

Chapitre 9

Expérimentations

Sommaire

9.1 Objectifs de l'expérience	2
9.2 Dispositif expérimental	3
9.2.1 Choix matériels et logiciels	3
9.2.2 Gameplay et courbes de difficulté	5
9.2.3 Déroulement de l'expérience	9
9.2.4 Données recueillies	10
9.3 Résultats	11
9.3.1 Evaluation des capacités du joueur et difficulté	11
9.3.2 Difficulté calculée et difficulté théorique	20
9.3.3 Difficulté calculée et difficulté subjective	22
9.3.4 Plaisir de jouer et niveau de difficulté	23
9.3.5 Plaisir de jouer et courbe de difficulté	27
9.4 Discussion et Conclusion	28

Dans les chapitres précédents de cette thèse (chap. ??, ?? et ??), nous avons proposé un modèle de mesure de la difficulté d'un jeu vidéo. Nous avons également soulevé certaines hypothèses quant aux caractéristiques optimales de la courbe de difficulté d'un jeu vidéo (chap. ??). Dans ce chapitre, notre objectif consiste à mettre à l'épreuve nos hypothèses ainsi qu'à tester l'efficacité et la validité de notre modèle sur un cas réel. Nous présentons donc une étude expérimentale de la difficulté d'un jeu vidéo, à partir d'un gameplay réalisé par nos soins, implanté dans un moteur de jeu et soumis à un panel de joueurs.

Dans la première partie de ce chapitre, nous présentons plus en détail les objectifs de cette partie expérimentale. Nous décrivons ensuite nos choix matériels, nos choix de gameplay, puis analysons les résultats obtenus. Nous terminons ce chapitre par une discussion des résultats obtenus.

9.1 Objectifs de l'expérience

L'objectif principal de cette thèse consiste à fournir aux game designers un modèle de mesure de la difficulté d'un jeu vidéo. Ce chapitre de la thèse se consacre tout d'abord à l'étude de la validité et de l'efficacité de notre modèle en l'employant effectivement dans un cas concret. Ensuite, notre travail a consisté à proposer un certain nombre d'hypothèses quant à l'impact de la forme d'une courbe de difficulté sur le plaisir du joueur (chap.??). Nous souhaitons donc également éprouver ces hypothèses.

Nous avons ainsi mis au point un gameplay, de manière à soumettre un échantillon de joueurs à différentes courbes de difficulté. Nous avons étudié le comportement des joueurs et recueilli leur ressenti. Nous avons ainsi pu éprouver notre modèle en tentant de mesurer la difficulté de ce gameplay, tout en vérifiant également la validité de nos hypothèses, par l'étude des résultats subjectifs suscités par chaque type de courbe de difficulté.

Nous souhaitons tout d'abord montrer qu'il est effectivement possible de définir et de mesurer certaines des capacités du joueur. Nous avons proposé d'encoder la détection des capacités sous la forme d'un algorithme de traitement des traces d'évènements. Cette expérience doit démontrer que sur un gameplay suffisamment représentatif des jeux commerciaux actuellement développés, cette technique est effectivement applicable.

Nous désirons ensuite montrer que notre modèle permet de mettre en évidence le lien qui unit les capacités du joueur à la difficulté du jeu. Un tel lien confirmerait, dans le cadre de cette expérience, que notre mesure de capacité fournit bel et bien la valeur d'une variable explicative de la difficulté du gameplay.

Nous souhaitons également montrer qu'il existe un lien entre la difficulté empirique, calculée par notre modèle à partir des capacités du joueur, et la difficulté théorique que nous avons évalué lors de la création du jeu. Nous avons en effet construit notre gameplay dans l'objectif de soumettre le joueur à une certaine difficulté, que nous sommes plus ou moins capables d'évaluer a priori. Nous allons étudier si la difficulté calculée par le modèle correspond à notre vision intuitive, prévisionnelle, de la difficulté du jeu.

Ensuite, nous souhaitons comprendre si notre mesure de difficulté permet d'évaluer la difficulté subjective, ressentie et rapportée par le joueur. Nous avons intégré à notre dispositif expérimental un questionnaire permettant de recueillir la perception qu'a le joueur de la difficulté du jeu. Cette mesure doit nous permettre d'étudier le lien entre la difficulté telle que nous la calculons et la difficulté telle que le joueur la perçoit.

Pour finir, nous souhaitons vérifier l'hypothèse formulée au chapitre ?? . Nous postulons dans ce chapitre qu'une courbe de difficulté optimale respecte les caractéristiques suivantes :

- une première phase de progression graduelle,
- une seconde phase de forte variabilité,
- les challenges les plus difficiles doivent à peine dépasser le niveau du joueur.

Notre expérience se déroule donc en deux phases consécutives. Une de ces phases propose une courbe de difficulté basique de suivi du niveau du joueur, tandis que l'autre propose une courbe respectant l'ensemble des caractéristiques pré-citées¹. L'ordre de ces phases est tiré au hasard en début de partie pour éviter tout effet d'ordre. Nous avons également intégré un questionnaire portant sur le plaisir ressenti par le joueur. L'étude du lien entre le type de courbe de difficulté et le plaisir rapporté par le joueur devrait nous éclairer sur la validité de notre hypothèse.

9.2 Dispositif expérimental

9.2.1 Choix matériels et logiciels

Pour réaliser cette expérience, nous devons avoir accès à un logiciel d'édition de jeu vidéo, de manière à construire notre univers et les lois qui le gouvernent. Mais au delà de l'édition des comportements des objets du jeu, nous devons également avoir accès au coeur du moteur de jeu. Nous devons en effet être en mesure d'implanter le code d'enregistrement des événements retraçant le parcours du joueur. En tant qu'académiques, l'accès aux moteurs de jeu nous est relativement restreint. Les collaborations avec les studios de jeu sont complexes à mettre au point pour ce type d'expérience. Il est en effet nécessaire de nous laisser accéder au code source du jeu, ce que les studios autorisent très rarement. Nous avons également besoin d'accéder à un jeu suffisamment abouti. Or, plus le jeu est abouti, plus le studio approche de la date de mise sur le marché et donc moins celui-ci est en mesure de consacrer du temps à des activités de recherche. Une fois le jeu terminé, l'accès au code reste problématique dans la mesure où le studio passe très rapidement à un nouveau projet. Le cycle de production d'un jeu est donc particulièrement complexe à synchroniser avec une activité de recherche.

Cependant, certains studios ont pour objectif de commercialiser leur moteur de jeu, comme c'est le cas de Valve ou d'Epic Games. Leurs moteurs de jeu sont parmi les plus performants, et gratuitement accessibles au grand public pour une utilisation non commerciale. Ces moteurs jouissent donc d'une grande popularité chez les universitaires. Nous avons choisi d'utiliser le moteur commercialisé par Epic Games, l'Unreal Engine. Ce moteur permet avant tout la réalisation de jeu de type **First Person Shooter**². Ce type de jeu nous convient parfaitement puisqu'il est particulièrement populaire et correspond au type d'une grande partie des jeux les plus récents. Nous avons donc réalisé un First Person Shooter à l'aide de l'Unreal Development Kit (UDK).

L'Unreal Development Kit permet de compiler des jeux exécutables par plusieurs types de plateformes. Nous avons choisi de développer notre jeu pour PC, de manière à simplifier les

1. Ces courbes sont décrites dans les sections suivantes
2. First Person Shooter, jeu de tir en vue subjective.

développements annexes nécessaires. En effet, UDK ne permet pas, à l'origine, d'enregistrer facilement une séquence d'évènements. Le seul accès possible au système de fichier est fourni par la fonction `'log()` qui permet d'ajouter un message au journal d'erreurs du moteur. Ce fichier est géré de manière automatique par le moteur, et par exemple effacé à chaque démarrage, ce qui ne nous permet pas de réaliser notre expérience. Nous avons donc développé une librairie dynamique qui nous permet d'enregistrer n'importe quel évènement dans un fichier XML unique, créé automatiquement à chaque début d'expérience. Cette librairie enrichit le langage de script d'UDK d'une fonction permettant d'accéder à ce log particulier. Cette fonction nous a permis d'éditer de nombreux scripts régissant le comportement du moteur de jeu pour rediriger les évènements utiles vers notre propre fichier de log.

Les ressources sonores et graphiques nécessaires au développement sont également très largement fournies par la version gratuite d'UDK. Ils proviennent du jeu *Unreal Tournament*, First Person Shooter développé par Epic Games. Le moteur dispose de suffisamment d'objets graphiques pour permettre de créer un environnement de jeu complet. Afin d'apporter une touche originale à l'univers et de s'éloigner de l'univers plutôt noir et violent d'Unreal Tournament, nous avons toutefois choisi d'utiliser un personnage populaire et décalé comme ennemi du joueur : le lapin crétin (Figure 9.1)³. L'univers d'Unreal Tournament cible particulièrement les joueurs dits *hardcore*, les plus passionnés, et nous souhaitons élargir ce public, ce que permet en partie l'utilisation du lapin crétin.



FIGURE 9.1 – L'ennemi du joueur

3. Le lapin crétin est une création originale d'Ubisoft, que nous avons importé ici uniquement à titre expérimental.

9.2.2 Gameplay et courbes de difficulté

Pour cette expérience nous avons conçu un gameplay proche du mode *Deathmatch* d'Unreal Tournament. Dans une partie de type *Deathmatch*, le joueur est face à un groupe d'ennemis. Chaque personnage du jeu, joueur ou non joueur, doit remplir les mêmes objectifs et peut s'attaquer à n'importe quel autre personnage. Une partie de ces objectifs concerne l'exploration et la collecte : se déplacer dans le **Niveau de jeu**⁴ pour découvrir et ramasser armes, munitions et boucliers. Tous les joueurs explorent le même terrain de jeu, et l'objectif principal consiste à tirer sur ses adversaires pour les détruire, tout en évitant de subir le même sort. A chaque fois qu'un personnage est détruit, il réapparaît à un endroit choisis aléatoirement dans le niveau de jeu, et le personnage qui l'a détruit marque un point. Le match se termine lorsqu'un score ou une durée particulière sont atteints. Le personnage avec le plus de points à la fin du match a gagné.

Nous avons apporté quelques modifications au mode *Deathmatch* d'Unreal Tournament. Nous avons modifié l'**Intelligence Artificielle**⁵ des personnages non joueurs, de manière à pouvoir maîtriser leur comportement. L'Intelligence Artificielle originale dispose de comportements tactiques complexes que nous souhaitons simplifier, ainsi que d'une adaptation automatique de niveau qu'il nous fallait bien sur désactiver. Le niveau d'un ennemi correspond à la précision de ses tirs, à son temps de réaction ainsi qu'au type d'actions qu'il peut réaliser. Nous avons également choisis d'assujettir la vitesse d'un ennemi à son niveau, en plus des autres paramètres déjà gérés par les scripts originaux de l'UDK, de manière à pouvoir offrir un niveau de difficulté le plus bas possible.

Dans le gameplay original d'Unreal Tournament, le niveau général des ennemis est fixé en début de partie, et chaque ennemi modifie ensuite son niveau automatiquement : à chaque fois qu'un personnage non joueur est détruit, son niveau augmente, et à chaque fois qu'il tue le joueur, son niveau diminue. Nous avons supprimé cette adaptation répartie de la difficulté pour la remplacer par un dispositif centralisé, qui permet d'attribuer un niveau à l'ensemble des ennemis du joueur en fonction de la difficulté que nous souhaitons donner au jeu.

Nous avons ensuite construit un mode de jeu plus segmenté dans le temps que ne l'est le mode *Deathmatch* original. Les personnages non joueurs sont tous des lapins crétins et sont tous ennemis du joueur, mais ne s'attaquent pas entre eux. Le joueur doit venir à bout de l'ensemble des ennemis du niveau pour marquer un point, et un ennemi détruit ne réapparaît plus automatiquement. De cette manière, le joueur affronte des vagues successives d'ennemis, ce qui nous permet d'avoir un contrôle plus précis de la courbe de difficulté. Au départ de chaque vague d'ennemis, le joueur dispose de cent points de vie, et aucun objet du niveau ne lui permet d'en récupérer. Chaque vague d'ennemis correspond ainsi à un challenge, dont

4. Le niveau de jeu désigne l'espace dans lequel évolue le joueur, plus particulièrement lorsque l'univers du jeu est présenté au joueur de manière clairement segmentée, sous espace après sous espace.

5. Intelligence Artificielle.

la liste est fournie figure 9.2.

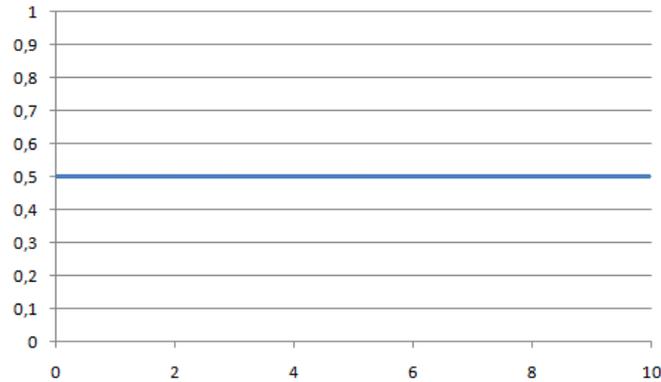
Challenge	Nombre d'ennemis	Niveau des ennemis
0	1	0
1	1	1
2	1	2
3	2	2
4	2	3
5	3	3
6	3	4
7	4	4
8	4	5
9	5	5
10	6	5

FIGURE 9.2 – Liste des challenges.

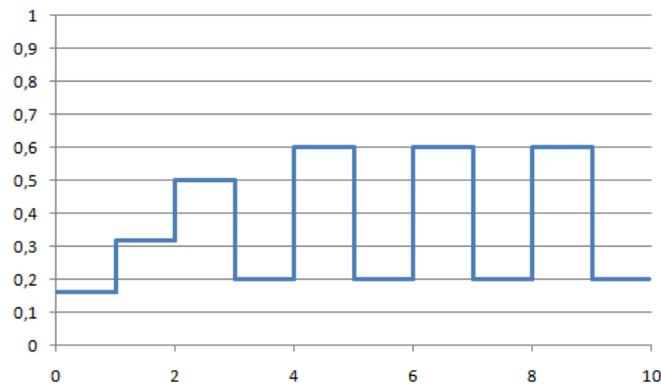
Ces challenges, classés figure 9.2 par ordre de difficulté que nous supposons croissante, nous permettent de composer notre courbe de difficulté. Nous soumettons le joueur à deux courbes différentes, pendant deux phases de jeu successives, l'ordre des phases étant aléatoire.

La première courbe de difficulté est une courbe de *suivi*, qui adapte en permanence la difficulté du jeu au niveau du joueur. Elle correspond à l'interprétation la plus directe des travaux de Mihaly Csikszentmihalyi sur le *flow* étudiée chapitre ?? [Csikszentmihalyi 91]. La courbe de difficulté relative théorique de *suivi* est représentée figure 9.3. On note que cette courbe décrit la difficulté comme nous la définissons, c'est à dire sous la forme d'une probabilité d'échec. Si le niveau du joueur et le niveau du jeu sont strictement équivalents, alors la probabilité que le joueur gagne est aussi forte que celle qu'il perde, ce qui correspond à une difficulté relative de 0.5.

Une courbe de difficulté relative qui correspond au suivi parfait du niveau du joueur est en permanence à 0.5, telle que tracée figure 9.3. Dans les faits, cette courbe est plus souvent oscillante, car la difficulté absolue du jeu ne peut évoluer pour suivre le niveau du joueur que lorsque ce changement de niveau a été constaté, et donc lorsque la difficulté relative constatée d'un challenge est différente de 0.5. Une courbe de suivi parfaite ne serait réalisable qu'en étant capable d'anticiper précisément les progrès du joueur.

FIGURE 9.3 – Courbe de difficulté relative et théorique de type *suivi*

Le second type de courbe de difficulté auquel un joueur est confronté dans notre expérience reprend les caractéristiques listées dans la section 9.1. C'est une courbe supposée optimale, avec une phase de progression graduelle suivie d'une oscillation autour du niveau du joueur. Les phases hautes d'oscillation ne dépassent que très légèrement le niveau du joueur. Cette courbe est représentée figure 9.4.

FIGURE 9.4 – Courbe de difficulté de type *progression puis oscillation*

Pour construire ces deux courbes de difficulté théoriques, nous avons évalué le niveau du joueur comme le numéro du challenge qui lui correspond le mieux. Nous mettons à jour le niveau du joueur, à chaque fois qu'il termine un challenge, selon l'algorithme présenté ci-après.

```
1 int evalNiveauJoueur(int niveauActuel, int numChallenge, string
   resultat)
```

```

2 {
3     int niveauJoueur = niveauActuel;
4
5     //Si le joueur perd alors que le challenge est à son niveau ou
6     //plus facile
7     if(resultat == "Echec")
8     {
9         if(numChallenge <= niveauActuel)
10            niveauJoueur = numChallenge-1
11    }
12
13    //Si le joueur gagne alors que le challenge est à son niveau ou
14    //plus difficile
15    if(resultat == "Valide")
16    {
17        if(numChallenge >= niveauActuel)
18            niveauJoueur = numChallenge+1
19    }
20 }

```

Nous avons construit nos courbes de difficulté en évaluant à priori les probabilités d'échec du joueur. Cette probabilité dépend de la différence entre le niveau du joueur et le numéro de challenge qu'il doit relever. Ces valeurs constituent la difficulté théorique de notre gameplay, définie à priori. Elles sont présentées figure 9.5.

Différence de niveau	Difficulté théorique
<-2	0.95
-2	0.8
-1	0.6
0	0.5
1	0.4
2	0.2
>2	0.05

FIGURE 9.5 – Difficulté théorique d'un challenge.

La différence de niveau correspond à la soustraction du niveau du challenge au niveau du joueur.

9.2.3 Déroutement de l'expérience

L'ensemble de l'expérience est réalisée dans l'univers du jeu, en limitant ainsi au maximum les interventions de l'expérimentateur. Un premier écran invite le joueur à fournir ses informations personnelles, son nom, prénom, âge et genre. Suit une fenêtre d'explication de l'interface de contrôle de l'avatar. Le joueur est ensuite invité à évoluer durant deux minutes dans l'environnement de jeu, limité pour l'instant à une petite arène centrale. Pendant toute cette phase, l'expérimentateur est à côté du joueur et lui apporte tout éclaircissement ou rappel nécessaire, si le joueur en fait la demande.

Vient ensuite la première phase de l'expérience, pendant laquelle nous évaluons une première fois le niveau du joueur. L'objectif du joueur consiste à éliminer le plus de lapin possibles en un temps limité. Les lapins apparaissent dans l'arène l'un après l'autre, et sont de niveau croissant. Leur Intelligence Artificielle est un peu particulière car ils ne choisissent jamais d'attaquer le joueur. Ils se contentent de fuir avec une agilité croissante. De cette manière, nous évaluons grossièrement la capacité du joueur, sans jamais le frustrer puisqu'il ne peut pas perdre.

À l'issue de cette phase de découverte et d'évaluation, le joueur se voit offert l'accès à l'ensemble du Niveau de jeu. Le premier challenge débute alors, avec une difficulté correspondant à une des deux courbes définies dans la section précédente. Au bout de 12 minutes, le système change de courbe. Au bout de 24 minutes, l'expérience est terminée.

Tout au long de l'expérience, à l'issue de chaque challenge, le joueur remplit un rapide questionnaire sur son plaisir et la difficulté qu'il perçoit. Ce questionnaire se compose de deux échelles de Lickert à 5 niveaux, présentées figure 9.6.

Pendant ce tout dernier groupe d'adversaires :

1) Prenez vous plaisir à jouer ?

1	2	3	4	5
pas du tout	un peu	moyen	beaucoup	parfait !

2) Le jeu était t'il difficile ?

1	2	3	4	5
très facile	plutot facile	moyen	plutot difficile	très difficile

FIGURE 9.6 – Questionnaire sur la difficulté et le plaisir du joueur

9.2.4 Données recueillies

Le système enregistre de nombreux événements qui retracent le parcours du joueur et son comportement, listés figure (Figure 9.7) page 10.

Numéro	Titre	Données
03	Player Dead	
04	Bot Dead	Name
05	Heal	Action.HealAmount
06	Player Pickup	Inv.ItemName
07	Touch	Identification
08	UnTouch	Identification
09	BornIn	TrigVolume.Identification
10	Name	Name
11	Change Weapon	NewWeapon.ItemName
12	Reset Skill	Skill :Pawn.GroundSpeed :Pawn.Name
13	Set specific level	Skill :Pawn.GroundSpeed :Pawn.Name
14	Bot Heal	Action.HealAmount
15	Damage Player	Damage :Health :by bot by player
16	Damage Bot	Damage :Health : by bot by player
17	Health Status Bot	Health
18	Health Status Player	Health
19	Damage Scaling Bot	DamageScaling
20	Damage Scaling Player	DamageScaling
21	Start of experience	
22	Possess	sBotName
23	UnPossess	sBotName :LogDead
24	Destroyed	sBotName :LogDead
25	Bot Dead	sBotName :LogDead
26	Identite	Nom :Annee :Homme :Femme
27	Report	Difficulte :Fun
28	Niveau	NivJoueur :NivJeu :NbBots :NivBots
29	Difficulte	SlowStart Oscillation Suivi
30	Start Challenge	NivJeu :NivJoueur
31	End Challenge	1 0
32	Fire	danger :cible
33	HasMoved	X :Y :Z :danger :cible
34	Danger	danger :cible

FIGURE 9.7 – Liste des événements.

Nous évaluons trois capacités du joueur à partir de la trace des différents événements : sa capacité à viser, sa capacité à rester en mouvement, et sa capacité à parcourir le niveau. La capacité du joueur à viser, sa précision, correspond au rapport entre le nombre de tirs supposés utiles qu'il effectue, c'est à dire lorsqu'il voit un ennemi ou lorsqu'un ennemi le pourchasse, et le nombre de fois ou un de ses tirs aboutit à blesser un ennemi, calculé grâce

aux trames 32 et 02. Sa capacité à rester un mouvement correspond au nombre d'évènements 33 reçus par seconde, lorsque le joueur est en danger ce dont nous informe l'évènement 34. Sa capacité à parcourir le niveau correspond au nombre de fois où le joueur change de zone par seconde, ce dont nous informe l'évènement 07, lorsqu'il est en danger, ce dont nous informe l'évènement 34.

L'évènements 27 nous fournit les résultat du questionnaire de fin de challenge, à savoir le niveau de difficulté et le plaisir perçus par le joueur.

9.3 Résultats

Nous avons réalisé trois sessions d'expérimentation. La première s'est déroulée dans notre laboratoire, au CEDRIC à Paris, et a regroupé 30 sujets. Une seconde s'est déroulée à dans les locaux de l'ENJMIN, à Angoulême, sur 35 sujets. La dernière a été réalisée au sein d'une association de joueurs⁶ et a regroupé 7 sujets. Les deux dernières expériences ont été encadrées par deux élèves ergonomes de l'ENJMIN⁷. Nous avons donc, en tout, pu étudier le comportement de 72 sujets, 56 hommes et 16 femmes. La moyenne d'âge des joueurs est de 26 ans, l'ensemble des âges étant répartis comme le présente la figure 9.8.

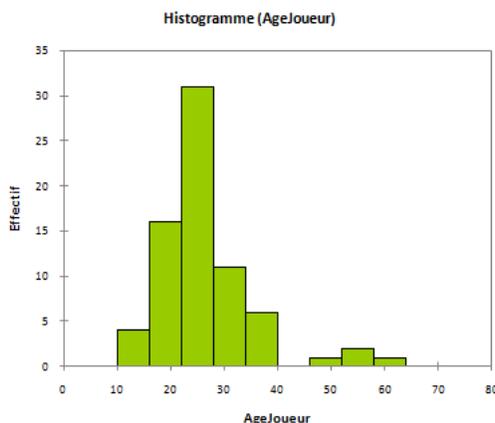


FIGURE 9.8 – Age des participants

9.3.1 Evaluation des capacités du joueur et difficulté

Chaque sujet a pu jouer à une suite de challenges, pendant les 24 minutes du test. Nous avons ainsi pu enregistrer 2401 tentatives de résolution de challenge, couronnées ou non

6. Association BLASS de Brioux sur Boutonne.

7. Merci à Joel Gueulin et Cédric Mondamert.

de succès. La figure 9.9 montre la distribution du nombre de challenges joués par joueur. Nous avons choisit de ne pas tenir compte des deux joueurs ayant enregistré un nombre de challenges anormalement bas. En effet, ces deux joueurs n'ont pas réalisé l'expérience jusqu'au bout mais ont demandé à arrêter au bout d'une dizaine de minutes. Leur abandon peut dénoter un manque de motivation et fausse donc nos analyses de performance, comme nous l'avons détaillé chapitre ???. Nous avons donc conservé 70 joueurs, c'est à dire 2368 tentatives de résolution d'un challenge.

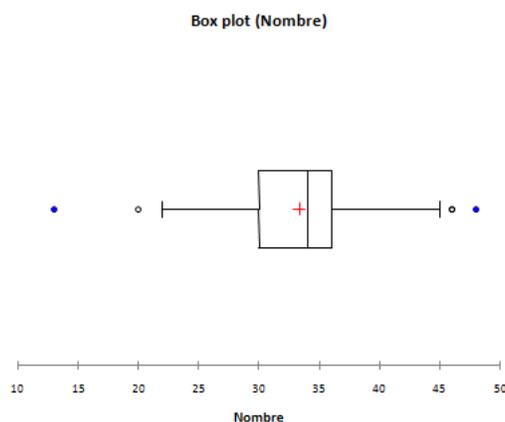


FIGURE 9.9 – Nombre de challenges joués par joueur

Pour chacune de ces tentatives, nous avons analysé, à l'aide des traces et du logiciel de calcul de la difficulté, les capacités démontrées par le joueur. Nous avons ensuite étudié le lien entre les capacités mesurées et l'échec des joueurs. Les résultats sont présentés par challenge. En effet, la réussite de chaque challenge ne dépend pas forcément des mêmes capacités. Certains challenges sollicitent plus les capacités de déplacement du joueur que d'autres, par exemple. La figure 9.10 montre la distribution du nombre de traces enregistrées pour chaque challenge.

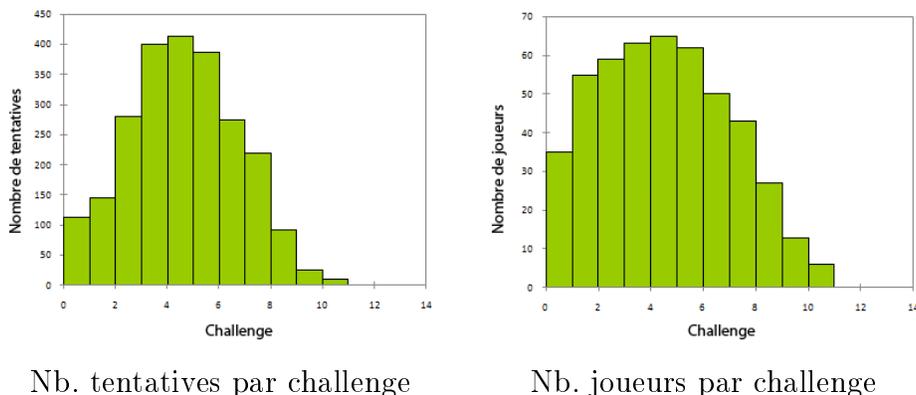


FIGURE 9.10 – Quantité de données par challenge.

Nous ne détaillons complètement ici que l'analyse du challenge le plus joué, c'est à dire le challenge 4. Nous présentons par la suite une synthèse des résultats obtenus pour l'ensemble des challenges.

9.3.1.1 Analyse détaillée du challenge 4

Nous avons analysé les traces du challenge 4 à l'aide du logiciel d'analyse de la difficulté présenté chapitre ???. Les algorithmes de mesure des capacités ont put extraire des valeurs correspondant au trois capacités que nous souhaitons mesurer, c'est à dire la précision des tirs, la fréquence de mouvement et la fréquence de déplacement du joueur. Pour rappel, le mouvement correspond pour nous à la capacité du joueur à ne pas rester statique, et sa capacité de déplacement correspond au fait que non seulement il bouge mais en plus se déplace de manière significative dans le niveau de jeu, et donc change de zone. Une fois ces valeurs extraites, nous avons analysé leur relation avec le résultat obtenu par le joueur.

Nous considérons les variables de capacité comme des variables explicatives du résultat, et recherchons des corrélations linéaires entre ces variables. La figure 9.11 montre les distributions des variables de capacités et de la variable de résultat. On remarque que ces distributions ne sont pas normales, ce qui proscrie l'usage des tests statistiques paramétriques. Nous avons choisit de tester nos hypothèses à l'aide d'un test d'indépendance χ^2 .

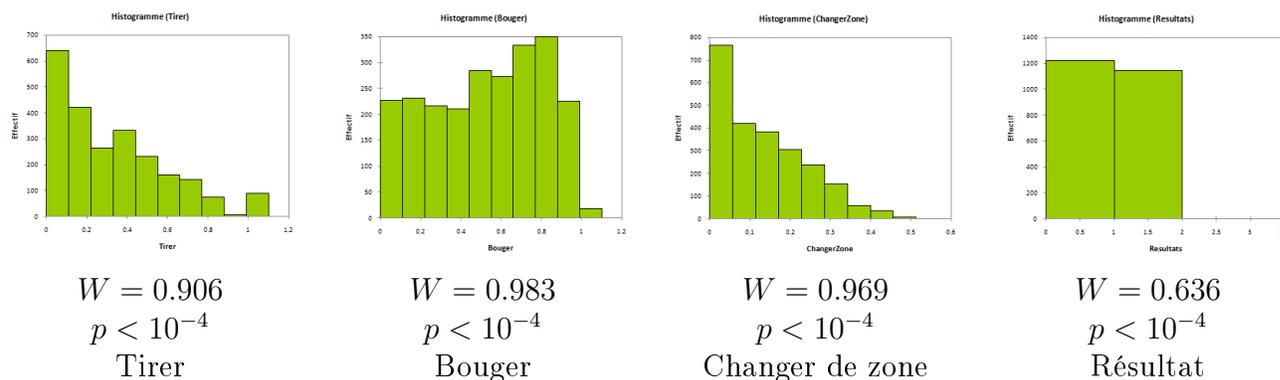


FIGURE 9.11 – Distributions des variables et tests de normalité (Shapiro/Wilks).

La figure 9.12 montre le coefficient de corrélation linéaire calculé pour les différents couples de variables. La capacité du joueur à tirer semble être la plus liée au résultat, avec une corrélation de -0.22 . Cette corrélation est faible mais la variable résultat est une variable binaire, qui reflète moins la difficulté d'un challenge que la probabilité d'échec par niveau, que nous avons établi comme notre mesure de la difficulté.

	Tirer	Bouger	Changer de zone	Résultat
Tirer	1	0.229	0.141	-0.215
Bouger	0.229	1	0.736	-0.111
Changer de zone	0.141	0.736	1	-0.068
Résultat	-0.215	-0.111	-0.068	1

FIGURE 9.12 – Tableau de corrélation capacités et résultat (Coefficient de Pearson).

Nous avons cherché à vérifier si'il existait un lien significatif entre la capacité du joueur à tirer et le résultat qu'il obtient à la fin du challenge. Cette hypothèse peut être formulée comme l'alternative de l'hypothèse nulle H_0 suivante : *La distribution du nombre de joueurs ayant gagné ou perdu est indépendante de leur précision.* Nous avons construit le tableau de contingence présenté figure 9.13 et réalisé un test χ^2 d'indépendance avec un seuil $\alpha = 10^{-2}$. Nous avons construit les classes du tableau de contingence à l'aide de l'algorithme de classification des k-moyennes.

	Gagne	Perdu	Total
Precision 1	29	32	61
Precision 2	35	60	95
Precision 3	13	18	31
Precision 4	25	10	35
Precision 5	15	26	28
Precision 6	18	10	61
Precision 7	22	11	33
Precision 8	25	10	35
Precision 9	11	12	23
Precision 10	20	12	32
Total	213	201	414

FIGURE 9.13 – Tableau de contingence entre précision et résultat.

Le résultat du test montre que nous pouvons rejeter l'hypothèse H_0 avec un risque inférieur à $3 * 10^{-4}$. Nous acceptons donc l'hypothèse alternative : *Il existe un lien entre le résultat obtenu par le joueur et sa précision au tir*. Nous avons réalisé le même test pour les autres capacités et obtenu le résultat inverse : nous ne pouvons pas rejeter H_0 pour les capacités de mouvement ($p = 0.26$) et de déplacement ($p = 0.085$) et considérer qu'il existe un lien entre ces capacités et le résultat. Seule la précision semble être liée significativement au résultat.

Il existe donc bien un lien significatif entre la capacité au tir du joueur et les résultats qu'il obtient au challenge 4. Ce lien peut par exemple être calculé comme une corrélation linéaire de -0.22 entre ces deux variables. Mais comme nous l'avons dit précédemment, la variable *résultat* reflète mal la performance du joueur, car elle la binarise. Par contre, à partir de multiples essais, nous pouvons calculer la probabilité d'échec du joueur, qui reflète plus précisément sa performance.

On notera qu'une partie des résultats a été obtenue *par le même joueur*, et qu'à ce titre, les différentes observations de la variable *résultat* ne sont pas totalement indépendantes, propriété nécessaire pour permettre dévaluer une probabilité à partir d'observations. En effet, les résultats successifs d'un même joueur sont dépendants puisque ce joueur apprend à chaque essai et donc modifie sa probabilité de réussite. Mais nous cherchons à calculer la probabilité de réussite des joueurs d'un même niveau, et satisfaisons donc l'hypothèse d'indépendance conditionnelle. En d'autres termes, nous supposons que la seule dépendance existant entre les différentes observations de la variables *résultat* vient de l'évolution du niveau du joueur, facteur que nous maîtrisons en regroupant les observations par classes de niveau similaires. Cependant, l'indépendance conditionnelle n'est totalement vérifiée que si nous tenons compte de l'ensemble des dimensions du niveau du joueur. Notre indépendance est donc approximative car nous nous contentons ici d'exprimer le niveau du joueur par sa

capacité principale, c'est à dire sa capacité à tirer pour le challenge 4. Nous contenterons de cette approximation dans la suite de l'expérience.

Nous évaluons la corrélation entre une capacité du joueur et la difficulté du challenge comme nous l'avons définie au chapitre ??, c'est à dire comme la probabilité d'échec par classe de niveau, pour cette capacité donnée. La capacité *Tirer* semble la seule liée au résultat, nous calculons donc la difficulté en fonction de cette capacité et considérons qu'elle est représentative du niveau du joueur. Pour établir les différents niveaux, nous utilisons l'algorithme de classification des k-moyennes, de manière à obtenir 10 classes de niveau. Ces classes sont représentées figure 9.14.

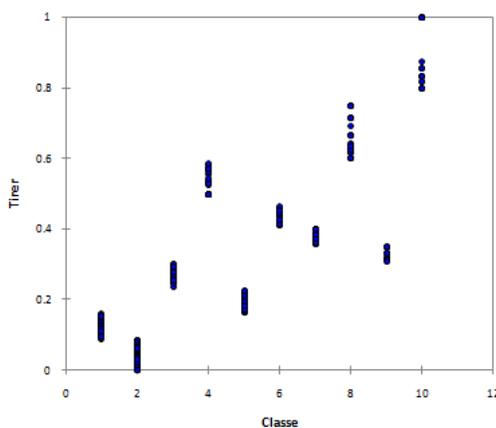


FIGURE 9.14 – Classes de niveau pour la capacité Tirer

Nous avons ensuite évalué les probabilités de réussite par classes, en calculant la moyenne des valeurs de *Résultats* au sein de chaque classe. Nous avons également représenté chaque classe de niveau de précision par la moyenne des valeurs de niveau qui la composent. Nous obtenons une corrélation linéaire de -0.76 entre la probabilité d'échec du joueur et sa capacité à tirer. La régression linéaire correspondante est tracée figure 9.15.

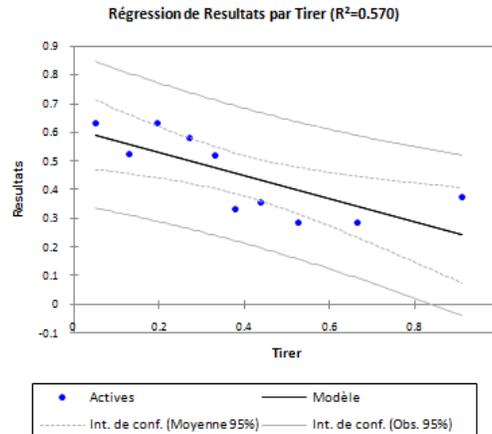


FIGURE 9.15 – Régression linéaire entre précision et difficulté

Notre méthode permet donc, dans le cadre de cette expérience, de mesurer certaines des capacités du joueur et de mettre en évidence une partie du lien qu’elles entretiennent avec la difficulté du gameplay, telle que nous la définissons. Pour le challenge 4, la capacité du joueur à tirer, sa précision, semble la plus déterminante. Dans la section suivante, nous mettons en évidence ce type de lien pour l’ensemble des challenges.

9.3.1.2 Analyse de l’ensemble des challenges

Les dix challenges qu’offrent notre gameplay ne peuvent pas, à la suite de notre expérience, être analysés avec la même précision. La figure 9.10 nous a montré que les challenges 9 et 10, par exemple, n’ont été que très peu joués, et seront donc écartés de cette analyse. En effet, notre échantillon de joueurs est centré autour d’un niveau donné. Notre gameplay s’adapte à ce niveau, en proposant des challenges dont la difficulté oscille plus ou moins fortement autour du niveau du joueur. Les challenges les plus proches du niveau moyen du groupe de joueurs sont donc les plus joués, et certains challenges ne sont quasiment jamais proposés.

La figure 9.16 montre l’ensemble des coefficients de corrélation linéaire calculés entre les variables décrivant les capacités du joueur et la performance obtenue. Les nombres en gras sont ceux pour lesquels le du χ^2 d’indépendance montre que résultats et capacités ne sont pas indépendants (** $p < 0.01$, * $p < 0.05$).

Challenge	Tirer(ρ)	Bouger (ρ)	Changer de zone (ρ)
0	-0.276**	-0.057	0.008
1	-0,184	-0,068	0,019
2	-0,229**	-0,084	-0,007
3	-0.245**	-0.277**	-0.162**
4	-0,215**	-0,111	-0,068
5	-0,307**	-0,058**	0,012**
6	-0,148*	0,080**	0,139**
7	-0,195	0,121*	0,095
8	-0,071	-0,112	0,125

FIGURE 9.16 – Corrélacion capacités et résultat par challenge (Coefficient de Pearson).

Les nombres en gras sont ceux pour lesquels le test du χ^2 d'indépendance montre que résultats et capacités ne sont pas indépendants (** $p < 0.01$, * $p < 0.05$).

Les résultats obtenus semblent montrer que pour 6 challenges sur 9, dont les challenges les plus joués et de difficulté la plus centrale par rapport aux niveaux des joueurs, la précision du joueur telle que nous la calculons est liée au résultat binaire qu'il obtient. A chaque fois, cette relation peut être calculée comme une corrélation linéaire comprise entre -0.148 et -0.307 . Pour les challenges 3, 5 et 6, on observe un lien entre les capacités de mouvement, de déplacement et le résultat binaire obtenu. Ce lien n'est exprimable sous forme d'une corrélation linéaire que pour le challenge 3. Les challenges 1, 7 et 8 ne semblent mettre en évidence quasiment aucun lien entre les capacités calculées et le résultat obtenu, si ce n'est la capacité de mouvement pour le challenge 7.

Les résultats précédents montrent en synthèse qu'il est possible, pour certaines parties du gameplay seulement, de mettre en évidence un lien entre capacités du joueur et résultat obtenu. Nous avons choisit de calculer ce lien comme une corrélation linéaire entre les variables décrivant le niveau du joueur et la variable de résultat. Parmi les liens mis en évidence par le test de χ^2 , certains ne semblent cependant pas modélisables sous cette forme.

Le modèle paraît ainsi incomplet, ne détectant pas de lien significatif entre capacités et résultat pour tout les challenges, et en partie naïf, puisque la corrélation linéaire ne semble pas suffisante pour exprimer les liens mis en évidence par le test de χ^2 . Nous avons donc choisit d'évaluer la pertinence du modèle de manière plus globale, en vérifiant dans quelle mesure celui ci permet de prévoir de *nouveaux* résultats, c'est à dire dans quelle mesure les corrélations établies à partir d'une partie des données permettent de prédire la valeur des résultats de l'autre partie des données.

De manière à évaluer la robustesse du lien entre résultat et comportement du joueur, nous avons construit un modèle de prédiction de la difficulté à partir de régressions linéaires

calculées sur 75% des données récoltées. Nous avons ensuite évalué la correspondance entre le résultat obtenu par le joueur et celui prédit par le modèle, sur les 25% de données restants, c'est à dire 587 traces. Notre algorithme de prédiction utilise de manière naïve les régressions linéaires calculées pour chaque capacité du joueur : chaque capacité donc le coefficient de corrélation est supérieur à 0.1 est utilisée pour prédire la difficulté du challenge. Nous réalisons donc une prédiction par capacité, qu'il nous faut combiner pour réaliser une prédiction globale. Ces prédictions sont combinées sous la forme d'une moyenne, pondérée par leur qualité relative, c'est à dire pour nous par la valeur du coefficient de corrélation de chaque capacité. Ainsi, la prédiction réalisée à partir d'une capacité qui corrèle faiblement avec le résultat du joueur sera moins prise en compte dans la prédiction globale que celle réalisée à partir d'une capacité avec un meilleur coefficient de corrélation.

Nous obtenons une corrélation de 0.48 entre difficulté prédite et résultat obtenus. La figure 9.17 présente les distributions de difficulté prédite pour chaque résultat obtenu par le joueur, pour les 25% des données utilisées pour valider le modèle. Si nous binarisons la prédiction en considérant que toute valeur supérieure à 0.5 est égale à 1, et 0 sinon, alors sur les 25% des données utilisées pour valider le modèle, notre modèle prédit le résultat du challenge avec une précision de 70% (411 prédictions correctes sur 587 résultats). Le modèle semble plutôt pessimiste, la plupart des erreurs étant des prédictions d'échec alors que le joueur a effectivement réussi.

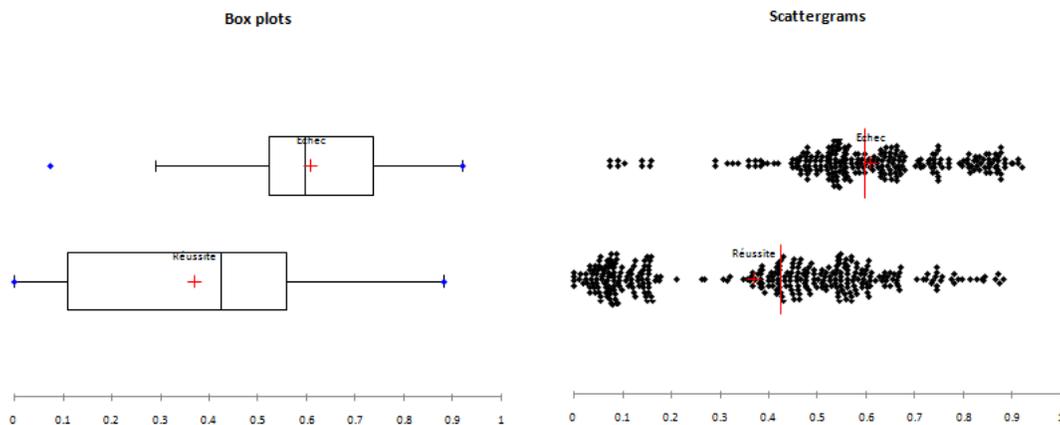


FIGURE 9.17 – Predictions de difficulté pour chaque résultat obtenu sur données de validation.

Ces premiers résultats montrent qu'il est possible de mesurer certaines des capacités du joueur ainsi que de mettre en évidence un lien entre ces capacités et le résultat obtenu. Pour une partie des challenges de cette expérience, une ou plusieurs des capacités du joueur est liée au résultat qu'il obtient lorsqu'il tente de le valider. Si nous exprimons ce lien sous forme

d'un coefficient de corrélation linéaire, alors le modèle est suffisamment robuste pour prédire le résultat du challenge avec une précision de 70%.

9.3.2 Difficulté calculée et difficulté théorique

Nous avons mis au point cette expérience à partir d'une évaluation à priori de la difficulté de chaque challenge. L'objectif principal de notre modèle consiste à permettre au game designer de confronter son évaluation de la difficulté à priori avec le résultat d'une étude expérimentale. Nous avons donc comparé la difficulté calculée par notre modèle à partir des capacités du joueur à la difficulté théorique, que nous avons évaluée comme présenté figure 9.5. La figure 9.18 montre les distributions des différentes valeurs de difficulté calculées, pour chaque niveau de difficulté théorique. Le test d'indépendance χ^2 met en évidence un lien entre difficulté théorique et difficulté calculée ($p < 10^{-4}$). Il existe également une corrélation linéaire de 0.51 entre ces deux difficultés.

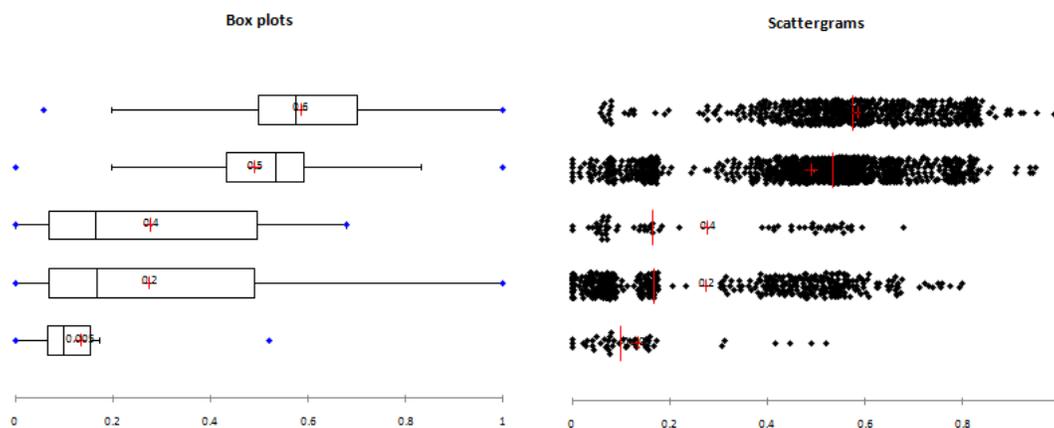


FIGURE 9.18 – Distributions des prédictions de difficulté pour chaque difficulté théorique.

En tant que designer, ces résultats nous permettent de pointer les parties de notre gameplay qui respectent le moins bien la difficulté théorique que nous souhaitons imposer au joueur. Très peu de challenges ont été joués avec une difficulté de 0.05 ou de 0.4, difficulté qui n'est proposée que lors des phases de progression graduelle de la difficulté. En effet, la courbe de suivi propose continuellement une difficulté de 0.5 et la partie oscillante de la courbe optimale propose une difficulté variant entre 0.2 et 0.6. Pour les valeurs les plus hautes, c'est à dire les difficultés de 0.5 et 0.6, la distribution des valeurs observées reste principalement groupée autour de la valeur théorique. Cependant pour la valeur de difficulté la plus basse, 0.2, on remarque que la distribution est plus étalée, avec un premier quartile

à 0.07 et un troisième quartile à 0.49, même si la médiane à 0.16 est proche de la difficulté théorique voulue.

Plusieurs erreurs dans la conception du gameplay peuvent expliquer ce résultat. Parmi ces erreurs, il en est une que nous pouvons étudier sans modifier le gameplay et réaliser une nouvelle expérience. En effet, nous pourrions sûrement améliorer notre gameplay en calculant le niveau du joueur d'une manière un peu plus stable et précise que celle utilisée pendant l'expérience. Comme le montre l'algorithme présenté section 9.2.2, nous calculons le niveau du joueur en fonction de son dernier résultats, sans tenir compte des résultats précédents. Un bon joueur victime de malchance ou à l'inverse, un joueur moins bon qui gagne par coup de chance vont voir leur niveau instantanément mis à jour et être soumis à un niveau de difficulté qui ne correspond pas à celui désiré. Nous avons donc recalculé les niveaux de difficulté théorique, toujours selon la table construite a priori (fig. 9.5) mais en se basant sur un niveau du joueur calculé comme une moyenne des 5 derniers niveaux qui lui ont été attribués. Ce calcul plus *juste* du niveau du joueur devrait permettre d'obtenir une meilleure évaluation de la difficulté théorique à laquelle il a été soumis.

Les résultats obtenus sont les suivants : le test de χ^2 montre qu'il existe un lien entre difficulté théorique et difficulté calculée ($p < 10^{-4}$). Il existe également une corrélation linéaire de 0.61 entre ces deux difficultés, supérieure à celle obtenue précédemment. La figure 9.19 montre les distributions obtenues avec la réévaluation du niveau du joueur.

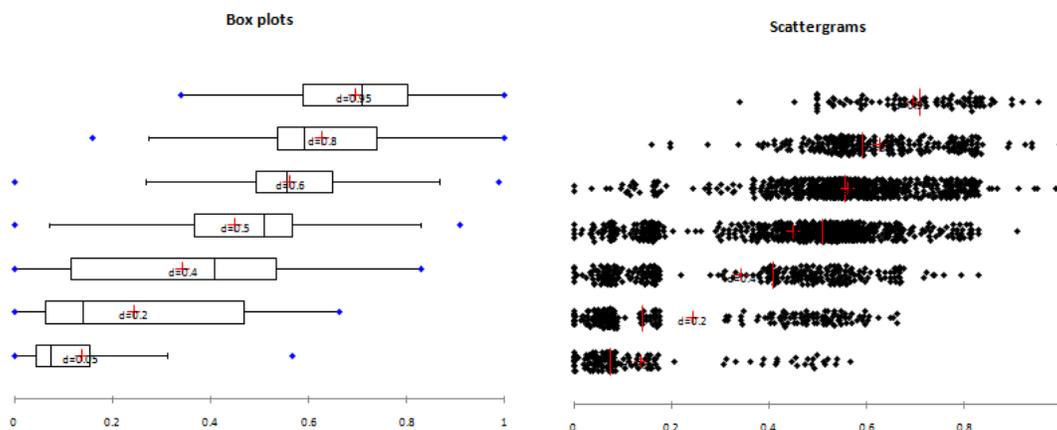


FIGURE 9.19 – Distributions des prédictions de difficulté pour chaque difficulté théorique (niveau du joueur réévalué).

Un calcul plus juste du niveau du joueur rapproche donc la difficulté théorique de la difficulté calculée par notre modèle. Ce résultat nous conforte en partie sur la validité du modèle de calcul de la difficulté : une amélioration de la précision du système de régulation de la difficulté permet d'obtenir des résultats plus proches de la difficulté calculée par notre

modèle. Réciproquement, ces résultats mettent en évidence le type de faiblesse que notre modèle est capable de mettre à jour au sein d'un gameplay.

9.3.3 Difficulté calculée et difficulté subjective

Une autre manière d'évaluer la pertinence de notre modèle consiste à vérifier à quel point la difficulté rapportée par les joueurs correspond à celle calculée par le modèle. La figure 9.20 montre les corrélations linéaires calculées entre la difficulté subjective d'une part, et le résultat obtenu par le joueur, la difficulté calculée par le modèle et la difficulté théorique⁸ d'autre part.

	Diff Subjective
Résultat	0.335**
Diff Calculée	0.318**
Diff Théorique	0.235**

FIGURE 9.20 – Tableau de corrélation de la difficulté subjective.

Le test d'indépendance du χ^2 montre un lien significatif entre les variables dépendantes et explicatives. Exprimé sous forme d'une corrélation linéaire, ce lien est le plus fort pour le résultat obtenu. En effet, on comprend facilement que le résultat final ait un impact sur la perception de la difficulté par le joueur. Ensuite, la difficulté calculée à partir du comportement du joueur corrèle au même niveau avec la difficulté subjective, de manière supérieure à la difficulté supposée à priori.

Ces résultats semblent montrer tout d'abord que la difficulté calculée par le modèle entretient un lien au moins aussi important que le résultat obtenu par le joueur sur son évaluation subjective de la difficulté, ce qui confirme en partie sa validité. Ensuite, ces résultats montrent que ce lien est plus important que celui entretenu par notre évaluation a priori de la difficulté, et que notre modèle nous permet donc d'évaluer une difficulté plus proche du sentiment du joueur que celle que nous avons évalué au départ.

Plus généralement, ces résultats montrent également que notre modèle ne permet pas d'expliquer totalement l'évaluation de la difficulté faite par le joueur. La figure 9.21 montre les distributions de difficulté calculées pour chaque niveau de difficulté subjective rapporté par le joueur. Ces résultats semblent montrer que l'évaluation de la difficulté par le joueur suit un processus plus complexe que la simple observation du résultat obtenu, ou du résultat qu'il aurait statistiquement du obtenir étant donné son comportement.

8. Cette difficulté théorique prend en compte les améliorations de la section 9.3.2 pour le calcul du niveau du joueur.

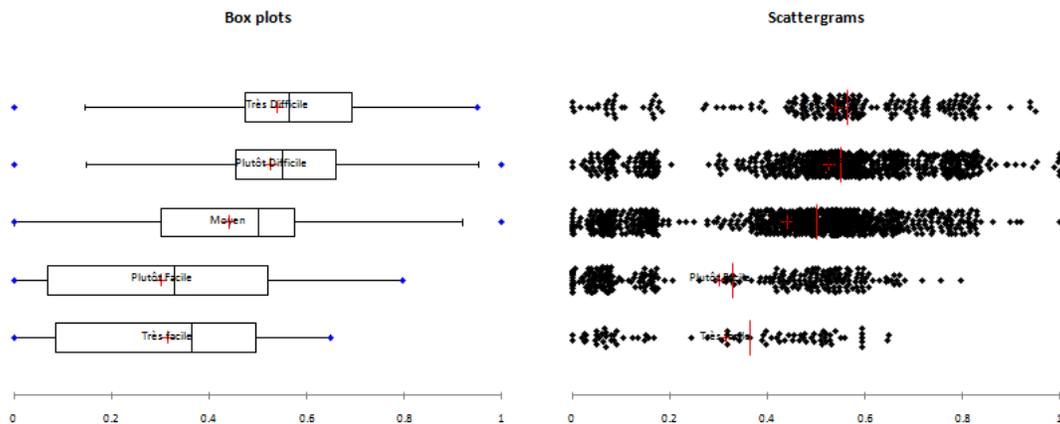


FIGURE 9.21 – Distributions des prédictions de difficulté pour chaque niveau difficulté subjective.

9.3.4 Plaisir de jouer et niveau de difficulté

Au delà de l'évaluation de notre modèle, cette expérience a pour but d'étudier la validité des hypothèses formulées au chapitre ?? et rappelées dans la section 9.1 de ce chapitre quant à la forme optimale d'une courbe de difficulté. Tout d'abord, il est intéressant d'étudier le lien qu'il existe entre le plaisir rapporté par les joueurs et la difficulté du jeu. Selon la théorie du *flow*, plus la difficulté s'écarte du niveau du joueur, c'est à dire plus le jeu devient trop dur ou trop difficile, moins le joueur est satisfait. Nous avons donc calculé le coefficient de corrélation linéaire existant entre le plaisir du joueur d'une part, et la valeur absolue des écarts de difficulté subjective, prédite et théorique par rapport à une difficulté équilibrée, c'est à dire 0.5, d'autre part. On notera cette valeur δD pour une difficulté quelconque D avec $\delta D = \text{abs}(0.5 - D)$. Selon la théorie du *flow*, plus δD est réduit et plus le joueur doit être satisfait.

La figure 9.22 montre les corrélations entre difficulté et plaisir du joueur, entre δD et plaisir du joueur, ainsi que les cases pour lesquelles le test du χ^2 a indiqué une dépendance significative. Ces résultats nous montrent que seule la difficulté subjective semble entretenir une relation linéaire avec le plaisir rapporté par le joueur. La mesure objective de difficulté ainsi que la difficulté théorique semblent être liées au plaisir du joueur de manière significative, mais ce lien n'est pas mis en évidence par le calcul d'un coefficient de corrélation. On peut donc tout d'abord uniquement noter que lorsque le joueur ressent une difficulté plus importante, son plaisir augmente légèrement.

	Plaisir		Plaisir
D Subjective	0.233**	δD Subjective	0.238**
D Prédite	0.069*	δD Prédite	-0.060*
D Théorique	0.066**	δD Théorique	-0.071**

FIGURE 9.22 – ρ entre plaisir de joueur et difficulté.

Les résultats en gras sont ceux pour lesquels le test du χ^2 d'indépendance montre qu'écart de difficulté et plaisir ne sont pas indépendants (**p<0.01, *p<0.05).

Les distributions de la figure 9.23 permettent de visualiser plus aisément la relation entre difficulté subjective et plaisir du joueur. On note que plus le plaisir du joueur est important, plus la difficulté est fréquemment évaluée comme importante ce qui est cohérent avec le coefficient de corrélation de 0.23 obtenu précédemment.

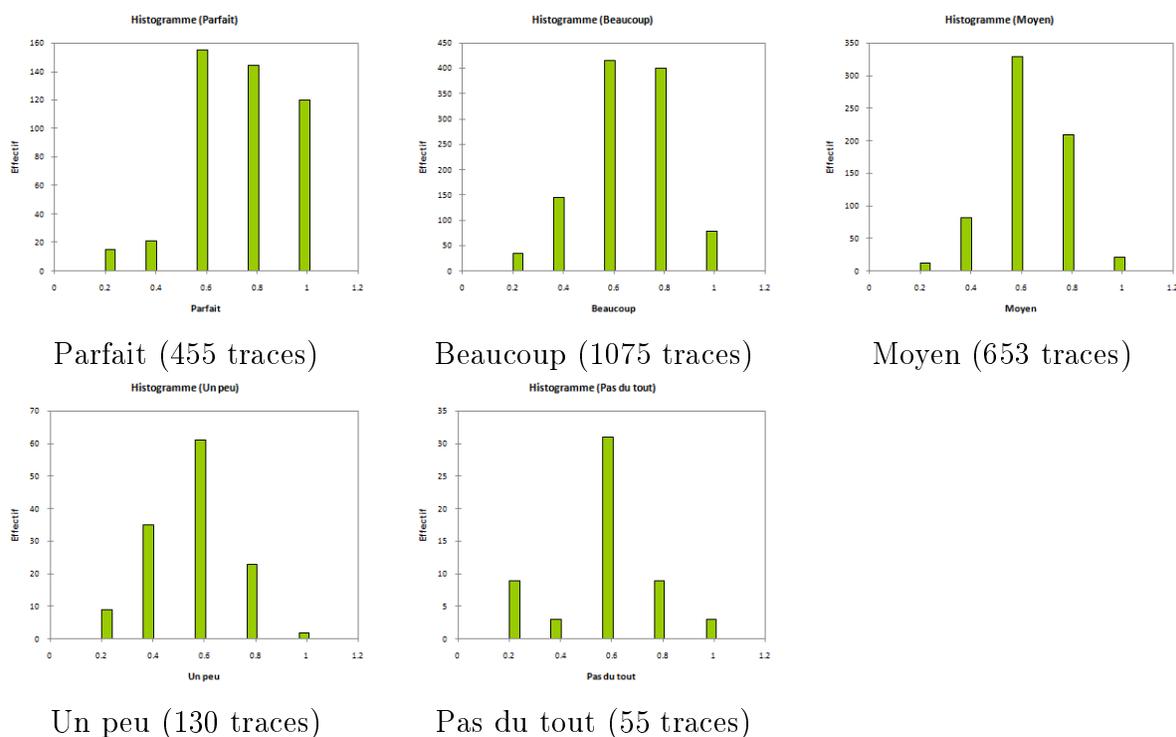


FIGURE 9.23 – Distributions des difficultés subjectives pour chaque niveau de plaisir.

Lorsqu'on étudie les distributions des difficultés objectives pour chacun des niveaux de plaisir (fig 9.24), on constate cependant le phénomène suivant : lorsque le joueur n'est pas

insatisfait, ce qui arrive dans la très grande majorité des cas⁹, on remarque qu'une partie importante de la distribution est centrée sur la difficulté équilibrée de 0.5. On remarque également que plus on s'éloigne de la note *Parfait*, plus la distribution présente un deuxième pic, centré sur les faibles difficultés. Puis lorsque le joueur s'avoue légèrement voire très insatisfait, le premier pic est plus important et le second se déplace légèrement vers la droite.

A la différence de la difficulté subjective, la distribution ne semble réellement évoluer que lorsque le joueur est insatisfait : pour les notes *Parfait*, *Beaucoup* et *Moyen*, le gros de la distribution reste centré sur 0.5. Par contre, on constate que la distribution évolue plus franchement pour les satisfactions *Un peu* et *Pas du tout*. Une corrélation linéaire calculée sur l'ensemble des données capture difficilement cette relation car les joueurs ont très rarement rapporté être insatisfaits.

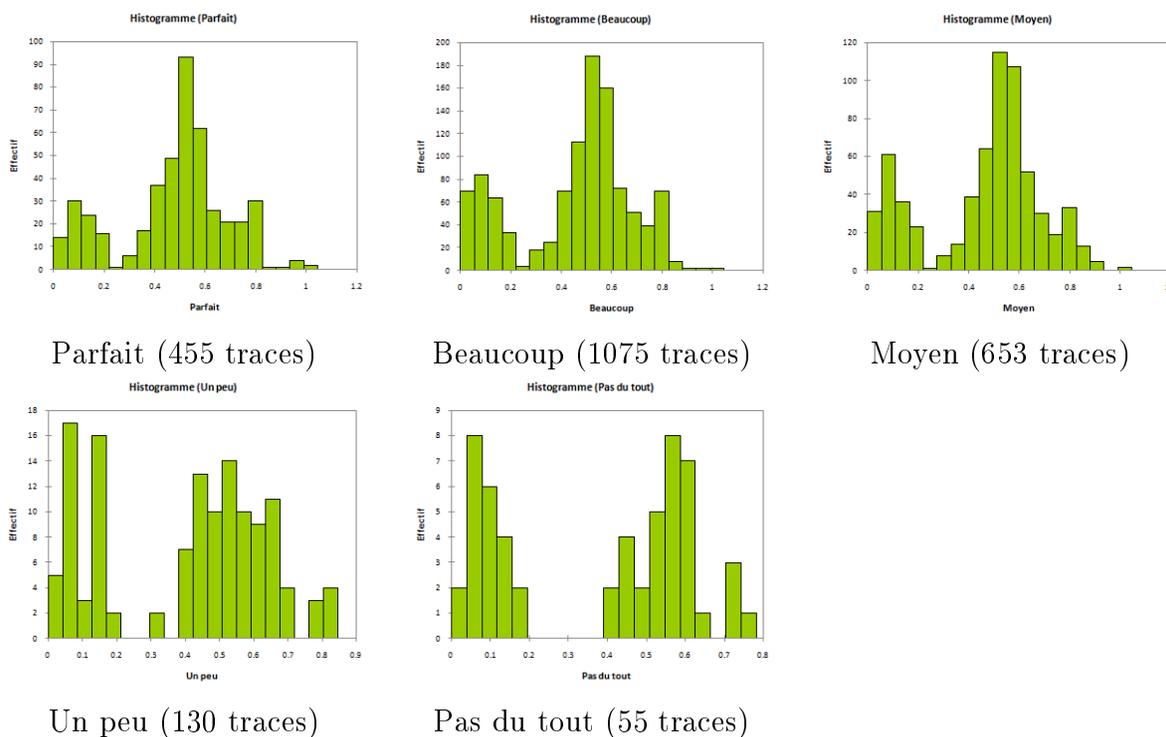


FIGURE 9.24 – Distributions des difficultés calculées pour chaque niveau de plaisir.

Si on calcule, pour chaque niveau de plaisir du joueur, la moyenne de la valeur absolue des écarts de la difficulté par rapport à 0.5, c'est à dire à quel point la difficulté s'écarte,

9. 92% des cas si l'on considère que lorsque le joueur répond *Moyen*, il n'est pas insatisfait.

en moyenne, d'une valeur équilibrée, alors la corrélation linéaire avec le plaisir est de -0.96 , mais sur seulement cinq points. La régression linéaire est présentée figure 9.25.

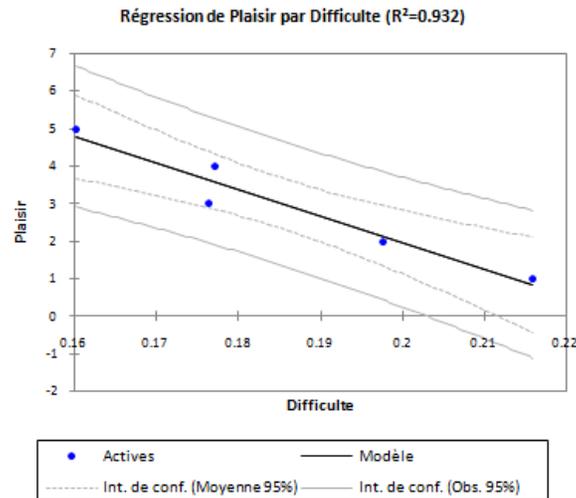


FIGURE 9.25 – Régression linéaire entre écarts de difficulté objective et plaisir.

On peut donc, en définitive, interpréter ces résultats de la façon suivante :

- On ne peut pas dire que plus la difficulté objective s'écarte de la valeur d'équilibre, moins le joueur apprécie le jeu. En l'occurrence, dans de très nombreux cas, le joueur a dit apprécier le jeu alors que la difficulté était effectivement déséquilibrée.
- On peut par contre dire que plus le joueur est insatisfait, plus les valeurs de difficultés semblent s'écarter, en moyenne, de l'équilibre. Les joueurs ont été rarement insatisfait, mais les valeurs de difficulté étaient plus extrêmes dans ce cas. De plus, il semble que ce soient les difficultés les plus basses qui évoluent le plus. Si le pic des difficultés fortes s'est légèrement déporté vers les hautes valeurs, celui des difficultés les plus basses a bien plus évolué.

On observe également que la difficulté subjective semble plus directement liée au plaisir du joueur que la difficulté objective. En effet, on observe directement une corrélation de 0.24 entre l'écart de difficulté subjective et le plaisir du joueur. De plus la corrélation positive de 0.23 entre difficulté subjective et plaisir indique que lorsque le joueur est satisfait, il indique plus souvent une difficulté importante.

9.3.5 Plaisir de jouer et courbe de difficulté

Tout au long de ce chapitre, nous avons rappelé que cette expérience pouvait nous permettre de tester nos hypothèses quand à la forme optimale d'une courbe de difficulté. Nous supposons donc en effet qu'une courbe respectant les caractéristiques suivantes est optimale :

- une première phase de progression graduelle,
- une seconde phase de forte variabilité,
- les challenges les plus difficiles doivent à peine dépasser le niveau du joueur.

Nous avons donc soumis les joueurs à deux phases de jeu de 12 minutes, chacune suivant une courbe de difficulté particulière, et ce dans un ordre aléatoire. La courbe de *Suivi* est une courbe standard, qui cherche à suivre le plus précisément possible le niveau du joueur. La courbe d'*Oscillation* respecte quand à elle les caractéristiques que nous considérons comme celles d'une courbe optimale. Durant chacune de ces deux périodes, les joueurs ont régulièrement reporté leur niveau de plaisir. La seule modification du gameplay entre ces deux phases est le changement de courbe de difficulté, ce qui nous permet donc d'évaluer l'impact de chaque type de courbe sur le plaisir du joueur.

Un test de χ^2 d'indépendance montre qu'à $p = 10^{-3}$, on peut accepter l'hypothèse selon laquelle le type de courbe influence le plaisir du joueur. La corrélation linéaire calculée entre le type de courbe et le plaisir du joueur n'est que de 0.075. La figure 9.26 donne la moyenne et l'écart type du plaisir du joueur pour chaque type de courbe de difficulté. La figure 9.27 donne les distributions du plaisir du joueur, pour chaque type de courbe de difficulté.

	Oscillation	Suivi
Moyenne	3.67	3.81
Ecart-Type	0.90	0.91

FIGURE 9.26 – Plaisir du joueur pour chaque type de courbe de difficulté.

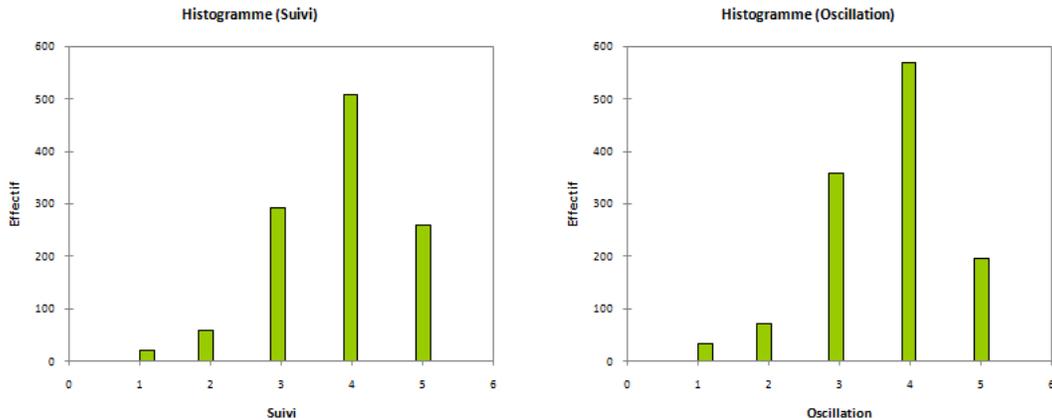


FIGURE 9.27 – Distributions du niveau de plaisir pour chaque type de courbe de difficulté.

Les différences entre les deux distributions sont assez minces, mais significatives selon le test du χ^2 . Elles vont à l’encontre de nos hypothèses : la moyenne du plaisir du joueur est plus faible dans la situation que nous considérons comme optimale. En observant les distributions, on remarque que pendant les phases d’oscillation, les joueurs ont moins répondu *Parfait* (5) et plus souvent *Beaucoup* (4) ou *Moyen* (3). En conséquence, ces données contredisent nos hypothèses et montrent que dans cette expérience, les joueurs ont préféré la courbe de suivi à la courbe d’oscillation.

9.4 Discussion et Conclusion

Nous avons tout d’abord évalué si notre modèle était capable, pour un gameplay représentatif d’une grande partie des jeux commerciaux produits aujourd’hui, d’évaluer les capacités du joueur. A partir d’un enregistrement d’évènements survenus au cours de la session du jeu, nous avons été en mesure de mesurer ces capacités. Nous avons ensuite pu constater que pour 7 challenges sur 9, dont les challenges les plus joués, il existait un lien entre les capacités observées et le résultat obtenu par le joueur. Nous ne sommes par contre pas capable d’expliquer pourquoi certains challenges ne semblent pas exploiter certaines capacités. Ces résultats montrent que le joueur influe effectivement sur le résultat des challenges, mais surtout que nous pouvons mesurer en partie sa capacité à le faire.

Nous avons ensuite choisit de représenter ce lien sous forme d’un modèle linéaire. Ces régressions linéaires nous permettent de prédire une probabilité d’échec du joueur en fonction de son niveau, c’est à dire de la mesure de ses capacités. En combinant les prédictions réalisées pour chaque capacité en les pondérant par leur qualité supposée, nous parvenons à prédire le résultat du challenge avec une fiabilité de 70%, sur 25% des données conservées uniquement

pour la validation. Ce test montre que notre modèle linéaire naïf mesure bien une propriété du gameplay que nous avons construit, et non un artefact particulier des données recueillies. Il serait maintenant particulièrement intéressant d'étudier la performance de modèles non linéaires plus complexes, comme par exemple les réseaux de neurones, à prédire le résultat du joueur. Il est de plus intéressant de noter que le modèle semble avoir plus de mal à prédire le résultat des challenges les plus faciles. Nous pouvons supposer qu'il est effectivement plus difficile de prédire le résultat des challenges les plus simples, et donc soumis au joueurs les moins expérimentés, au cours desquels le comportement du joueur est désorganisé, incohérent et dont le résultat repose peut être sur une part de chance plus importante que pour les challenges de niveau plus élevé.

Nous avons pu vérifier que la difficulté théorique, c'est à dire évaluée à priori, correspondait en partie à celle calculée par le modèle. Les deux difficultés sont liées selon le test de χ^2 d'indépendance, et ont une corrélation linéaire de 0.51. De plus, l'étude de la relation entre difficulté calculée et difficulté théorique nous permet d'envisager diverses améliorations. Par exemple, les valeurs théoriques de difficulté les plus hautes (fig. 9.19) semblent correspondre à des difficultés calculées plus faibles. Nous devrions surement augmenter la différence de niveau entre joueurs et ennemis pour atteindre effectivement une difficulté de 0.95. Nous avons également pu étudier une amélioration apportée au calcul du niveau du joueur, que nous avons choisit de moyennner sur une fenêtre glissante. Cette amélioration du calcul du niveau du joueur a rapproché les valeurs de difficulté théoriques des valeurs calculées par le modèle, ce qui souligne la qualité des résultats qu'il fournit.

L'étude de la difficulté perçue par le joueur confirme également la pertinence de notre modèle. Il existe une relation significative entre difficulté subjective et calculée, et leur corrélation linéaire est de 0.32. On remarque cependant que pour les valeurs hautes et basses, l'évaluation de la difficulté par le joueur semble plus complexe que celle que fournit notre modèle. Le joueur note comme *Très Difficile* une distribution de traces qui ne semble pas très différente de celle notée comme *Plutôt Difficile*. Les distributions des traces notées *Très Facile* et *Plutôt Facile* semblent plus décalées, mais dans le mauvais sens. Il serait donc particulièrement intéressant d'étudier plus précisément de quelle manière le joueur attribue des notes de difficulté très élevées ou très basses.

L'étude du plaisir rapporté par le joueur en fonction de la difficulté relative à laquelle il est soumis confirme la théorie du flow d'une manière intéressante. Les difficultés subjectives et calculées ne semblent pas être liées de la même manière au plaisir du joueur. Le joueur semble en effet plus satisfait lorsqu'il évalue une difficulté plus importante : la distribution des difficultés subjectives des tentatives correspondant à un plaisir *Parfait* est principalement répartie sur les valeurs *Moyenne*, *Plutôt Difficile* et *Très Difficile*. Par contre, lorsqu'on observe les niveaux de difficulté calculée, on remarque que lorsque le joueur est mécontent, la difficulté calculée se répartit sur les valeurs très faibles et moyennement hautes, alors qu'elle reste principalement centrée sur 0.5 lorsque le joueur est satisfait. En synthèse, dans notre

expérience, il semble que le joueur apprécie une difficulté objective adaptée et subjective haute, et apprécie moins les difficultés objectives et subjectives faibles. Il serait d'autant plus intéressant de comprendre comment le joueur évalue les valeurs importantes de difficulté, de manière à pouvoir exploiter ces résultats : fournir une difficulté réelle adaptée tout en simulant une difficulté importante. Il est possible que la désaffection du joueur pour les difficultés basses s'explique également par une baisse des rétributions intrinsèques : pour créer un challenge simple, nous avons limité le nombre d'ennemis. Cette méthode limite sûrement également le plaisir du joueur, car le faible nombre d'ennemis signifie clairement la baisse de difficulté et limite les occasions de signifier la réussite du joueur et donc de le récompenser, ce qui arrive à chaque fois qu'il tue un ennemi.

Finalement, l'étude du plaisir du joueur en fonction du type de courbe de difficulté semble contredire nos hypothèses. La courbe de *Suivi* semble être légèrement plus appréciée que celle d'*Oscillation*. On peut expliquer ces résultats à la lumière des résultats précédents. Tout d'abord, la courbe de *Suivi* n'est pas une courbe de suivi parfaite, mais oscille autour du niveau du joueur en s'y adaptant constamment. Ensuite, la relation entre difficulté relative et plaisir du joueur semble montrer que le joueur n'apprécie pas les niveaux de difficulté trop bas, ce que notre courbe d'oscillation propose régulièrement. De plus, notre modification de la difficulté pourrait être améliorée pour permettre une modification de la difficulté objective plus difficile à percevoir, en gardant un nombre d'ennemis constant mais en modifiant leur niveau. Globalement, la courbe de suivi respecte en grande partie nos hypothèses d'optimalité : la phase de tutorial préliminaire à l'expérience permet de toute façon une phase d'ascension de la difficulté, et le suivi offre une oscillation qui ne dépasse que de très peu le niveau du joueur. De plus la courbe de *Suivi* ne lui propose pas de challenges objectivement et subjectivement trop faciles. Il serait particulièrement intéressant de réaliser à nouveau cette expérience, mais avec la courbe d'*Oscillation* respectant les hypothèses suivantes :

- une première phase de progression graduelle,
- une seconde phase de variabilité moyenne,
- les challenges les plus difficiles doivent à peine dépasser le niveau du joueur mais paraître plus difficiles.
- les challenges les plus faciles ne doivent pas paraître trop simples ni limiter les occasions d'afficher la réussite du joueur.

Nous ne sommes donc pas réellement en mesure de tester l'hypothèse selon laquelle une courbe de suivi est moins efficace qu'une courbe d'oscillation, puisqu'une courbe de suivi est aujourd'hui impossible à mettre en place. Les recherches effectuées sur les algorithmes d'adaptation de la difficulté le permettront cependant peut être un jour. En attendant, nous pouvons néanmoins essayer de construire une oscillation optimale, à partir des modifications apportées à nos hypothèses.

Glossary

First Person Shooter First Person Shooter, jeu de tir en vue subjective.. 3, 4

Intelligence Artificielle Intelligence Artificielle.. 5, 8

Niveau de jeu Le niveau de jeu désigne l'espace dans lequel évolue le joueur, plus particulièrement lorsque l'univers du jeu est présenté au joueur de manière clairement segmentée, sous espace après sous espace.. 4, 8

Bibliographie

[Csikszentmihalyi 91] Mihaly Csikszentmihalyi. Flow : The psychology of optimal experience. Harper Perennial, March 1991.