback général sur l'intérêt des différentes méthodes et les

premières évaluations recueillies auprès des entreprises pour une validation du système.

Axes de travail développés dans le cadre du projet :

Dans ce cadre, le LIP6 a développé plusieurs axes de travail qui couvrent un large spectre de compétences :

la récolte des avis des joueurs par le biais d'un système de questionnaire intelligent
l'étude des signaux physiologiques et leur association à des mécanismes du jeu et des états émotionnels des joueurs (en collaboration avec les spécialistes de Paris 8) par des techniques d'apprentissage supervisé,

le développement d'un expert pour le diagnostic d'un jeu reposant sur les deux types précédents d'information (réponses subjectives aux questionnaires et signaux objectifs enregistrés sur les joueurs), dont les conclusions nourrissent à différents niveaux l'expert artificiel prédictif mis en œuvre en parallèle dans le projet.

Les différentes étapes de ce travail sont résumées dans le schéma de la figure 1 ci-dessous (les éléments notés en gras indiquent la participation du LIP6).

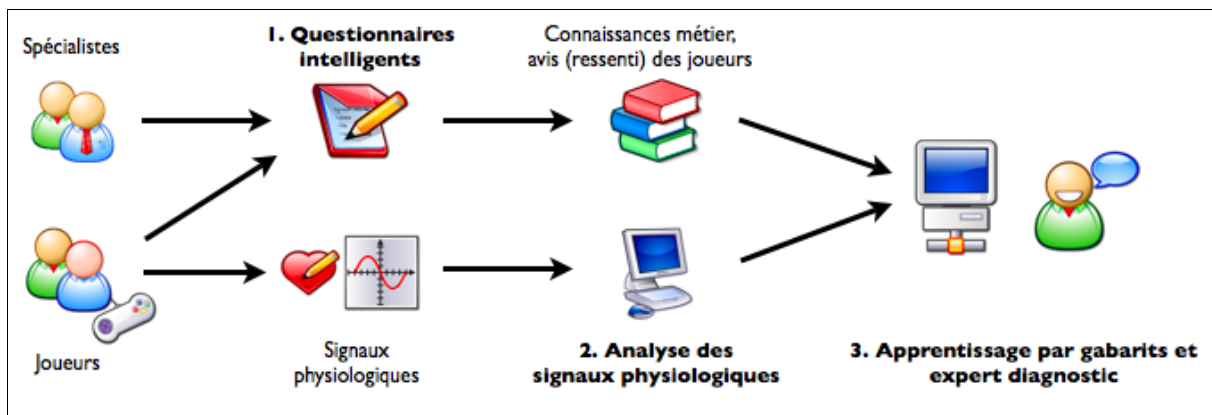


Figure 1. Solutions proposées par le LIP6 : analyse conjointe des signaux physiologiques et des questionnaires utilisateurs pour permettre l'évaluation automatique d'un jeu vidéo dans les différentes phases de sa conception.

Structure du rapport

Dans ce rapport, nous présentons les travaux conduits pour répondre à ces axes de recherche : la détermination des questions à poser dans les questionnaires et plus généralement des critères utilisés par l'expert artificiel a été réalisée à partir de connaissances expertes modélisées dans les heuristiques d'évaluation des jeux vidéo. Nous avons proposé une organisation et une hiérarchisation des éléments constitutifs de ces évaluations de jeux vidéo, ainsi qu'une modélisation de cette structure hiérarchique dans le formalisme des gabarits. Au-delà de la modélisation, ce formalisme permet l'automatisation du processus d'évaluation en fonction de données recueillies lors de tests d'utilisation d'un jeu fixé conduisant à l'obtention d'une évaluation numérique. Nous avons testé ce modèle dans le cadre d'expérimentations sur des joueurs réels, qui ont conduit à valider à la fois l'utilisation des questionnaires et des signaux physiologiques. Nous présentons également les logiciels développés dans le cadre du projet qui permettent de mettre en œuvre la démarche proposée, tant pour la récolte des données, subjectives ou objectives, que pour leur traitement pour le processus d'évaluation.

Plus précisément, le rapport s'organise comme suit : la section 1 présente les états de l'art sur les

heuristiques d'évaluation des jeux vidéo et l'étude des signaux physiologiques pour la caractérisation des émotions et du ressenti des joueurs. La section 2 présente la hiérarchisation des critères d'évaluation proposée, dans un modèle global et dans un modèle spécifique à chacun des types de jeux vidéo étudiés dans le projet, ainsi leur formalisation dans le cadre des gabarits, en discutant les avantages de cette approche. La section 3 décrit les résultats expérimentaux obtenus dans le cadre d'expérimentations conduites sur des données issues de questionnaires ainsi que sur des signaux physiologiques. La section 4 présente les réalisations logicielles développées par le LIP6 dans le cadre du projet Gamelab. Enfin la section 5 est consacrée au bilan et aux perspectives qui restent ouvertes.

1. Etats de l'art

1.1 Les heuristiques d'évaluation des jeux vidéo

Il existe de nombreux travaux sur l'évaluation des jeux vidéo, qui fournissent des critères d'évaluation suivant différents principes : certains sont basés sur l'adaptation des critères d'évaluation de la qualité de logiciels [8, 5, 7], d'autres se concentrent des aspects spécifiques des jeux vidéo selon différents points de vue, comme l'immersion dans le jeu [2], l'amusement du joueur, dans le cadre de la ludologie, [14], ou le rôle de la compétition [15]. Dans des sites web portant sur les jeux vidéo, les notes données permettent de lister d'autres critères par exemple [17, 18, 19, 20].

En général, les heuristiques sont organisées en quatre catégories :

- le *game play* regroupe les aspects liés aux problèmes, énigmes, défis, obstacles auquel le joueur est confronté durant la partie,
- le *game story* concerne le développement de l'histoire, des intrigues et des personnages,
- le *game mechanics* représente la façon dont le joueur interagit avec l'environnement,
- le *game interface* concerne les outils que le joueur utilise pour interagir avec le jeu.

Ces catégories ont été testées et éprouvées et sont fréquemment utilisées dans la littérature.

L'annexe A présente à titre d'exemple la grille d'évaluation issue de [5]. Nous avons laissé les heuristiques telles qu'elles sont formulées dans la littérature. Elles sont donc en anglais.

Les heuristiques prennent des formes très différentes : certaines sont très intuitives telles que l'importance du scénario ("*great interesting story line*") ou des personnages ("*player is interested in the characters*"), l'organisation de l'interface ("*menu well organized and minimalist*") ou encore la clarté des objectifs ("*clear goals*"). D'autres sont plus recherchées et concernent par exemple la relation entre l'apprentissage du jeu et les compétences du joueur ("*learning should provide skills*", "*challenges require skills*") ou encore la place de la cohérence du monde dans le processus d'immersion du joueur ("*game should react consistently*", "*make consequences of actions predictable*").

On constate que certaines heuristiques provenant de différentes sources sont proches les unes des autres, quoique différentes, et participent à des mêmes éléments généraux. De plus, on observe que certains critères appartiennent à plusieurs catégories, ou sont à mettre en relation les uns avec les autres. Aussi, comme détaillé dans les sections 2.1 et 2.2, nous avons proposé un modèle hiérarchique regroupant plus de 70 critères, afin de mettre en évidence leurs relations et leurs rôles dans les principes de jeu. Le but d'une telle classification est de présenter une vision globale et compréhensible des règles à vérifier pour identifier un "bon" jeu vidéo : elle fournit une organisation structurée des critères permettant en particulier de distinguer des propriétés élémentaires ou complexes.

Par ailleurs, le plus souvent, les travaux ne donnent pas d'indication sur la façon dont ces critères peuvent être utilisés pour évaluer la qualité d'un jeu et l'exploitation pratique de ces listes d'heuristiques n'est pas décrite. Lorsqu'elle est spécifiée, il apparaît que la qualité globale d'un jeu est définie comme la moyenne de l'évaluation des critères individuels [14], et elle ne prend pas en compte la complexité et la spécificité des différents types de jeux. Exploitant la structuration hiérarchique des critères d'évaluation, nous avons proposé de la modéliser dans le formalisme des gabarits, comme décrit dans la section 2.3, pour fournir un outil d'évaluation automatique. Ce sont ensuite les feuilles du gabarit, qui représentent les critères élémentaires, qui définiront les questions posées aux utilisateurs dans les questionnaires.

1.2 Emotions et signaux physiologiques

Les signaux physiologiques doivent permettre de mieux comprendre les causes d'un comportement humain observable. Dans le cas des jeux vidéo en particulier, ils ont pour objectif de donner des indications sur le ressenti des joueurs, qui permettent de compléter les informations fournies par les questionnaires : ces derniers donnent des informations subjectives, provenant de l'opinion des joueurs, les signaux physiologiques fournissent des informations objectives, mesurées par leurs réactions physiologiques. Ce couplage doit également permettre d'intégrer à la fois des réactions immédiates et des réactions a posteriori à la qualité du jeu.

L'idée des signaux physiologiques est de mettre au jour des mécanismes émotionnels personnels et de les analyser pour déterminer dans quelle mesure ils sont la cause ou la conséquence des actions de la personne. Ces travaux sont à la frontière d'un ensemble de compétences allant du traitement du signal, de l'apprentissage artificiel et bien sûr des sciences cognitives.

De nombreux travaux voient actuellement le jour dans le cadre d'un thème de recherche nommé *affective computing* ou *emotional engineering* et qui a pour vocation la mise en lumière des émotions / sentiments éprouvés par les utilisateurs d'un système.

Les applications sont nombreuses comme dans le domaine du design industriel (vêtements, lignes de voiture qui évoquent la sympathie ou la puissance par exemple), le domaine des communications intelligentes (standards téléphoniques artificiels qui doivent pouvoir détecter la colère, l'impatience dans la voix des interlocuteurs humains), ou dans des domaines plus artistiques, par exemple pour adapter l'ambiance d'une exposition en fonction de ce que l'on veut faire ressentir à des visiteurs ou pour analyser des textes pour déterminer l'opinion (positive / négative) des auteurs vis-à-vis du sujet abordé.

Du fait des progrès en électronique, plusieurs types de signaux peuvent être captés de manière peu intrusive et à moindre frais (voir [9] pour une discussion argumentée des possibilités).

Nous retenons dans notre étude les principaux signaux suivants :

- les traces oculaires obtenues par enregistrement des mouvements de la rétine de l'œil,
- les signaux électrodermaux (obtenus par mesure de la conductance électrique de la peau),
- les électrocardiogrammes qui retranscrivent la fréquence cardiaque.

Dans ce qui suit, nous présentons un bref état de l'art sur ces signaux et l'exploitation qui en est faite. Dans les sections 3.2 et 3.3, nous décrivons les expérimentations réalisées dans le cadre du projet avec de telles données, ainsi que les résultats qui ont été obtenus.

1.2.1. Les traces oculaires

L'analyse des traces oculaires, ou oculométrie, s'est principalement développée en sciences cognitives dans le cadre de l'*étude des processus*, ayant pour but d'évaluer l'attention d'une personne, sa perception, le cheminement de son raisonnement voire sa compréhension.

Parmi les premiers travaux dans ce domaine, on peut citer Yarbus [16] qui a réalisé des expérimentations visant à suivre les mouvements des yeux de sujets auxquels différentes questions ont été posées. Il a émis l'idée que les mouvements oculaires sont liés à un contexte comportemental.

L'oculométrie est une technique d'enregistrement des mouvements des yeux qui consiste à repérer en temps réel la position du regard. Il existe pour cela différentes méthodes, parmi lesquelles l'utilisation d'un détecteur optique ou d'une caméra vidéo calés sur le reflet émis par un rayon infrarouge envoyé sur la cornée oculaire.

Ce dispositif couplé à un système informatique échantillonne régulièrement la position spatiale de l'œil et dans certains cas le diamètre pupillaire. La quantité considérable de données enregistrées est ensuite réduite pour ne retenir que deux types d'information :

- les pauses de l'œil appelées *fixations* témoignent des traitements cognitifs,
- les sauts d'une fixation à l'autre, qui sont appelés *saccades*, et qui sont davantage sous le contrôle de la perception et des mécanismes oculomoteurs.

Les fixations et les saccades représentent les éléments fondamentaux de l'étude oculométrique à partir desquels sont calculées plusieurs mesures spatiales et temporelles du déplacement du regard :

- les mesures spatiales sont les distances saccadiques, les localisations ou le tracé des zones inspectées par le regard « scanpath »,
- les mesures temporelles concernent les durées des fixations globales ou locales, c'est-à-dire limitées à une information précise ou à une zone.

Ces indicateurs permettent de distinguer des mouvements oculaires de différents types :

- la poursuite, qui consiste à suivre un objet et est constitué d'une succession de petites saccades et une petite fixation très rapprochée,
- les tremblements, qui sont dus à des micro-déplacements de l'œil qui, par nature, doit perpétuellement être en mouvement,
- la rotation,
- la dérive (« drift » en anglais) : la fatigue musculaire induite par une fixation prolongée entraîne une dérive du point de regard

De nombreuses analyses des fixations et des saccades ont en particulier eu lieu dans le cadre de la lecture de textes. Elles ont montré qu'il faut distinguer deux types de fixations qui ne possèdent pas la même signification au niveau du comportement que l'on cherche à expliquer :

- les fixations progressives se produisent souvent de gauche à droite en lecture et révèle un besoin d'acquisition d'information
- les fixations régressives sont caractérisées par un retour en arrière, ayant pour but de compléter/affirmer une information précédemment acquise

Des travaux ont montré que, lorsque l'on considère une succession de fixations (comme la lecture d'une phrase par exemple), celles-ci ne sont ni indépendantes ni équivalentes : elles ne sont pas indépendantes car il a été observé qu'il y a une influence des mots précédents (faisceau attentionnel) sur ce qui est en cours d'observation et qu'il y a également un effet d'anticipation sur les mots suivants dans la phrase qui ne permettent donc pas de conclure sur l'importance cognitive d'une fixation hors contexte. De plus, les fixations ne sont pas équivalentes comme l'ont montré des travaux statistiques sur la corrélation entre la durée des fixations et l'amplitude des saccades qui les suivent ; ainsi dans une interface bien organisée, pour des fixations sur les mêmes éléments d'interface, on peut observer soit des comportements de *découverte* dans le cas où les fixations sont courtes et suivies de longues saccades ou bien un comportement de *recherche ciblée* quand les fixations sont longues et suivies de courtes saccades. La non équivalence peut également être observée au niveau physiologique en utilisant un électroencéphalogramme qui montre que les fixations n'entraînent pas la même charge de travail au niveau cognitif.

Il existe donc un large panel de mesures qui peuvent être effectuées pour caractériser l'activité oculaire d'un utilisateur. On peut citer par exemple la durée moyenne des fixations, ou leur nombre, la durée moyenne des saccades, leur taille ou leur fréquence, la distance parcourue à l'écran (*scan path*), l'aire convexe qui est la surface dans laquelle se concentrent les fixations, la somme des durées sur des régions d'intérêts définies manuellement par l'expérimentateur, ou encore la séquence de ces régions d'intérêt.

Enfin, le diamètre pupillaire est également étudié actuellement pour savoir s'il permet d'inférer des

informations sur l'état de concentration de la personne, sur son stress ou ses émotions, bien qu'il soit difficile de conclure étant donné qu'il varie principalement avec la luminosité ambiante.

1.2.2. Analyse des autres signaux physiologiques

Notre étude bibliographique a principalement porté sur l'analyse des signaux oculaires, car c'est dans ce domaine que la littérature est la plus abondante et que plusieurs systèmes d'oculométrie sont accessibles au sein du LUTIN.

Plusieurs auteurs proposent des pistes de travail pour la manipulation des signaux physiologiques. Boucsein [1] interprète le niveau des signaux électrodermiques comme un indicateur de la force de l'émotion ressentie. Collet [4] insiste sur le fait que le signal électrodermal doit être accompagné d'une autre mesure, telle que l'électro-encéphalogramme, pour pouvoir être interprété. Enfin, Levenson [11] associe les différentes variations à des sentiments positifs ou négatifs et approfondit l'étude pour distinguer des émotions de base comme la colère ou la peur.

A l'heure actuelle, il ne semble pas exister de modèle formel décrivant des indicateurs plus précis issus de ces signaux et les liens éventuels avec le ressenti du joueur et ses émotions.

Nous avons effectué quelques expérimentations conduites sur le jeu Halo 3, décrites dans la section 3.2, pour essayer de valider les quelques pistes précédentes et construire des indicateurs associés à des phases de jeu pertinentes pour l'étude.

2. Modèle proposé

Dans cette section, nous présentons la structuration hiérarchique des critères et heuristiques d'évaluation existant pour les jeux vidéo, tels que présentés dans la section 1.1, dans un modèle qui permet la combinaison avec des informations provenant des signaux physiologiques. Nous distinguons deux types de critères, suivant qu'ils sont généraux et peuvent s'appliquer à tout jeu vidéo (section 2.1), ou qu'ils dépendent du type du jeu (section 2.2). Nous détaillons ensuite dans la section 2.3 leur modélisation dans le formalisme des gabarits qui fournit des outils permettant de réaliser l'évaluation quantitative d'un jeu fixé, en fonction de données fournies par des joueurs, qu'elles soient explicites, comme les réponses aux questionnaires, ou implicites, comme les signaux physiologiques.

2.1 Structuration hiérarchique des heuristiques de jeux vidéo

Les critères génériques sont ceux décrits de façon générale dans la littérature et mentionnés dans la section 1.1. Ils peuvent être appliqués à tout type de jeu vidéo et sont, comme indiqué dans la section 1.1, organisés en quatre catégories, qui correspondent aux quatre aspects principaux qui doivent être pris en compte pour l'évaluation des jeux. Nous avons proposé de compiler ces heuristiques en une organisation précise et bien classée basée sur ces catégories, représentée sur la figure 2, et détaillée ci-dessous.

La procédure de construction de l'organisation que nous avons utilisée est illustrée par l'exemple suivant : parmi les heuristiques classiques de la catégorie *game mechanics*, on trouve "*use real world formula*" [7], "*mechanics should feel natural and have correct weight and momentum*" [8], "*the world is going on whether your character is there or not*" [5]. Ces différents critères se réfèrent à la cohérence du monde dans le jeu. Aussi, nous avons introduit une sous-

catégorie nommée "*consistent world*" et les deux critères de base "*the world is going on whether your character is there or not*" et "*natural mechanics, real formulas*". Ces critères contribuent à définir et à enrichir la catégorie introduite.

La compilation globale que nous avons obtenue est présentée sur la figure 2. Elle regroupe plus de soixante-dix critères, en dix-sept sous-catégories et quatre catégories, la figure 3 détaille l'organisation de la catégorie *game mechanics*. On peut observer par exemple que la catégorie *game story* a été décomposée en quatre sous-catégories, *story*, *outcome*, *characters* et *pace*. On peut noter aussi que la catégorie *pace* fait également partie de la catégorie *game play*. Chaque point listé dans une sous-catégorie est une règle (heuristique) à appliquer lors de la conception d'un jeu vidéo.

Les heuristiques pouvant être interprétées comme des relations entre deux catégories sont numérotées et symbolisées par des flèches. Par exemple dans la sous-catégorie *learning*, l'heuristique "*discovering the story is part of the game play*" a pour numéro 2 et est un lien entre la sous-catégorie *learning* et la catégorie *game story* (flèche numéro 2).

Le formalisme des gabarits présenté dans la section 2.3 inclut des outils d'agrégation propres à exploiter cette représentation.

2.2 Spécification du modèle pour chaque type de jeu

En plus des critères précédents, des critères spécifiques doivent être pris en compte, afin d'adapter le processus d'évaluation aux différents types de jeux. Ces critères caractérisent un type de jeu et doivent être vérifiés pour que le type soit reconnu. Ainsi, dans le cas d'un FPS, la réactivité des actions du joueur doit être élevée, et la phase d'apprentissage des contrôles doit être courte. Ces deux critères sont propres aux FPS, et sont amenés à jouer un rôle majeur dans l'évaluation des FPS.

Dans ce qui suit, nous détaillons ces critères spécifiques pour chacun des quatre types de jeux, basés en particulier sur des analyses de site internet d'évaluation de jeux vidéo [17, 18, 19].

2.2.1 Les jeux de tir, First Person Shooter (FPS)

En ce qui concerne les jeux de tir, on peut tout d'abord remarquer que les testeurs, dans leur notation, dégagent cinq points essentiels dans la majorité des sites [17, 18] : les graphismes, la jouabilité, la bande son, la durée de vie et le scénario.



Figure 2 : Organisation hiérarchique des critères et heuristiques de la littérature proposée.

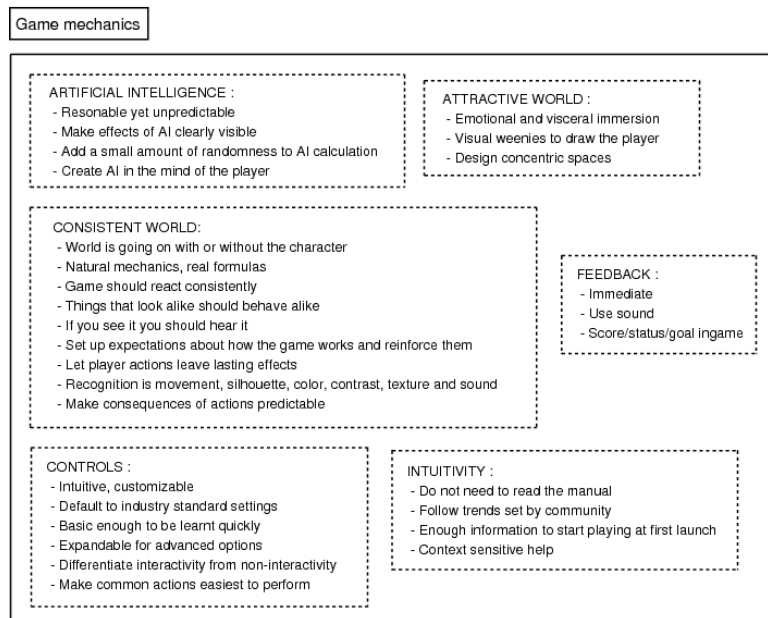


Figure 3 : Organisation hiérarchique des critères et heuristiques de la littérature pour l'évaluation de *game mechanics*.

Les testeurs attachent une très grande importance au réalisme : ils souhaitent que le jeu se rapproche le plus de la « réalité » (au sens de l'immersion du joueur dans un univers crédible : certains jeux prennent place dans des mondes fantaisistes qui n'offrent que peu d'éléments de comparaison avec la « réalité »). Ici, plusieurs facteurs sont pris en compte :

- les graphismes : les détails sont importants («La poussière virevolte, les explosions tout comme la fumée sont impressionnantes » [17]), mais aussi la prise en compte dans le jeu du cycle jour / nuit,
- une sensation de liberté : le parcours ne doit pas être linéaire et doit pouvoir permettre aux joueurs d'emprunter plusieurs voies possibles [18],
- la bande son quant à elle n'est pas un facteur essentiel mais peut améliorer sensiblement le ressenti et l'immersion des joueurs lorsqu'elle est réussie (par exemple le lancement de musique lors d'une phase d'action intense).

Un autre point crucial de l'évaluation de ce type de jeux est la jouabilité : elle doit à la fois être originale et proposer une prise en main rapide. Il est également bienvenu d'intégrer des éléments de jouabilité issus d'autres types de jeu : « une aventure riche en éléments issus du jeu de rôle (quêtes principales, secondaires et annexes, inventaire détaillé et même quelques dialogues à choix multiples) » [19].

Un autre facteur important qui concourt à la richesse de l'expérience vidéoludique et à l'immersion du joueur est l'intelligence artificielle des personnages contrôlés par le jeu. Ce facteur est crucial dans le cadre des campagnes « solo » dans lesquelles un seul joueur humain intervient. Cette importance est moindre lors de parties en réseau où plusieurs joueurs humains sont connectés en même temps et peuvent donc combler les déficiences du jeu à ce niveau.

D'autres points doivent finalement être pris en compte : l'ergonomie (interface accessible, pas de problème pour interagir avec le jeu et réaliser les actions complexes), le scénario qui doit présenter une histoire suffisamment claire pour que le joueur puisse définir ses objectifs et la durée de vie, qui n'est préjudiciable que si le jeu peut être terminé en moins de 5 heures (du fait du prix de vente des

jeux).

2.2.2 Les jeux de sport

Dans les jeux de sport, les critères sont principalement les mêmes que ceux des FPS. Les experts mettent toutefois en avant deux catégories bien distinctes de jeu de sport : les simulations et les jeux d'arcades.

Les simulations misent tout sur le réalisme, quitte à paraître rebutant pour les néophytes : « gameplay difficile au départ mais exceptionnel à l'arrivée » [17]. Dans ce cas, la profondeur du jeu est un critère important et on cherche à retrouver l'ensemble des gestes sportifs possibles de la discipline, la prise en compte des conditions météorologiques si besoin, la simulation des interactions avec l'environnement (déformations des voitures après les chocs, pistes glissantes après la pluie dans les jeux de courses automobiles).

Les graphismes et la bande son doivent être soignés pour favoriser l'immersion du joueur (par exemple les commentateurs dans une simulation de football) et peuvent défavoriser le jeu en cas d'erreur (par exemple dans Pro Evolution Soccer 6, jeu de football « les commentaires de Christian Jean-pierre et de Jean-Luc Arribart sont toujours aussi inaudibles » [18]).

Les jeux de types arcade favorisent le plaisir de jeu en recherchant en premier lieu le divertissement et une prise en main intuitive même si cela n'est pas réaliste (au sens d'une simulation).

La jouabilité doit être la plus originale possible (quitte à ne pas être la plus précise), ce qui peut être illustré par la console Wii de Nintendo. Celle-ci utilise une manette intégrant un gyroscope et un accéléromètre qui permet d'interagir avec des gestes plus que par des successions de frappes sur les boutons d'une manette conventionnelle. Cette recherche constante d'originalité doit être bien dosée de façon à conserver une prise en main aisée et la moins approximative possible (Nintendo va bientôt proposer à cet effet de nouvelles manettes plus précises, ce qui montre que le plaisir ressenti sur un jeu peut également être limité par les périphériques utilisés).

Les graphismes doivent également être soignés, mais adoptent plus généralement un style visuel « cartoon » plus être plus en phase avec le type de jeu.

Enfin, pour ces deux types de jeux (simulation / arcade), le scénario joue une place très peu importante (voire inexistante dans les simulations) et la durée de vie est dépendante du nombre de modes de jeu proposés.

2.2.3 Les jeux de stratégie, Real Time Strategy (RTS)

Les RTS sont des jeux de stratégies temps-réel qui opposent plusieurs factions dont l'une est contrôlée par le joueur. On distingue principalement deux modes de jeu : en « solo », un joueur est confronté à des troupes dirigées par le jeu. Dans ce cas, l'IA est prépondérante dans l'évaluation du jeu. A l'inverse et comme pour les FPS, en mode réseau, plusieurs joueurs s'affrontent et dans ce cas, l'IA n'est plus un critère prépondérant.

Comme dans tous les autres jeux, les graphismes sont importants pour aider à l'immersion et à la lisibilité des batailles entre factions. Les unités doivent être immédiatement identifiables visuellement de façon à ne pas créer de frustration chez les joueurs qui pourraient commettre des erreurs.

D'un point de vue *game play*, chaque faction doit posséder également ses propres caractéristiques et

autoriser des styles de jeu bien différents (centrés sur l'attaque, la défense, l'utilisation du terrain ...).

L'interface du jeu doit être la plus simple possible et la moins intrusive visuellement, pour laisser plus de place à la visualisation des troupes. L'ergonomie doit permettre une prise en main suffisamment rapide, pour ne pas rebuter les débutants. Pour ce faire, des tutoriaux sont généralement proposés lors des premières missions pour former le joueur à l'interface. L'interface doit également comporter des raccourcis claviers, qui permettent aux joueurs plus expérimentés de gagner en efficacité (par rapport aux joueurs débutants notamment).

Le scénario est, comme l'IA, prépondérant en mode « solo ». Un mauvais scénario ou une histoire trop courte pénaliseront le jeu. D'autres rebondissements sont appréciables, comme le fait de jouer les différentes factions au cours du jeu.

Enfin, comme dans les autres types de jeu, la bande son n'est pas un élément prépondérant mais concourt à faire d'un jeu une incontestable réussite quand elle est bonne.

2.2.4 Les Role Playing Game (RPG)

Les RPG permettent au joueur d'incarner un ou plusieurs personnages impliqués dans une histoire composée de multiples quêtes (héroïques). Dans ce cadre, le scénario doit être très riche (même si l'amorce de l'histoire peut ne pas l'être : par exemple « sauver la princesse ») et comporter de nombreuses quêtes qui relancent l'intérêt pour le jeu. Le scénario ne doit pas être linéaire et laisser au joueur la liberté d'explorer les lieux dans l'ordre qu'il souhaite. Les RPG peuvent également proposer plusieurs fins possibles conditionnées par les choix réalisés par le joueur tout au long du jeu.

Cette liberté doit se retrouver dans tous les aspects du jeu et notamment lors de la création de son personnage. Deux écoles sont possibles :

- la création d'un personnage complètement unique pour chacun des joueurs (choix du « métier », des attributs (force, intelligence ...) des aptitudes (magie ...), du style visuel de son avatar),
- le choix entre plusieurs personnages (généralement tous jouables) et proposant dans ce cas un style visuel très travaillé, une personnalité attachante et une jouabilité propre, de façon à permettre au maximum de joueurs de s'identifier.

La liberté de choix apparaît également dans les discussions avec les personnages non-joueurs (PNJ) dirigés par l'IA du jeu qui laissent généralement plusieurs choix en fonction de la personnalité du joueur.

L'IA est très importante également pour rendre le monde crédible et permettre des interactions complexes entre le joueur et les PNJ (et non uniquement des affrontements).

D'un point de vue technique, les graphismes doivent être soignés pour faciliter l'immersion et offrir une grande profondeur de champ pour permettre l'évasion apportée par ce type de jeu, de même que la musique qui ne doit pas être ennuyeuse ou répétitive pour ne pas lasser les joueurs. Il en va de même pour les voies digitalisées utilisées durant les discussions qui doivent être crédibles et réussies pour ne pas pénaliser le jeu.

L'interface du jeu doit également être travaillée de façon à permettre une gestion aisée de son personnage et de ses différents attributs et objets transportés.

Enfin, un RPG se doit d'avoir une durée de vie relativement longue en comparaison d'autres types de jeu (au moins 20-30 heures) et permettre éventuellement de reprendre l'aventure avec un autre personnage pour la vivre différemment ou obtenir une autre fin.

2.3 Modèle d'évaluation automatique : formalisme des gabarits

A partir de ces listes de critères, génériques et spécifiques, nous avons proposé de modéliser leur structure hiérarchique¹⁵ dans le formalisme des gabarits. Ce formalisme, proposé initialement comme un outil de reconnaissance de scénarios [3, 12, 13], fournit en effet à la fois une organisation hiérarchique de critères contribuant à un phénomène complexe, et des outils d'agrégation permettant d'évaluer un degré avec lequel le phénomène global est reconnu. Dans le cas des jeux vidéo, il s'agit de chercher à reconnaître le concept de « jeu vidéo de grande qualité », le degré de reconnaissance permettant d'obtenir une évaluation de la qualité du jeu.

Le principe général des gabarits consiste à décomposer un phénomène complexe, que l'on souhaite reconnaître, en éléments moins complexes, eux-mêmes décomposés, dans une procédure récursive, jusqu'à ce que des phénomènes élémentaires, qui peuvent être observés directement des données, soient atteints. Ce modèle est donc naturellement basé sur une structure d'arbre qui offre une représentation des données selon différents niveaux d'abstraction. La racine du gabarit est le scénario général à reconnaître (dans le cas des jeux vidéo, la notion « jeu vidéo de grande qualité ») et les feuilles sont des événements basiques concrets facilement identifiables (correspondant aux critères de base pour les jeux vidéo). Les nœuds internes correspondent à des niveaux d'abstraction intermédiaires.

Cette structuration correspond à la problématique des jeux vidéo : le principe consiste à extraire, des questionnaires ou des signaux physiologiques, des informations de base et peu complexes sur les jeux évalués, et à combiner et agréger ces informations pour en déduire une évaluation globale. Ainsi, les données issues des questionnaires doivent permettre de vérifier certaines heuristiques élémentaires exposées précédemment : on demandera par exemple au joueur s'il trouve les contrôles complexes (*"controls should be intuitive"*), si les buts sont nombreux (*"multiple goals"*) ou s'il s'intéresse aux personnages (*"player is interested in the characters"*). L'agrégation de certaines de ces données par le gabarit permet de reconnaître automatiquement des aspects propres à certains types de jeux et plus abstraits (comme par exemple facilité de prise en main, importance de l'histoire, liberté du joueur) et contribue à l'évaluation de leur qualité globale.

Outre la possibilité de modéliser naturellement la structure hiérarchique des critères et heuristiques, le formalisme des gabarits offre donc un mécanisme d'évaluation qui permet de propager, et de faire remonter vers la racine de l'organisation, les évaluations de ces critères élémentaires. Il offre par là une méthode d'évaluation interprétable et proche du raisonnement des experts. De plus, il permet de prendre en compte des informations hétérogènes, et de fusionner des informations de types différents comme les informations issues de questionnaires et les signaux physiologiques interprétés comme des indices pour les feuilles de l'arbre.

Dans ce qui suit, nous présentons avec plus de détails le formalisme des gabarits et soulignons les avantages qu'il présente pour l'évaluation des jeux vidéo. Enfin nous présentons une extension à un modèle qui généralise la structure arborescente à une structure de type treillis.

¹⁵ Ce travail a donné lieu à une publication : Automated video games evaluation based on the template formalism François Nel, Marc Damez, Nicolas Labroche and Marie-Jeanne Lesot, Proc. of the International Conference on Information Processing and Management of Uncertainty (IPMU) 2008.

2.3.1 Présentation des gabarits

Nous rappelons ici brièvement la méthode d'évaluation des gabarits (pour plus de détails voir [3, 12, 13]). A titre d'exemple, nous considérons la figure 2 qui illustre un gabarit incomplet dans le cas des jeux vidéo : il fait dépendre la reconnaissance de la qualité du réalisme de celle de la qualité graphique du jeu et de la cohérence du monde, cette dernière étant à son tour exprimée en fonction de deux critères plus simples à reconnaître.

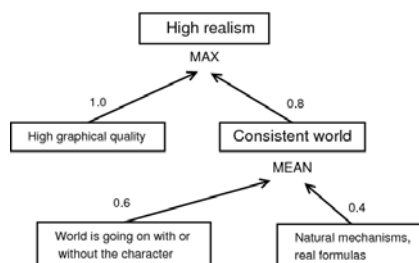


Figure 4 : Exemple de gabarit pour la reconnaissance de *high realism*.

Le mécanisme d'évaluation du gabarit, c'est-à-dire le calcul du degré de reconnaissance du phénomène global, fonctionne alors comme suit : chaque nœud de l'arbre est associé à un opérateur d'agrégation, tel que la moyenne ou le maximum par exemple. Le degré de reconnaissance du nœud est calculé comme l'agrégation des degrés de reconnaissance de ses fils. L'opérateur choisi pour l'agrégation exprime la sémantique de la dépendance : l'opérateur maximum par exemple a une sémantique de « ou logique » et indique que la reconnaissance d'un seul des fils suffit à reconnaître le phénomène père, l'opérateur moyenne permet d'exprimer un effet de compensation entre les nœuds fils. De plus, l'agrégation peut être pondérée.

Le degré de reconnaissance des feuilles, qui détermine le processus global fonctionne de la façon suivante : les feuilles du gabarit sont associées à des contraintes exprimées par des sous-ensembles flous. Pour l'exemple de la figure 2, on peut imaginer une contrainte associée à la qualité graphique du type « très bonne évaluation », qui signifie que la note donnée par l'utilisateur à ce critère doit correspondre à une « très bonne » note, entre 8 et 10 sur une échelle de 10 par exemple.

L'évaluation d'une feuille est alors définie par un degré de compatibilité (critère de satisfiabilité d'Isomoto [10] par exemple) qui mesure à quel point la contrainte est vérifiée. La propagation de ces degrés dans le gabarit jusqu'à la racine induit l'évaluation globale.

Aussi un gabarit est défini par sa structure, les opérateurs d'agrégation et les pondérations de ses nœuds, et les contraintes associées à ses feuilles.

2.3.2 Avantages du formalisme des gabarits pour l'évaluation des jeux vidéo

Comme nous l'avons souligné ci-dessus, le premier avantage des gabarits provient de leur structuration hiérarchique qui correspond parfaitement à la structure d'évaluation des jeux vidéo, telle qu'elle peut être déduite de l'étude des heuristiques existantes (cf section 2.1). Elle permet par là de traduire le raisonnement intuitif des experts, ce qui augmente son interprétabilité. Les mécanismes flous à l'œuvre dans les gabarits, dans l'expression de contraintes floues, sont également particulièrement adaptés pour atteindre ce but, de même que l'intégration des variables linguistiques : celle-ci permet à l'utilisateur de nommer les aspects positifs et négatifs d'un jeu. Aussi, l'évaluation est interprétable et peut être expliquée linguistiquement.

De plus, les gabarits constituent un outil d'agrégation qui permet de justifier une évaluation globale à partir de critères élémentaires. Leur intérêt est accru du fait de la possibilité d'analyser les nœuds qui n'ont pas été validés lors de la remontée de l'arbre : celle-ci permet d'identifier les points qui

devraient être améliorés dans le jeu pour lui permettre d'atteindre ses objectifs. L'étude de ces nœuds non validés forme le diagnostic attendu.

En outre, les gabarits permettent de fusionner des données hétérogènes et de prendre en compte des informations résultant de critères objectifs comme de critères subjectifs. En effet, l'évaluation des feuilles de l'arbre est faite en mesurant la compatibilité à une contrainte floue, qui peut porter aussi bien sur des évaluations linguistiques effectuées par les utilisateurs que sur des valeurs numériques déduites des signaux physiologiques.

Enfin, les jeux vidéo sont d'ordinaire classés en différentes catégories. Certaines, comme les FPS ou les RTS, possèdent des caractéristiques distinctes et spécifiques, qui doivent donc être évaluées différemment comme détaillé dans la section 2.2. Les mécanismes de pondération qui peuvent être utilisés dans les gabarits permettent d'intégrer ces attentes au formalisme : étant donné un gabarit général contenant toutes les heuristiques identifiées, il est possible de pondérer leur importance relative, voire d'éliminer certains critères qui ne sont pas à prendre en compte pour un type de jeu particulier. Aussi un expert artificiel pour le diagnostic des jeux vidéo peut être défini par une structure commune de gabarit modélisant la structuration hiérarchique des critères, qui est ensuite déclinée, selon les opérateurs d'agrégation et la pondération des nœuds, en fonction des différents types de jeux.

2.3.4 Extension aux treillis

De part la nature continue des données objectives recueillies pour le projet (oculométrie et physiologique), un processus d'extraction de descripteurs pour l'établissement des scores sur les critères d'évaluations du gabarits doit être effectué.

Ce processus d'extraction de descripteurs sur les données objectives permet d'obtenir de très nombreuses informations à analyser. En effet, plusieurs méthodes d'extraction peuvent être combinées à toutes les informations purement descriptives du signal : min, max, moyenne, écart-type, augmentation/diminution globale, etc. (voir 3.3.1. Formulation du problème et solutions proposées). Ces méthodes sont par exemple :

- l'extraction sur des séquences homogènes de jeux,
- l'extraction sur des séquences hétérogènes du jeu
- les signaux permettant plusieurs extractions simultanées (utilisation de certaines parties de l'écran pour les signaux oculométriques par exemple)
- les paramétrage d'extraction comme l'échantillonnage du signal continu.
- l'extraction sur différentes catégories de population de joueurs (un « hardcore gamer » n'a pas du tout le même comportement ou les mêmes sensations qu'un novice)

Chaque passation peut ainsi être représentée par plusieurs centaines de descripteurs. Pourtant, les interprétations significatives permettant d'obtenir un score sur un critère d'évaluation ne sont pas très nombreuses. Nous avons par exemple étudié des interprétations pour qualifier le rythme, le changement de rythme, l'immersion, l'ergonomie de l'interface (l'information cherchée est facilement trouvée) et l'utilisabilité (les informations que le jeu communiquent sont facilement perçues et comprises), la facilité de prise en main du jeu.

Aussi, nous avons distingué deux possibilités pour prendre en compte ces interprétations dans l'évaluation diagnostic formulé par le gabarit : remonter le score calculé sur les critères à des nœuds de hauts niveaux ou permettre à une seule évaluation issue des données objectives de pouvoir influencer sur plusieurs critères de bas niveau.

La seconde solution nous a semblé plus pertinente. En effet, elle rend l'utilisation des gabarits plus

souple et autorise l'interprétation diagnostic sur un maximum de critères. Un critère de bas niveau peut donc être influent sur plusieurs critères de haut niveau, et la structure du gabarit n'est plus arborescente mais en treillis. Elle permet des exécutions du gabarits sur des données homogènes, en gardant la même capacité d'interprétation diagnostic. En effet, si suffisamment de critère sont connectés au données objectives, l'exécution du gabarit peut se faire sans les données subjectives (et réciproquement).

Relation d'ordre R :

$xRy \Leftrightarrow \{x \text{ est fils de } y\}$

Tri dans l'ordre croissant des nœuds du gabarit par la relation R

Pour chaque nœud n

Calculer le score de n en fonction des scores des fils de n

Fin Pour

Algorithme d'exécution d'un gabarit.

3. Mise en œuvre du modèle : expérimentations

Dans cette section, nous décrivons les différentes expérimentations que nous avons réalisées pour tester les modèles proposés : une première expérience, décrite dans la section 3.1, porte sur l'exploitation des données objectives et se base sur les réponses fournies par des joueurs à des questionnaires qui alimentent les gabarits. Une seconde expérimentation, présentée dans la section 3.2, porte sur les informations objectives issues de signaux physiologiques. Enfin, la section 3.3 décrit les premiers travaux réalisés pour établir des relations entre les signaux physiologiques et les états émotionnels des joueurs.

3.1 Données objectives

L'expert flou pour le diagnostic du jeu vidéo a été évalué sur des données de questionnaires afin de valider l'approche retenue. Les résultats succinctement rappelés ci-après ont donné lieu à une publication scientifique dans la conférence internationale IPMU 2008 (Information Processing and Management of Uncertainty).

Les objectifs de l'étude étaient doubles :

- d'une part, montrer que la méthode des gabarits est apte à traiter le problème applicatif proposé,
- d'autre part valider le modèle du type de jeu construit à partir des heuristiques précédentes (et applicables, dans l'absolu, à l'ensemble des types de jeux).

L'étude a porté sur trois jeux de type FPS (jeu de tir ou First Person Shooter) sortis au moment de notre évaluation : Bioshock (Take 2 Interactive, 2007), Far Cry (Ubisoft, 2004) et Half Life 2, Episode 2 (Valve, 2007).

Il a été demandé à des joueurs d'évaluer les propriétés élémentaires définies dans le gabarit pour les FPS (cf section 2.3). Ces informations ont été mises en entrée de l'évaluateur de jeux qui a alors calculé une évaluation globale de la qualité du jeu. Les résultats obtenus sont présentés dans le tableau 1. Pour chaque jeu, le diagnostic global est détaillé par cinq notes intermédiaires *game play*, *interface*, *high pace*, *realism* et *story*. Elles représentent les nœuds qui précèdent directement la racine du gabarit des FPS.

De façon à évaluer la pertinence du modèle et plus généralement de l'approche des gabarits, nous avons comparé les résultats d'évaluation de notre gabarit et ceux d'un site spécialisé dans l'évaluation de jeux vidéo <http://www.gamerankings.com>. Les scores attribués par ce site sont

indiqués dans la dernière ligne du tableau.

On constate que les résultats fournis par les gabarits sont très proches des évaluations des experts du domaine, ce qui valide l'approche et plus spécifiquement le modèle de FPS développé dans le cadre de cette approche.

De plus, le principe hiérarchique des gabarits permet de nuancer le résultat global et d'examiner plus en détails l'évaluation des jeux : on constate que dans les trois cas, les évaluations des critères *high pace* et *realism* sont très élevées. Or ces aspects sont déterminants pour la qualité de jeux de type FPS, et ces résultats sont en adéquation avec nos attentes. Les scores pour *game play*, *interface* et *story* sont plus variables. Ils reflètent des aspects positifs et négatifs de ces jeux. En particulier la note obtenue par Bioshock pour le critère *story*, supérieure à celle de *Far Cry*, est cohérente avec les évaluations des spécialistes du domaine [20]. Aussi les gabarits permettent un diagnostic précis, au-delà de l'évaluation globale.

Critères / Jeux	Bioshock	FarCry	Half-Life 2
Game play	5.7	8.0	7.0
Interface	7.3	6.7	9.1
High pace	9.0	8.9	8.7
Realism	7.0	7.9	9.1
Story	7.3	6.0	7.0
Diagnostic global	8.5	8.4	8.9
Evaluation Gamerankings	8.3	8.6	8.9

Tableau 1 : Evaluation de trois jeux vidéo fournie par les gabarits à partir de données issues de questionnaires et comparaison avec les évaluations d'un site web spécialisé dans l'évaluation de jeux vidéo.

3.2 Données subjectives

Une seconde expérimentation a été consacrée aux informations objectives issues de signaux physiologiques. Nous présentons dans cette section le protocole expérimental ainsi que le détail des passations qui ont été réalisées au Laboratoire LUTIN, en collaboration avec le laboratoire CHART, dans le but d'observer les relations entre les signaux physiologiques enregistrables dans le cadre du projet et le ressenti ou les émotions des joueurs.

Pour ce faire, nous avons demandé à deux joueurs non expérimentés de réaliser une session de 25 minutes environ avec enregistrement :

- du signal électrodermal (conductance de la peau) (EDA ci-après),
- du signal électrocardiogramme (fréquence cardiaque) (ECG ci-après),
- du signal vidéo correspondant à ce que joueur voit à l'écran.

La séquence vidéo de la session est utilisée dans notre expérience pour aider à étiqueter et retrouver les différentes phases de jeu vécues par le joueur et pouvant avoir des implications sur les deux autres signaux.

Du fait de l'enregistrement des signaux physiologiques, chaque passation comporte deux phases principales :

- une phase où le sujet visionne un ensemble d'images relaxantes, qui permet de déterminer le niveau des signaux enregistrés au repos pour cette personne et donc de calibrer le système pour tenir compte des variations inter-individus,

- une phase de jeu où le sujet est amené à parcourir le niveau 1 du jeu Halo 3.

Durant la phase de jeu à proprement parler, plusieurs événements marquants ont eu des répercussions significatives sur les signaux physiologiques. Nous détaillons ci-dessous les constatations que l'on peut faire sur les relations entre signaux physiologiques et phases de jeu :

- durant les phases de repos, la conductance baisse pour atteindre un niveau proche de la phase d'observation des images relaxantes.
- Les phases d'exploration, c'est-à-dire lorsque le joueur déplace son personnage entre deux zones de combats, possèdent des similitudes avec la phase de repos : EDA et ECG sont proches de leur minimum observé ce qui dénote un rythme lent ou une phase de concentration diffuse.
- Durant les phases de tir, EDA et ECG sont élevés par rapport au repos ; un pic dans EDA indique soit une phase de tir, soit un changement brusque.
- Durant les phases d'échec, c'est-à-dire lorsque les personnages sont tués par les adversaires, deux cas sont à distinguer : en cas de mort prévisible, on observe une augmentation (faible) de la fréquence cardiaque puis un retour à la normale, en cas de mort imprévue, on observe une baisse minime de la conductance. Sur un large panel de joueurs le nombre de décès pourrait également servir d'indicateur sur le niveau du joueur.
- Durant les phases d'égarement, c'est-à-dire lorsque le joueur est perdu dans le niveau et s'énerve, comme pour les phases de tir, on observe des moyennes plus élevées des signaux EDA (très marqué) et ECG (moins marqué).
- Enfin, lors des phases « hors d'atteinte », c'est-à-dire lorsque le joueur est dans un milieu protégé (milieu aquatique), les signaux reviennent à un état proche du repos.

Le tableau suivant présente un résumé de ces premières constatations. Les cas pour lesquels l'observation n'a pas permis de déterminer l'état émotionnel sont marqués d'un symbole « ? ».

ECG // EDA	Faible	Normal	Fort
Faible	Repos / rythme lent	Rythme lent	?
Normal	Lassitude	-	Concentration accrue / stress (pics EDA)
Fort	?	Énervement	Rythme rapide

Tableau 2 : Caractérisation de l'état du joueur en fonction du niveau des signaux ECG et EDA

Cette courte expérimentation montre qu'il est possible de caractériser, même sommairement, l'état du joueur à partir de quelques signaux physiologiques comme l'électrodermal (EDA) et l'électrocardiogramme (ECG). Il est donc possible d'intégrer des informations extraites des signaux physiologiques pour renseigner certaines heuristiques de bas niveau des gabarits en lieu et place des questionnaires utilisateurs.

3.3 Fouille de données dans les signaux physiologiques

Comme nous l'avons indiqué dans les sections précédentes, nous avons à notre disposition lors du projet Gamelab plusieurs types de données : les données subjectives issues de questionnaires et des données objectives idéalement extraites à partir des signaux physiologiques. Comme déjà mentionné, il n'existe pas encore de modèle général et reconnu indiquant quel type de signal enregistrer et comment l'analyser pour apprendre l'état émotionnel d'une personne, son intérêt, son plaisir à effectuer une tâche.

Nous présentons donc dans ce paragraphe les premiers travaux conduits par le LIP6 pour franchir ce fossé sémantique entre la représentation des signaux physiologiques, leur interprétation et leur

pouvoir discriminant dans le cadre du diagnostic d'un jeu vidéo.

L'objectif de l'étude présentée ci-après s'attache à l'analyse exclusive des traces oculaires, car nous avons à disposition des passations complètes sur deux jeux de type RTS (Real Time Strategy) nommés respectivement *Dark Crusade* et *Perimeter*.

3.3.1. Formulation du problème et solutions proposées

La première difficulté rencontrée lors de cette étude est que la majorité des travaux sur les traces oculaires ne s'intéressent qu'à des interfaces fixes qui n'évoluent pas dans le temps ou bien à la manière dont un humain appréhende une photographie ou un texte statique. Le travail à réaliser est, dans notre cas, bien plus complexe car il implique l'analyse de séquences de jeux vidéo affichant plusieurs dizaines d'images par secondes. La seconde difficulté, toujours liée à cet aspect dynamique d'un jeu vidéo - par opposition aux études plus traditionnelles des traces oculaires, est que la signification même de ce que le joueur observe change au cours du temps car un jeu est composé de phases : préparation / introduction, exploration du terrain de jeu, combats, gestion des ressources (munitions, soldats selon le type de jeu). Enfin, la liberté offerte aux joueurs fait que deux joueurs différents n'ont pas nécessairement les mêmes phases de jeu durant leur passation ou tout du moins pas dans le même ordre ni avec la même importance (car cela dépend de la manière dont la personne joue).

Pour résoudre ces problèmes, nous avons donc proposé de découper les passations de jeu en séquences et pour chacune d'entre elle d'associer une ou plusieurs étiquettes en fonction des phases de jeu observées ou de l'état du joueur, telles que exploration ou recherche d'un objet ou bien encore combat, mort, ou stress. Pour cela, nous avons utilisé le logiciel GamEval, développé par la société IObjects dans le cadre du Sous-Projet 3 mené par le laboratoire CHART. De cette façon, il devient possible de résumer l'information et de passer d'un flot continu de données (vidéo) à un ensemble de « clichés » reflétant l'état du joueur à un instant donné en fonction de ces objectifs vidéoludiques. Chaque séquence ainsi étiquetée est représentée par un ensemble de descripteurs issus des signaux oculaires comme la durée des fixations sur cette période, la fréquence des saccades, la longueur du scan path.

Il devient alors possible de réintégrer ces informations dans le modèle de gabarit selon trois modalités :

- On peut comparer des séquences ayant la même étiquette entre deux jeux différents, un bon et un mauvais, pour essayer de les différencier et d'apprendre des descripteurs physiologiques qui reflètent qu'un jeu est bon ou non. C'est ce type d'approche que nous avons retenu dans nos travaux car nous avons à disposition des passations pour les jeux *Dark Crusade* et *Perimeter*, différemment appréciés par les critiques de jeux vidéos.
- On peut caractériser un des aspects d'un jeu vidéo par la comparaison des séquences observées au sein d'une même passation. Ainsi le rythme d'un jeu vidéo peut être déduit de l'alternance entre des séquences durant lesquelles les rythmes cardiaques et la conductance de la peau soit augmentent fortement soit diminuent fortement.
- On peut caractériser le niveau d'expertise des joueurs en comparant des séquences possédant les mêmes étiquettes (même phase de jeu par exemple) sur le même jeu. Des expériences ont été conduites au sein du projet Gamelab par le laboratoire CHART de façon à différencier les experts des novices.

3.3.2. Construction du jeu de données

Les passations ont été réalisées sur deux jeux de stratégie temps réel (RTS) nommés *Dark Crusade* et *Perimeter* en collaboration avec le laboratoire CHART. Les passations ont été découpées en séquences en fonction des phases de jeu observées. Pour chacune des séquences de une à plusieurs annotations ont été proposées pour chacune des zones d'intérêt formant l'interface utilisateur.

Ensuite, chacune des annotations est représentée par un ensemble de descripteurs statistiques simples calculés à partir des traces oculaires. Lorsque des annotations apparaissent plusieurs fois au cours de la même passation, les valeurs des descripteurs sont calculées comme la moyenne des valeurs des descripteurs sur l'ensemble de ces annotations.

Au total, cette représentation par annotation utilise plus de 1000 descripteurs statistiques (ou attributs) pour représenter une passation.

3.3.3. Expérimentations et résultats

Un premier test a été conduit à partir des données précédentes en utilisant l'algorithme d'apprentissage ID3. Celui-ci travaille à partir d'un ensemble de données étiquetées et construit un arbre de décision dont les nœuds en partant de la racine correspondent aux attributs les plus discriminants dans le processus d'apprentissage (au sens de l'entropie de Shannon) et dont les feuilles correspondent à une décision. Dans notre cas, la décision revient à différencier un bon jeu d'un mauvais jeu.

Les résultats de ce premier test n'ont pas été bons : 62 % de reconnaissance d'un bon jeu par rapport à un mauvais. Cela s'explique par le trop grand nombre d'attributs comparé aux nombres d'exemples pour apprendre et du fait du domaine de représentation des attributs trop grand également.

Un second test a été réalisé avec l'algorithme C4.5 qui est une amélioration de l'algorithme ID3 travaillant sur des données symboliques. Afin de l'utiliser sur nos données de passations, une discrétisation des valeurs numériques a été réalisée, ce qui a réduit considérablement les domaines de définition des attributs. Nous avons ensuite appliqué une analyse en composante principale (ACP) qui conduit à la définition de 28 attributs au pouvoir discriminant accru qui améliorent les performances de notre méthode. En revanche la combinaison des attributs initiaux interdit désormais toute interprétation simple des règles apprises par l'algorithme d'apprentissage C4.5. Les résultats obtenus sont bons puisque le taux de bonne classification s'élève à 86%.

3.3.4. Conclusion

Ce paragraphe a décrit les travaux visant à analyser les signaux physiologiques pour en tirer des informations objectives qui puissent être intégrées dans les gabarits pour compléter ou remplacer les données subjectives issues des questionnaires. Les premiers résultats obtenus après filtrage et réduction du nombre des attributs décrivant des passations oculaires sont très bons et permettent la discrimination dans plus de 86% des cas entre un bon et un mauvais jeu (à partir des données oculaires à disposition). Ce premier résultat est prometteur et ce type d'étude pourrait donc être reproduite sur d'autres types de signaux et dans d'autres contextes pour nourrir le gabarit dès lors que des modèles liant signaux physiologiques et émotions ou ressentis des joueurs auront été validés par la communauté.

4. Réalisations logicielles

Cette partie présente les différents logiciels qui ont été développés au LIP6 dans le cadre du projet Lutin Gamelab.

4.1 GamelabQuest

La plateforme GamelabQuest de récolte des données subjectives a été développée pour la tâche 2 du sous projet. GamelabQuest est une interface web permettant la création et la gestion de questionnaires en ligne. Ses principales fonctionnalités sont la création de questionnaires (sous la forme d'un arbre XML dans un format standard), l'affichage d'un questionnaire pour la saisie des réponses, la recherche d'un questionnaire précis et un accès personnalisé aux différentes

fonctionnalités selon les utilisateurs.

4.1.1 Structure des questionnaires

Les questionnaires de GamelabQuest sont structurés dans le format décrit par IMS Global dans le standard QTI et qui fait office de norme dans le domaine. Plusieurs structures de questionnaires simplifiés ont été développées. Ils respectent tous le standard QTI et permettent notamment aux développeurs de rendre accessible un certain nombre de types de question, qui respectent le standard QTI, tout en continuant de développer d'autres types de question plus complexe. L'un des intérêts d'utiliser le format IMS QTI dans GameLabQuest réside dans la possibilité offerte de communiquer des questionnaires avec d'autres applications (notamment l'application EvalGen développée également dans le cadre de ce projet, voir section 4.2) dans un format standard. Cette utilisation a donc pour but de garantir la pérennité (car IMS QTI est un standard et cela depuis 1999), la compatibilité (entre les fichiers provenant de GamelabQuest, d'EvalGen et plus largement tous les fichiers respectant le format IMS QTI), l'interopérabilité (entre les différentes versions du logiciel : la création des questionnaires ne dépend plus du logiciel mais de l'évolution du standard), et la fiabilité (IMS QTI garantit le fait d'avoir un questionnaire bien conçu).

Plusieurs types de réponses peuvent être apportées :

- « Choice » correspond à l'ensemble des questions à choix multiples classiques,
- « Slider » correspond à une réponse sous forme de « slider », une échelle horizontale sur laquelle on peut faire glisser un curseur pour moduler sa réponse,
- « Extended Text » dans lequel la réponse prend la forme d'un texte, ce qui permet les questions ouvertes,
- « Match » correspondant à un tableau de réponse, très souvent représenté dans les questionnaires.

Afin d'exploiter quelques améliorations dans les QCM classiques proposées par des chercheurs du LIP6 [6], un nouveau type de questionnaire a été introduit de façon à pouvoir prendre en compte l'incertitude des réponses des utilisateurs.

4.1.2 Les QCM évidentiels

Les QCM évidentiels proposent une généralisation des QCM classiques pour permettre à la personne interrogée d'exprimer plus finement sa réponse : les QCM classiques ne laissent pas la place au doute ou à l'indécision, ils exigent une réponse précise et certaine. Les QCM évidentiels introduits par Diaz et al. Dans [6] permettent à la personne interrogée d'exprimer un doute, une incertitude, une ignorance tout en répondant à la question. Les notions de doute, de préférence des réponses (ex : 80% de A et 20% de B), ou d'ignorance sont des informations que les QCM classiques sont incapables de prendre en charge. Les QCM évidentiels introduisent le moyen de pallier ce manque et permettent à la personne interrogée d'exprimer plus précisément son opinion. Pour cela, ils font appel à la notion de masse de croyance attribuée aux réponses.

L'hypothèse est que l'utilisateur possède initialement une certaine masse de croyance. Lors de sa réponse, il affecte tout ou partie de sa masse de croyance à une ou plusieurs réponses proposées par le QCM. Il peut partager sa masse de croyance en plusieurs réponses et n'est aucunement obligé de répartir toute sa masse de croyance. Ainsi, le fait de partager sa masse de croyance apporte des informations sur le doute, et le fait de ne pas répartir toute sa masse de croyance reflète son manque de confiance ou son ignorance dans la question.

La figure ci-dessous présente l'affichage d'un QCM évidentiel dans GameLabQuest. La première ligne grise représente la zone de croyance à répartir. Cette zone est partagée en deux. Tout d'abord, la partie qui contient la croyance déjà distribuée (zone c), et la partie restante à distribuer (zone u). On peut remarquer que la personne qui répond aux questionnaires a partagé sa croyance sur deux

questions. Elle pense que la bonne réponse est la réponse A, mais a un petit doute avec la réponse B. On peut remarquer aussi qu'elle n'est pas convaincue de sa réponse, car elle n'a pas distribué toute sa masse de croyance (zone u). Par contre, elle est sûre que les réponses C, D et E sont des mauvaises réponses.



Figure 5 : Une question d'un QCM évidentiel

4.1.3 Utilisateurs de GamelabQuest

Comme il a été mentionné précédemment, GamelabQuest autorise la gestion de différents utilisateurs dont les droits varient. Nous les présentons rapidement ci-dessous :

- Administrateur : il a tous les droits. Il crée les comptes utilisateurs et leur affecte un rôle (Chef de Projet, Utilisateur, Invité). Il peut modifier les droits de tous les utilisateurs. Il peut aussi créer des questionnaires.
- Chef de Projet : il a pour rôle de créer des questionnaires. Il peut aussi créer des comptes utilisateurs.
- Utilisateur : il a pour rôle de répondre aux questionnaires. Lors de la création d'un nouvel utilisateur, le chef de projet remplit un formulaire de renseignement sur cette personne. Les chefs de projet peuvent donc solliciter les utilisateurs de la base de données afin qu'ils répondent aux questionnaires. Lorsque le questionnaire est rempli, les réponses de l'utilisateur sont enregistrées en attente d'être récupérées ultérieurement (traitement des données, analyse...).
- Invité : il ne peut accéder aux questionnaires que par le biais d'une URL (un seul questionnaire par URL) qui peut être déposé sur un site ou un forum, ou être envoyé par e-mail. En cliquant sur ce lien, l'invité arrive sur une page de connexion. Lors de sa première participation, il doit s'inscrire en se créant un compte. Il lui suffit ultérieurement de se connecter avec son identifiant. Le compte invité est temporaire, et lorsque le questionnaire a été rempli, l'invité n'a plus accès au questionnaire.

4.1.4 Modules fonctionnels de l'application GamelabQuest


- Module de création et d'édition des questionnaires

Lors de l'ouverture de ce module, le chef de projet (seul habilité à créer des questionnaires avec l'administrateur) dispose d'une arborescence de fichiers représentant un espace personnel où il sauvegarde ses questionnaires, la liste de ses questionnaires partagés, et la liste des questionnaires qu'il a créés. Selon ses besoins, il peut modifier l'arborescence de son espace personnel ou créer, déplacer, éditer, modifier, supprimer ou partager ses questionnaires avec d'autres utilisateurs. Un lien lui permet d'accéder à l'éditeur de questions présenté dans la figure ci-dessous.

Ajouter

Prénom :

Nom :

Date de naissance : 

Email :

Niveau :

Rôle :

- Chef de Projet
- Default
- Invité

Pseudo :

Mot de passe :

Retapez le mot de passe :

Figure 9 : Formulaire de création des utilisateurs

La partie de gestion des rôles permet de gérer les différents rôles des utilisateurs comme l'illustre la figure ci-dessous.

Ajouter

Nom du rôle :

Editer

Nom du rôle	Supprimer
Chef de Projet	<input type="checkbox"/>
Default	<input type="checkbox"/>
Invité	<input type="checkbox"/>
Utilisateur	<input type="checkbox"/>

Figure 10 : Gestion des rôles des utilisateurs

La partie de gestion des structures des questionnaires permet d'importer les schémas de données au format XSD (XML Schema Definition) utilisés pour la création des questionnaires ainsi que pour l'import et l'export des données.

Ajouter & Editer - Binding

Ajouter

URL :

ou

Chemin local : (.xsd, .zip)

Editer

URL	Supprimer
http://colargol:8080/Gamelab/xsd/qtifrsimplifie.xsd	<input type="checkbox"/>
http://colargol:8080/Gamelab/xsd/qtifrsimplifie2.xsd	<input type="checkbox"/>
http://colargol:8080/Gamelab/xsd/qtifrsimplifie3.xsd	<input type="checkbox"/>
http://colargol:8080/Gamelab/xsd/qtivosimplifie.xsd	<input type="checkbox"/>

Figure 11 : Gestion des XSD structurant le questionnaire

4.2 EvalGen

EvalGen est une application permettant la création de questionnaires, ainsi que la récolte des réponses aux questionnaires, pour évaluer tout type d'objets (au sens large : une voiture, un jeu

vidéo, un tissu, un projet de loi ...). Il représente la version sur poste fixe (ou « standalone ») du logiciel précédent GamelabQuest. Nous présentons ici rapidement ses fonctionnalités.

EvalGen travaille à partir des entités conceptuelles suivantes :

- Les projets : un projet peut être comparé à un répertoire partagé avec des fichiers, des utilisateurs et leurs droits d'accès. Chaque projet possède des objets à évaluer, des évaluateurs, des droits d'accès à ces objets à évaluer ainsi que la possibilité d'ajout, de modification et de suppression d'un évaluateur ou d'un objet à évaluer. Comme dans GameLabQuest la création, la suppression et la modification sont limitées à certains utilisateurs.
- Les utilisateurs : les utilisateurs sont ici les évaluateurs dont le rôle est de répondre aux questionnaires.
- Les groupes permettent de regrouper les utilisateurs dès lors qu'ils partagent un certain nombre de caractéristiques communes. On peut par exemple, en fonction des cas d'application, définir les groupes « spécialistes textiles », « enseignants » ou encore « ergonomes ».
- Les objets à évaluer sont de nature diverse et cette notion peut aussi bien recouvrir un logiciel, une fonctionnalité, un produit ou une personne.
- Les critères sont les caractères ou les signes particuliers qui permettent de distinguer un objet et qui autorisent un utilisateur à formuler un jugement d'appréciation. Un critère peut être décrit par une structure hiérarchisée composée de sous-critères. Les critères sont les entités qui sont évaluées dans les réponses aux questionnaires. Un système de gestion de référentiels de critères a été mis en place de façon à pouvoir réutiliser les critères d'un référentiel déjà existant ou bien en créer de nouveaux pour l'objet de l'étude.
- Échelle d'appréciation : l'évaluateur peut attribuer une note à un critère d'évaluation à partir d'une échelle d'appréciation. Cette échelle peut être numérique ou symbolique (par exemple par des adjectifs tels que « bon » ou « mauvais »). La gestion de l'échelle d'évaluation est souple et paramétrable : ainsi, il est possible de déterminer, pour une note comprise entre 3 et 8, une appréciation symbolique équivalente étiquetée, par exemple, « Médiocre ». Ces échelles d'appréciations utilisent la notion de sous-ensembles flous et sont appelées Fuzzy Partitions.

Nous présentons ci-dessous les principales fonctionnalités du logiciel EvalGen illustrées de captures d'écran de l'interface.

L'arbre de navigation permet de naviguer simplement à l'intérieur des projets, et ainsi de facilement retrouver un objet à évaluer ou un utilisateur. Il est présent dans la partie gauche de l'application.



Figure 12 : Arbre de navigation

Les tableaux des entités conceptuelles permettent d'accéder à toutes les informations importantes. Ces tableaux ont été utilisés dans plusieurs interfaces de l'application, et servent à lister ce que possède l'entité sélectionnée dans l'arbre de navigation. Par exemple dans la fenêtre donnant toutes les informations relatives à un groupe d'utilisateurs, on trouve la liste des utilisateurs sous forme d'un tableau. L'utilisateur a la possibilité de cliquer sur l'une des lignes du tableau pour obtenir toutes les informations disponibles sur l'utilisateur.

Nom	Date de création
Dupond jean	2007/5/18-14:24:56
Dupont toto	2007/5/18-10:58:52

Figure 13 : Tableau des entités conceptuelles

Une fenêtre d'édition s'ouvre lorsque l'utilisateur réalise un clic souris droit sur l'un des éléments de l'arbre de navigation. Un menu contenant des possibilités d'actions différentes selon l'entité sur laquelle l'utilisateur a cliqué s'ouvre. Ainsi, appliqué à un projet, un clic droit permet de modifier le projet, de lui ajouter un groupe d'utilisateurs ou un objet à évaluer au projet ou encore de supprimer ce projet. Appliqué à un groupe d'utilisateurs il permet de modifier ce groupe, de lui ajouter un utilisateur ou de le supprimer.

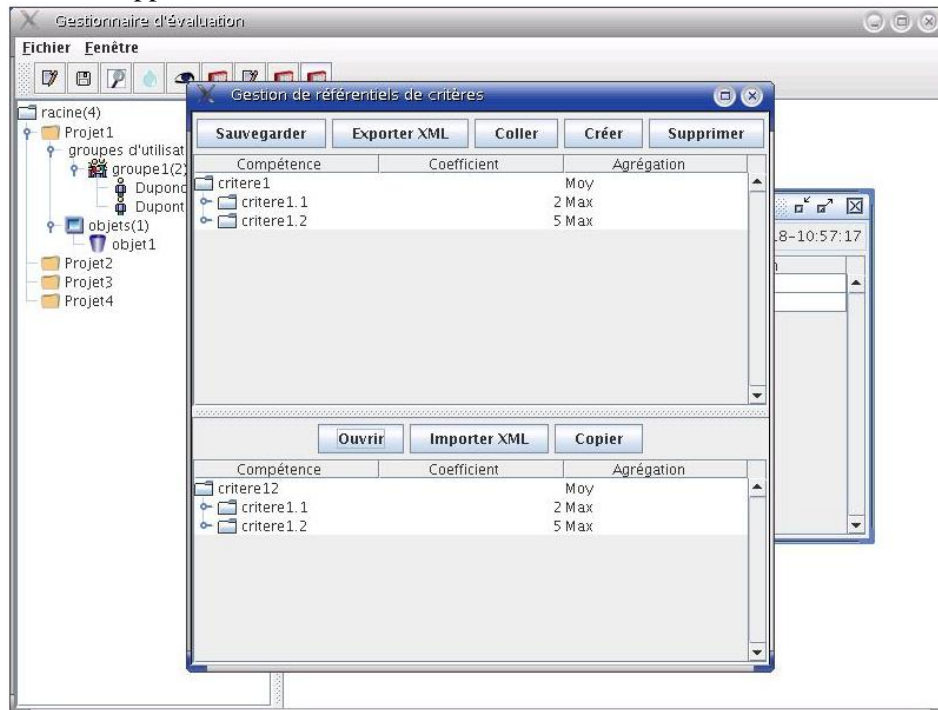


Figure 14 : Option de gestion de référentiels de critères

4.3 Plugin GamEval

Un plugin pour l'extraction formalisée de données issues des signaux physiologiques a été développé sur le logiciel GamEval. Le logiciel GamEval a été implémenté par la société Iobjects pour une commande du laboratoire CHART. Ce logiciel permet la synchronisation des différents signaux issus de capteurs physiologiques (électroencéphalogramme, signal électrodermal) et du capteur oculométrique. Les annotations produites sur ce logiciel permettent à des experts de qualifier une séquence peuvent ainsi être traduite en données numérique directement exploitable par le logiciel GamExpé.

Tout comme pour l'exportation des analyses descriptives ou de l'exportation des observations, cette option est accessible via le menu « Fichier » de l'application. Une fois cette option choisie, une fenêtre de dialogue apparaît et permet de réaliser les calculs sur la globalité des sessions ou sur les annotations.

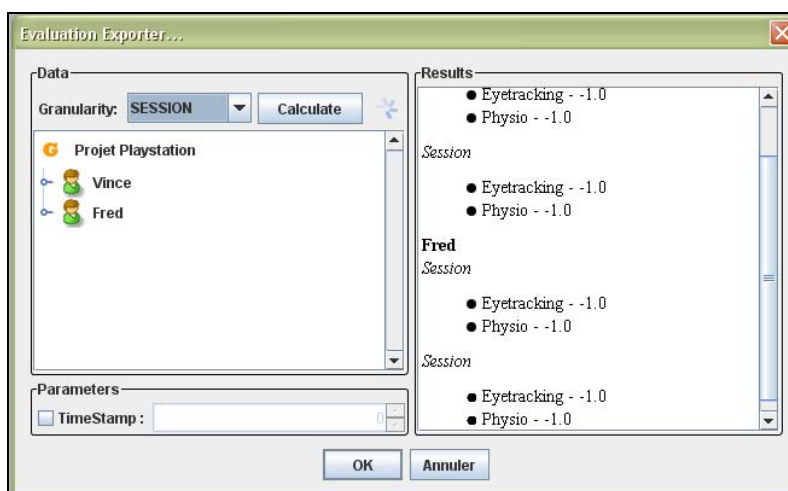


Figure 15 : Fenêtre pour l'exportation des données descriptives vers GamExpé

Une fois la granularité sélectionnée et les différents paramètres saisis, l'utilisateur peut lancer le calcul afin de voir si les résultats par session correspondent à ce qu'il attendait. Si cela lui convient, il peut exporter le résultat vers un fichier d'évaluation en appuyant sur « OK » sinon il est libre de changer les différents paramètres.

Cette fonctionnalité a pour but de donner des étiquettes sur les données objectives recueillies lors de session dûment annotées au vue de compléter celle sorties par une autre application récoltant soit des données subjectives soit d'autres données objectives.

Les évaluations sont principalement réalisées sur les données physiologiques (électrocardiogramme et électrodermale) ainsi que sur les données oculométriques.

Les critères extraits sont par exemple des valeurs statistiques comme le minimum, le maximum, la moyenne ou l'écart type prise par le signal sur la fenêtre choisie par l'utilisateur. Ils comprennent de plus des tendances comme l'augmentation ou la diminution à différents niveaux d'intensité. Par ailleurs, pour les signaux oculométriques, des attributs spécifiques sont ajoutés tels que le nombre de changement entre zones d'intérêt, le nombre de fixations oculaires, la vitesse moyenne ou la distance moyenne.

4.4 GamExpé

Comme indiqué dans la première partie de ce rapport, le LIP6 a retenu le modèle des gabarits pour

la conception de son expert artificiel capable de diagnostiquer la qualité d'un jeu vidéo. Un des nombreux avantages de cette méthode réside dans la possibilité de personnaliser les mécanismes d'agrégation qui conduisent à ce diagnostic.

La structure générale d'un gabarit forme un modèle générique capable de synthétiser l'information contenue dans les réponses d'un questionnaire ou bien de se nourrir d'informations extraites des signaux physiologiques pour évaluer la qualité globale d'un jeu-vidéo.

Un exemple de gabarit est présenté dans la figure ci-dessous. La personnalisation de ce modèle intervient dans le choix des fonctions d'agrégation (ici principalement des « et logique ») et éventuellement le poids donné à chacun des nœuds agrégés.

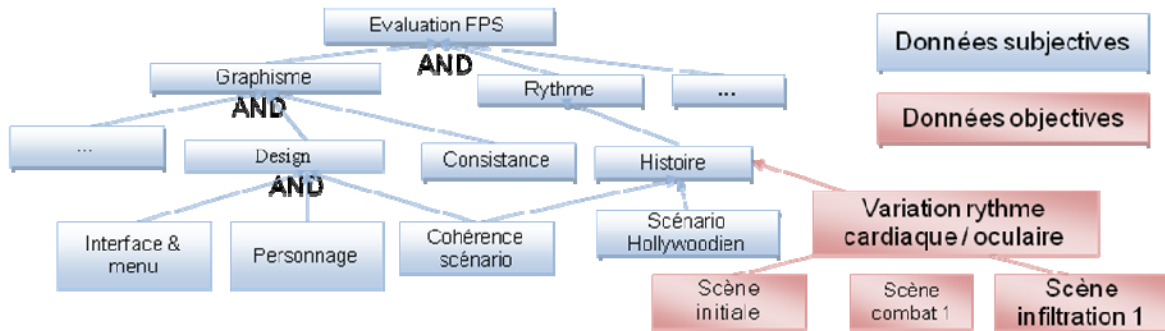


Figure 16 : Exemple de gabarit spécialisé pour les jeux de type First Person Shooter (FPS)

Le logiciel GameExpé permet de réaliser la conception d'un modèle de gabarit spécifique et le diagnostic d'un jeu à partir du modèle.

4.3.1. Conception d'un modèle spécifique de gabarit

La conception d'un modèle de gabarit doit pouvoir se faire de façon itérative comme représenté dans la figure ci-dessous. Durant les différentes phases du processus, un expert est amené à modifier les fonctions d'agrégation ainsi que les poids des différents critères pour refléter au mieux sa connaissance d'un type de jeu en particulier. A l'issue de ce processus, le modèle de gabarit ainsi spécialisé doit être à même de diagnostiquer un type de jeu en indiquant si le jeu est bon (pour cette famille de jeu) et sinon quels sont les critères qui pénalisent le plus son évaluation finale.

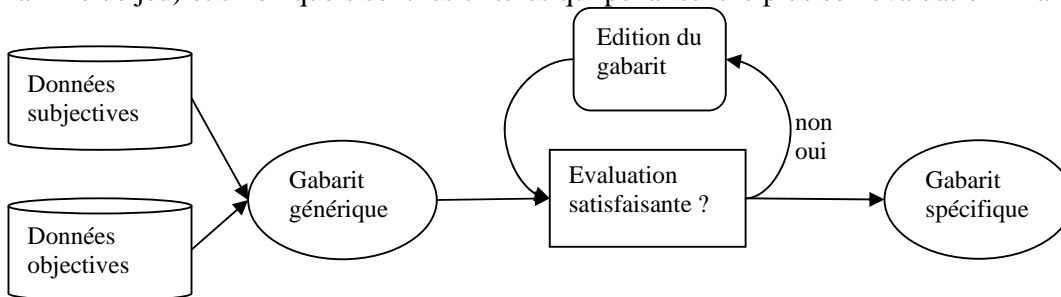


Figure 17 : Protocole de construction d'un gabarit spécifique

La tâche de conception d'un gabarit spécifique nécessite de pouvoir comparer facilement les résultats produits par des gabarits différents. L'appréciation d'une évaluation de la qualité d'un jeu étant toujours subjective, il est nécessaire de comparer nos résultats à un standard d'évaluation d'un jeu de la même catégorie.

4.3.2. Diagnostic d'un jeu

Afin d'évaluer un jeu, il est nécessaire de posséder des données issues de ce jeu soit sous la forme de questionnaires (données subjectives), soit d'informations issues des signaux physiologiques (données objectives). Les informations extraites des signaux sont obtenues principalement par étiquetage des phases de jeu puis calcul de mesures statistiques simples pour chacun des signaux sur ces périodes.

Une étiquette est l'association d'un label (critère) et d'une note qui sont rattachés à une évaluation faite par un sujet (de manière subjective ou objective). Le gabarit exige un ensemble d'étiquettes afin de pouvoir réaliser le diagnostic.

La figure suivante représente la chaîne de traitement pour le diagnostic d'un jeu.

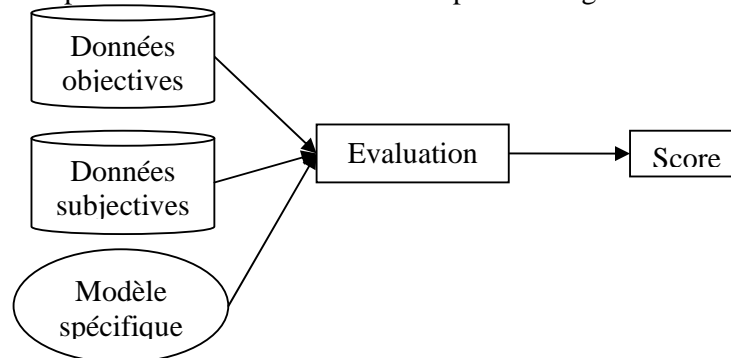


Figure 18 : Processus d'évaluation d'un jeu

Dans GameExpé, l'évaluation d'un gabarit est un ensemble de sujets accompagnés de leurs notes sur des critères constituant tout ou en partie le gabarit chargé. On parle d'évaluation partielle si les notes apportées ne sont pas suffisantes pour lancer une exécution du gabarit. Il est alors permis de charger plusieurs évaluations provenant de différentes sources pour permettre l'évaluation du nœud final.

Une exécution est le résultat de l'application de l'algorithme sur le gabarit courant et avec les données sélectionnées en entrée. L'exécution contient toutes les notes de ses nœuds ainsi que l'état du gabarit lors du lancement de l'algorithme.

5. Conclusions et perspectives

Ce rapport présente l'ensemble des méthodes, algorithmes et logiciels développés par le Laboratoire d'Informatique de Paris 6 dans le cadre du projet Gamelab.

L'objectif premier, qui consistait à réaliser un expert artificiel diagnostique de jeux vidéo, a été atteint en proposant une structuration des heuristiques les plus utilisées dans la littérature sur le gameplay des jeux vidéos et en associant cette connaissance à un modèle de gabarit alliant une structure générale représentative de l'ensemble des types de jeux vidéos à une possibilité d'adaptation pour tirer partie des spécificités de chacun des types de jeux. Ce modèle a été validé à partir de questionnaires sur un panel de trois jeux de types FPS pour lesquels le gabarit spécialisé a proposé des évaluations comparables à celles données par le site de référence *GameRankings.com*.

Trois logiciels ont été développés dans le cadre du projet pour répondre aux besoins de toute la chaîne de traitement de notre expert diagnostique et notamment la collecte des deux types de données prises en compte, données subjectives et données objectives, ainsi que l'analyse a posteriori par des méthodes de traitement de données automatique :

- la collecte et la gestion de données de questionnaires est réalisée par les logiciels GameLabQuest (interface Web implémentant la méthode des QCM évidentiels) et EvalGen (version standalone pour poste de travail),
- l'extraction de descripteurs de signaux physiologiques est faite par un plugin du logiciel GamEval pour l'intégration de données objectives issues des signaux physiologiques dans les gabarits,
- la construction, l'édition des structures de gabarit spécifiques et leur exécution sont rendues possibles par l'application GamExpé.

Enfin, d'un point de vue plus prospectif, les discussions et le travail réalisé au cours de ce projet ont permis de dégager des axes de travail nouveaux sur l'apprentissage automatique à partir de signaux physiologiques, qui sont encore sous-exploités du fait de leur complexité. Une première étude a été conduite avec succès validant notre approche et différenciant deux jeux à partir de descripteurs statistiques issus de traces oculaires.

Ces méthodes ont été présentées devant des industriels experts du domaine et de nouvelles pistes de recherche voient ainsi le jour. Les réalisations ont notamment permis de mesurer l'importance des travaux visant à caractériser les émotions humaines à l'aide de capteurs physiologiques et oculaires, couplé à des données subjectives ressenties par les personnes.

L'ensemble de ces résultats positifs autorise désormais à réfléchir à de nouveaux projets plus ambitieux impliquant une étude plus approfondie des utilisateurs d'un jeu (ou de manière plus large d'une interface logicielle) pour le caractériser plus efficacement et éventuellement l'améliorer.

Les données récoltées, en collaboration avec le laboratoire CHART durant le projet, et l'ensemble des protocoles expérimentaux pourront également être utilisés au LIP6 par un étudiant en thèse sur le domaine de l'analyse des signaux physiologiques et de l'*affective computing*

Références

- [1] W. Boucsein.. Electrodermal activity as an indicator of emotional processes, 1999.
- [2] E. Brown and P. Cairns. A grounded investigation of game immersion. In *Proceedings of the International Conference on Human Factors in Computing Systems*, pages 1297-1300. ACM New York, NY, USA, 2004.
- [3] E. Collain. Technique des gabarits. Rapport technique, Thomson-CSF, 1995.
- [4] C. Collet. 1997. Autonomic nervous system patterns specificity to basic emotions.
- [5] H. Desurvire, M. Caplan, and J. Toth. Using heuristics to evaluate the playability of games. In *Proceedings of the International Conference on Human Factors in Computing Systems*, pages 1509-1512. ACM New York, NY, USA, 2004.
- [6] J. Diaz, M. Rifqi, and B. Bouchon-Meunier. QCM évidentiels pour le diagnostic des apprenants. Actes des rencontres francophones sur la Logique Floue et ses Applications (LFA2007), 2007.
- [7] N. Falstein and H. Barwood. The 400 project. Available at <http://theinspiracy.com/400project.htm>
- [8] M. Federoff. Heuristics and usability guidelines for the creation and evaluation of fun in video games. Master's thesis, University Graduate School of Indiana University, 2002.
- [9] P. Goulev. An Investigation into the Use of AffectiveWare in Interactive Computer Applications, PHD Thesis, Supervisors: Prof. Abe Mamdani, Dr. Jeremy Pitt, Intelligent Systems & Network Group, Department of Electrical & Electronic Engineering Imperial College of Science, Technology & Medicine, 2005
- [10] Y. Isomoto, K. Yoshine, H. Nakatani, and N. Ishii. Data model and fuzzy information retrieval for scenic image database: theoretical extension of a traditional crisp model to fuzzy model. *Fuzzy information engineering*, pages 283-289, 1997.
- [11] R. W. Levenson. 1994. The search for autonomic specificity.
- [12] L. Mouillet. *Modélisation, reconnaissance et apprentissage de scénarios de conflits éthno-politiques*. Thèse de doctorat, Université Paris 6, 2005.
- [13] L. Mouillet, B. Bouchon-Meunier, and E. Collain. Automated identification of political conflicts with a scenario recognition technique. In *Proceedings of IPMU*, volume 3, pages 1609-1616, 2004.
- [14] P. Sweetser and P. Wyeth. Gameflow: A model for evaluating player enjoyment in games. *ACM Computers in Entertainment*, 3(3), 2005.
- [15] P. Vorderer, T. Hartmann, and C. Klimmt. Explaining the enjoyment of playing video games: the role of competition. In *ACM International Conference Proceedings Series*, volume 38, 2003.
- [16] A. L. Yarbus. Eye movements during perception of complex objects. In L. A. Riggs, ed., *Eye*

Movements and Visio', Plenum Press, chapter 7, pp. 171-19, 1967.

[17] <http://www.jeuxvideo.com/>

[18] <http://www.jeux-france.com/>

[19] <http://www.gamekult.com/>

[20] Gamerankings. 2007. Available at <http://www.gamerankings.com/>.

Annexe

Liste des heuristiques utilisées pour évaluer la jouabilité d'un jeu vidéo par H. Desurvire, M. Caplan et J. Toth [5]

Game Play	
1	Player's fatigue is minimized by varying activities and pacing during game play.
2	Provide consistency between the game elements and the overarching setting and story to suspend disbelief.
3	Provide clear goals, present overriding goal early as well as short-term goals throughout play.
4	There is an interesting and absorbing tutorial that mimics game play.
5	The game is enjoyable to replay.
6	Game play should be balanced with multiple ways to win.
7	Player is taught skills early that you expect the players to use later, or right before the new skill is needed.
8	Players discover the story as part of game play.
9	Even if the game cannot be modeless, it should be perceived as modeless.
10	The game is fun for the Player first, the designer second and the computer third. That is, if the non-expert player's experience isn't put first, excellent game mechanics and graphics programming triumphs are meaningless.
11	Player should not experience being penalized repetitively for the same failure.
12	Player's should perceive a sense of control and impact onto the game world. The game world reacts to the player and remembers their passage through it. Changes the player makes in the game world are persistent and noticeable if they back-track to where they've been before
13	The first player action is painfully obvious and should result in immediate positive feedback.
14	The game should give rewards that immerse the player more deeply in the game by increasing their capabilities (power-up), and expanding their ability to customize.
15	Pace the game to apply pressure but not frustrate the player. Vary the difficulty level so that the player has greater challenge as they develop mastery. Easy to learn, hard to master.
16	Challenges are positive game experiences, rather than a negative experience (results in their wanting to play more, rather than quitting).
Game Story	
1	Player understands the story line as a single consistent vision.
2	Player is interested in the story line. The story experience relates to their real life and grabs their interest.
3	The Player spends time thinking about possible story outcomes.
4	The Player feels as though the world is going on whether their character is there or not.
5	The Player has a sense of control over their character and is able to use tactics and strategies.
6	Player experiences fairness of outcomes.
7	The game transports the player into a level of personal involvement emotionally (e.g., scare, threat, thrill, reward, punishment) and viscerally (e.g., sounds of environment).
8	Player is interested in the characters because (1) they are like me; (2) they are interesting to me, (3) the characters develop as action occurs.
Mechanics	
1	Game should react in a consistent, challenging, and exciting way to the player's actions (e.g., appropriate music with the action).
2	Make effects of the Artificial Intelligence (AI) clearly visible to the player by ensuring they are consistent with the player's reasonable expectations of the AI actor.
3	A player should always be able to identify their score/status and goal in the game.
4	Mechanics/controller actions have consistently mapped and learnable responses.
5	Shorten the learning curve by following the trends set by the gaming industry to meet user's expectations.
6	Controls should be intuitive, and mapped in a natural way; they should be customizable and default to

	industry standard settings
7	Player should be given controls that are basic enough to learn quickly yet expandable for advanced options.
Usability	
1	Provide immediate feedback for user actions.
2	The Player can easily turn the game off and on, and be able to save games in different states.
3	The Player experiences the user interface as consistent (in control, color, typography, and dialog design) but the game play is varied
4	The Player should experience the menu as a part of the game.
5	Upon initially turning the game on the Player has enough information to get started to play.
6	Players should be given context sensitive help while playing so that they do not get stuck or have to rely on a manual.
7	Sounds from the game provide meaningful feedback or stir a particular emotion.
8	Players do not need to use a manual to play game.
9	The interface should be as non-intrusive to the Player as possible.
10	Make the menu layers well-organized and minimalist to the extent the menu options are intuitive.
11	Get the player involved quickly and easily with tutorials and/or progressive or adjustable difficulty levels.
12	Art should be recognizable to player, and speak to its function.

