

Gabriel Sauvage

UFR Ingémédia, Licence Pro TSI, Parcours NTS

Université de Toulon

Mémoire de Licence Professionnelle

**Le rapport entre le son et l'état de Flow
dans l'expérience vidéoludique**

Travail réalisé sous la direction de Jean-Michel DENIZART

Mois et année de soutenance: Juillet 2021

Gabriel Sauvage

UFR Ingémédia, Licence Pro TSI, Parcours NTS

Université de Toulon

Mémoire de Licence Professionnelle

**Le rapport entre le son et l'état de Flow
dans l'expérience vidéoludique**

Travail réalisé sous la direction de Jean-Michel DENIZART

Mois et année de soutenance: Juillet 2021

Le rapport entre le son et l'état de Flow dans l'expérience vidéoludique

Résumé (2000 caractères)

Ce mémoire se concentre sur le rapport entre le son et l'état de Flow au sein d'une expérience vidéoludique. Le "flow", un état psychologique théorisé par Mihály Csikszentmihályi en 1975 puis notamment exploré dans son ouvrage "Flow: The Psychology of Optimal Experience", constituerait selon l'auteur "l'expérience optimale", celle qu'un être humain recherche au travers de toute pratique ou activité quelle qu'elle soit. L'atteinte fréquente de cet état serait alors directement corrélée au bonheur et au bien être psychique d'un individu.

Dans ce mémoire, j'explore la manière dont les concepteurs sonores contribuent à recréer cet état de flow dans l'expérience vidéoludique, au travers de sa direction sonore. Tout d'abord, je réalise une introduction au flow et entreprends de définir cet état. Puis, je m'emploie à souligner la distinction entre "immersion" et "flow", deux états différents et pourtant interconnectés dans leur fonctionnements et conditions d'apparition. Suite à cela, j'explique en quoi l'aspect sonore d'un jeu constitue le pilier de l'apparition de l'état de flow, et comment l'intégration de la conception sonore dans le processus de game design, ainsi que certaines utilisations des caractéristiques d'un son, permettent de faciliter l'apparition de l'état de flow. Je parle ensuite de la relation intrinsèque entre le "game feel", le son et le flow. Puis, j'aborde la question de la nécessité de l'interactivité sonore vis-à-vis de l'apparition de l'état de flow, via (notamment) l'utilisation de la musique, l'interprétation des données de jeu en temps réel, l'agencement des matériaux sonores sous forme de systèmes proactifs et réactifs, et enfin des questions budgétaires liées à la conception de ces systèmes. J'explore par la suite les différents dangers de rupture du Flow liés à la direction sonore, et conclus par la question suivante: "Sans l'image, le son suffit-il à atteindre le flow?", avant de procéder à une conclusion.

Mots clés: Flow, Jeu Vidéo, Expérience Vidéoludique, Sound Design, Composition, Conception Sonore, Interactivité, Interactivité Sonore, Game Feel, Wwise

Remerciements

Merci à Arnaud Roy, Masami Komuro, Joonas Turner, Christophe Héral, Paul Weir et enfin Bjorn Jacobsen, pour m'avoir, à tour de rôle, conseillé, aiguillé et apporté de solides ressources ayant grandement contribué à la conception de ce mémoire. Merci également à Jean-Michel Denizart, mon directeur de mémoire, qui m'a aidé à concevoir le fil conducteur de mes entretiens avec les membres du corps professionnel cités plus haut, et aiguillé tout au long de ce travail.

Table des matières

Préambule

1. Le Flow, introduction et définition

2. Distinction entre "Immersion" et "Flow"

3. Le Son, pilier du Flow

A - Les problématiques techniques et humaines liées à la réalisation de la conception sonore d'un jeu vidéo

B - La direction sonore conçue comme partie intégrante du game design

C - L'utilisation des caractéristiques sonores comme vecteur de flow

4. Game Feel, Flow, Son: une relation intrinsèque

A - Définition du "Game Feel"

B - La conception sonore au centre du game feel

C - L'identification à l'avatar comme moteur de flow

5. L'implémentation et la conception de l'interactivité sonore, ciment du Flow

A - L'implémentation de transitions fluides et naturelles

B - L'utilisation de la musique interactive comme vecteur de flow

C - L'interprétation des données de jeu comme moteur de l'interactivité

D - La nécessité de concevoir une implémentation modulaire et flexible

E - L'optimisation de l'implémentation sonore et de ses systèmes

F - Le budget, obstacle à une intégration sonore complexe selon l'échelle des productions

G - La dé-corrélation entre complexité et efficacité d'une implémentation sonore

6. Les dangers de rupture du Flow liés à la direction sonore

A - L'abus de stimuli sonores et la peur du vide (parler de silent hill)

B - La question de la latence numérique, séparant l'input de sa réponse sonore

C - L'ignorance des phénomènes psychoacoustiques

D - Le concept de score de nuisance et de l'unité de mesure "NUFS"

E - Les problématiques de décorrélation entre un événement sonore et son intention

7. Sans l'image, le son suffit-il à atteindre et maintenir l'état de Flow?

Conclusion

Introduction

Ce mémoire est un travail réalisé dans le cadre d'une licence professionnelle à l'Université de Toulon, au sein du parcours Nouvelles Technologies du Son. A une époque où le potentiel d'attention d'un individu moyen est en chute libre, et où tout type de média lutte pour s'approprier et maintenir l'intérêt du spectateur/lecteur/joueur, il convient de s'intéresser aux phénomènes et outils permettant de le conserver. Depuis les prémices du cinéma et de la diffusion télévisuelle, on a pu constater que le rapport image son constituait l'un de ces outils. Mais naturellement, le jeu vidéo se démarque ici par sa dimension interactive, permettant d'adopter un point de vue et une posture impossibles dans toute autre forme de média. A l'heure où le jeu vidéo vient de souffler sa 46e bougie (si l'on retrace l'origine de sa démocratisation avec la mise en vente de l'Atari Pong en 1975), il reste malgré tout un média très récent. Les questions de perception d'une œuvre vidéoludique chez le joueur, ainsi que les mécanismes mis en place afin de susciter un intérêt chez ce dernier, constituent autant de sujets remplis de zones d'ombres, attendant d'être éclaircies. Pour comprendre la relation entre le joueur et le jeu vidéo, il faut alors remonter à ce qui lie l'humain au divertissement, jusqu'à venir interroger l'expérience sensorielle qu'il constitue.

A travers ce mémoire et les réponses de professionnels ayant été interrogés sur le sujet, nous tenterons de comprendre le phénomène du flow, son application dans une expérience vidéoludique et comment l'utilisation du son peut faciliter, voir susciter son apparition. Dans un premier temps, nous nous concentrerons sur la définition du flow en tant que tel, et son application dans le domaine du jeu vidéo; dans un second temps, nous tenterons de discerner l'état de Flow de l'Immersion; puis nous explorerons les différentes spécificités, méthodes et dangers liés à l'utilisation du son dans une optique d'apparition et maintien de l'état de flow, avant de poursuivre par diverses études de cas dans un but illustratif. Les opinions et idées avancées dans ce travail seront, sauf citation, miennes, et n'engageront que son auteur.

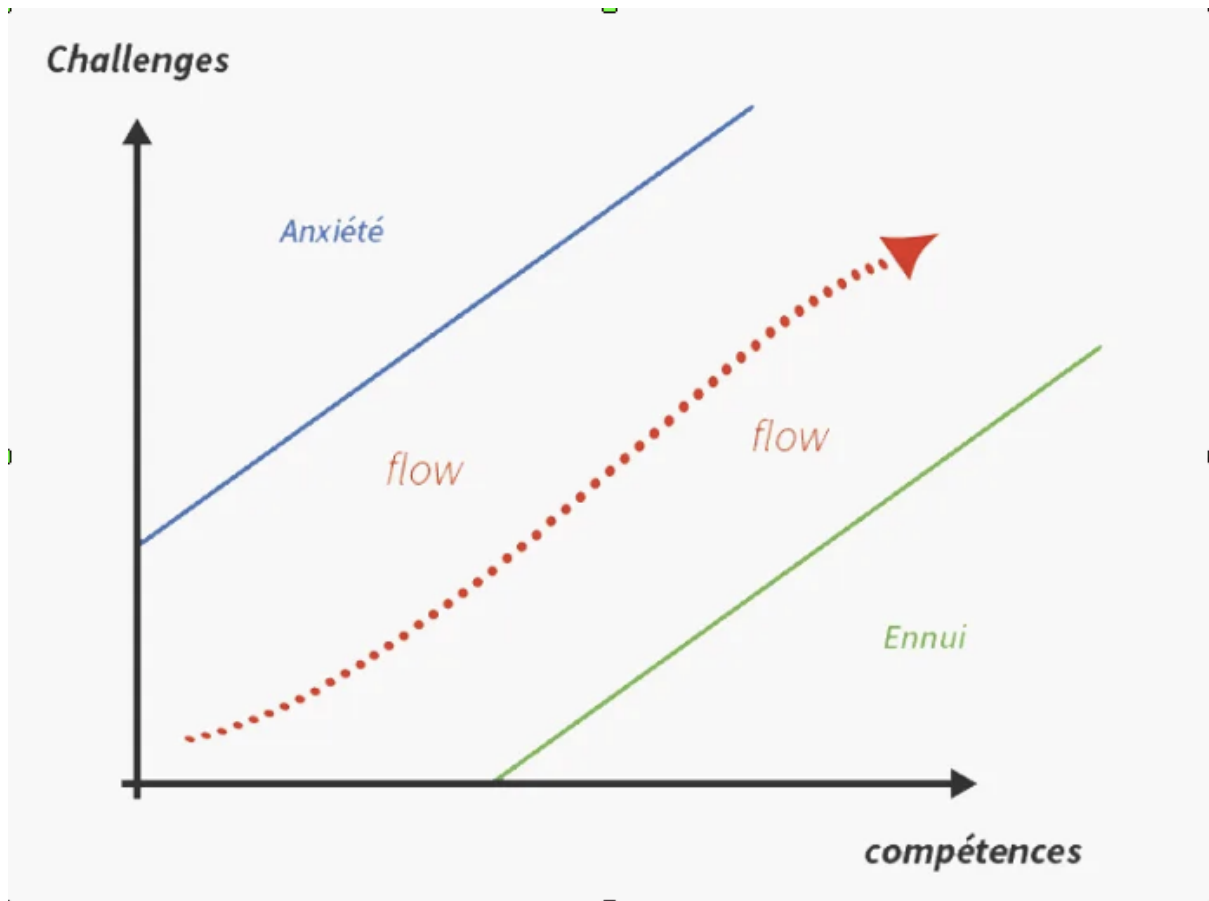
1. Le Flow, introduction et définition

Une propriété comme celle de l'hystérèse pourrait être rapprochée de l'état de flow: hystérèse, du grec ὑστερος/*hústeros* (soit « après », « plus tard »), est la propriété d'un système dont l'évolution ne suivrait pas le même chemin selon qu'une cause extérieure augmente ou diminue. Beaucoup de concepteurs de jeux utilisent le terme "flow" afin de désigner un flot d'informations véhiculées comme un processus de communication au joueur, auquel celui-ci répondrait d'une manière ou d'une autre. Or, cette conception vient se heurter à ce que constitue en réalité le flow, ou plutôt "l'état de flow". Théorisé et nommé par Mihály Csikszentmihályi en 1975, le concept de "flow" est caractérisé comme un état constitué de 8 composantes précises (bien que la présence de la totalité de ces composantes ne soit pas requise pour qu'un état de flow soit vécu):

- Un challenge qui requiert une ou plusieurs compétences.
- La fusion entre l'action et la conscientisation de celle-ci.
- La clarté de l'objectif en cours.
- Une sensation directe de feedback.
- Une concentration intense sur la tâche en train d'être effectuée.
- La sensation de contrôle permanent.
- La perte de la conscience du Soi*.
- Et enfin, une déformation de la perception temporelle.

La notion de challenge et de performance est donc intrinsèquement liée à l'apparition de ce phénomène: si la difficulté d'une tâche entreprise est trop importante par rapport aux compétences du sujet, ce dernier sera confronté à un sentiment d'anxiété; à l'inverse, si cette difficulté est trop faible, l'ennui guettera le sujet. Selon Mihaly, l'état de flow absolu nécessite donc un équilibre entre les compétences du sujet et la difficulté de la tâche en cours afin d'apparaître.

**"Soi", ou "Moi", théorisé par Freud comme la partie visible, consciente d'un individu; celle à laquelle il s'assimile et se définit.*



Toutefois, une certaine tolérance existe, même si elle est difficilement mesurable en raison de la pluralité de profils des joueurs: s'approcher plus ou moins de l'ennui ou de l'anxiété sans l'atteindre (dans le contexte imagé de ce graphique) ne crée pas la disparition du flow; elle augmente par contre les risques de diminuer son intensité, voir le faire disparaître. Ainsi, pour Mihaly, l'état de Flow constitue "l'expérience optimale", celle qu'un être humain recherche au travers de toute pratique ou activité quelle qu'elle soit. L'atteinte fréquente de cet état serait donc, selon lui, directement corrélé au bonheur et au bien être psychique d'un individu.

Si cet état de Flow apparaît donc couramment au cours d'actions effectuées dans la vie réelle, il doit être recréé artificiellement dans le contexte d'une œuvre vidéoludique. La capacité de concentration de chaque joueur sur une période prolongée étant différente pour chaque profil, et toute session de jeu étant obligatoirement vouée à prendre fin, faire apparaître et maintenir l'état de flow le plus longtemps possible permet en revanche de maximiser le potentiel de plaisir et d'investissement pouvant être ressenti par le joueur au cours de l'expérience vidéoludique. Dans la thèse de Jenova Chen nommée "Flow in Games" (transposition des travaux de Csikszentmihalyi au domaine du jeu vidéo, et qui sera citée plusieurs fois lors de ce travail), l'auteur décrit 3 conditions permettant l'apparition de l'état de Flow au sein d'un jeu vidéo:

1. Tout d'abord, le jeu doit être intrinsèquement digne d'intérêt, et le joueur doit s'y essayer de son plein gré.
2. Le jeu doit offrir un challenge équilibré, et adéquat aux capacités du joueur, qui lui permettent de plonger profondément dans l'expérience.
3. Le joueur doit ressentir une sensation de contrôle personnel au travers de son expérience.

Toujours selon Chen, la résultante de ces 3 éléments ferait perdre la notion du temps et de pleine conscience au joueur (ou perte du Soi, évoquée plus haut). Au fil du temps, le concept de difficulté adaptative (ou plus communément appelée DDA, pour Dynamic Difficulty Adjustment) a fini par apparaître, consistant à calibrer la difficulté de l'expérience en se basant sur une mesure de la performance et des compétences du joueur. Cependant, continue l'auteur, l'utilisation de la difficulté adaptative "viendrait grandement diminuer le contrôle des game designers sur l'expérience", rendant l'utilisation de ce système un peu trivial et plutôt rare dans le spectre vidéoludique actuel. Ainsi, un système de difficulté adaptative à lui seul ne pourrait être la solution à l'apparition de l'état de flow, là où un système se basant sur les éléments principaux d'un jeu (la sensation de feedback, le game design, etc.) serait plus direct et utile afin d'obtenir l'effet recherché.

Toutefois, si susciter le flow au sein d'une expérience constitue un challenge en soi, la rétention de ce phénomène chez le joueur est d'autant plus importante: dans une ère où l'on incite parfois le joueur à rester "captif" d'un même titre durant plusieurs centaines d'heures, il n'est pas rare que les jeux à gros budgets, et notamment les mondes ouverts, deviennent de moins en moins finissables pour une majorité de consommateurs. Si ces derniers peuvent tolérer une légère frustration ou moment de flottement, la clé d'une expérience optimale serait donc de maintenir le sujet dans la "zone de Flow". Cependant, chaque profil est différent, et possède sa propre zone en fonction de ses compétences. "Afin de produire un contenu pouvant être apprécié par une majorité, il conviendrait alors d'étendre le panel d'expériences potentielles présentes dans un jeu vidéo, afin de rentrer en adéquation avec les différents profils de joueurs" (Jenova Chen). Les mondes ouverts, de par leur conception intrinsèque, ainsi que via leur boucle de gameplay, proposent généralement un choix varié de "jeux dans le jeu" (capture de territoires, puzzles, et autres activités annexes), permettant au joueur d'adapter son expérience en fonction de ses besoins et désirs.

Un des procédés permettant d'anticiper ou d'adapter des systèmes de jeu au joueur en fonction de ses actions consiste à mesurer ces dernières sous forme de données, et de prévoir des réactions ou modifications en jeu en fonction de la variation de celles-ci. Par exemple: en fonction de la barre de vie du joueur, la musique pourrait

s'intensifier ou se calmer, afin d'offrir un feedback plus percutant qu'un simple changement visuel sur une interface. Cependant, des données plus complexes (comme la notion de challenge perçu par le joueur), sont sujettes à l'interprétation, et donc potentiellement en contradiction vis-à-vis de l'expérience utilisateur réelle. Cela limite donc l'utilisation de ces données comme source d'une altération de gameplay ou de feedbacks. L'auteur conclut donc avec ceci: "La seule solution est d'inclure des choix dans le gameplay, laisser le joueur traiter ces choix comme une part de l'expérience et éventuellement les ignorer. Alors, ses décisions deviendront intuitives et reflèteront ses véritables désirs." Ainsi, hormis par l'analyse de données bas niveau, susciter le Flow via des mécaniques de jeu reviendrait à créer une multitude de systèmes, se répondant ou non entre eux, tout en testant de manière empirique sur un panel d'utilisateurs plus ou moins exhaustif la pertinence de ces systèmes. Une autre solution, cette fois post-développement, consisterait à interpréter de manière statistique certaines données utilisateur par le développeur, afin de les synthétiser en statistiques pouvant être utilisées pour réajuster certaines mécaniques, ou bien être prises en compte pour le prochain titre produit. Même si ce procédé se base sur des données plus concrètes et ancrées dans la réalité, il s'ancre toujours dans une démarche plus ou moins subjective, et ne peut donc suffire à elle seule afin d'agencer des mécaniques permettant d'atteindre un état de flow de la manière la plus optimale possible chez la/les différentes cible(s) de joueur(s) et joueuse(s).

Cependant, si la conception des mécaniques et les choix de game design constituent une part capitale dans l'apparition du flow, la direction sonore n'est pas en reste: "Le son étant un matériel très facilement analysé par le cerveau, il véhicule rapidement une notion de qualité et de confort d'écoute. Il constitue donc le premier sens vecteur de flow au sein d'un jeu vidéo: il englobe le joueur dans un cocon, un cadre rassurant, qui lui permet de lâcher prise.", nous dit Masami Komuro. Le son serait alors indissociable de l'oubli du Soi (l'une des composantes majeures de l'état de Flow), car il donnerait l'opportunité au joueur de s'évader de sa conscience, tout en gardant le contrôle physique de sa personne. "On pourrait également établir un parallèle entre la composition fréquentielle d'un son et sa faculté à plonger le joueur dans ce dit cocon, suivant le niveau de confort et de qualité perçu par le joueur vis-à-vis de ce son." continue-t-il. Ainsi, le son permettrait, entre autres, de plonger le joueur dans un état propice à l'apparition du flow. Toutefois, il faut alors distinguer l'appréciation même d'un son (si il est agréable ou désagréable) de l'état de confort dans lequel il plonge le sujet: un jeu comme Doom Eternal comportant des sonorités industrielles agressives, pourra parfois proposer un panel d'éléments sonores très désagréables à l'audition du joueur, tout en le plaçant dans un état de confort propice à l'apparition du Flow de par ses qualités (textures riches, travaillées et équilibrées malgré le type de sonorités proposées). Le spectre fréquentiel (et donc, le niveau d'agréabilité) d'un signal sonore serait donc plus corrélé à un but scénaristique qu'à une aide ou un obstacle vis-à-vis de l'état de Flow.

On peut alors supposer que le son, de par sa nature physique et beaucoup plus directe en terme de transmission vers le cerveau humain, serait un moyen plus efficace pour susciter l'état de flow dans l'expérience du joueur, comparé à l'analyse de données récupérées par des phases de test ou par les actions d'un ou plusieurs joueur(s) en temps réel. Toutefois, une bande son ou un système interactif prévu dans une direction sonore ne peut s'appuyer que sur "ce qui est perçu par le développeur comme efficace dans un certain contexte de jeu et/ou pour une cible type de joueur": la question de l'utilisation du son dans le jeu vidéo comme moteur de flow, en particulier dû à son caractère artistique, reste donc un élément subjectif. Cependant, au même titre qu'au fil des expériences de jeu vidéo de ces 40 dernières années, certaines composantes liées à l'appréciation d'un jeu (dues à certaines mécaniques ou phénomènes cognitifs) ont pu être décodées et analysées, la même idée s'applique à la direction sonore d'un jeu vidéo. Nous essaieront donc, au cours de ce mémoire, de comprendre, lister et établir des éléments de réponse qui permettraient de nous guider dans la réalisation d'une direction sonore, afin de pouvoir plus efficacement susciter, et surtout maintenir l'état de flow chez le joueur.

Mais tout d'abord, il convient d'établir une différenciation entre le concept "d'immersion" et de "flow", souvent incompris et regroupés à tort, jusque dans le milieu professionnel.

2. Distinction entre "Immersion" et "Flow"

Si nous avons pu, au cours de notre première partie, expliquer avec plus ou moins de détail l'état de flow ainsi que son contexte d'apparition, celui d'immersion est autrement plus complexe, car dépendant de plusieurs phénomènes et facteurs variables selon les profils. La définition la plus communément admise (notamment dans le dictionnaire Français actuel "Le Robert") consisterait en: "l'action d'immerger, de plonger (dans un liquide ou un milieu)". De là, on peut évidemment considérer l'espace virtuel de jeu comme un milieu, dans lequel le développeur immerge le joueur par l'utilisation de différents procédés mécaniques, narratifs, haptiques, visuels et sonores. L'on peut alors avancer une définition de l'immersion propre au jeu vidéo, qui serait la suivante: "L'immersion dans le jeu vidéo est un processus qui permet, via la somme de différents procédés mécaniques, narratifs, et sensoriels, de plonger mentalement et physiquement le joueur dans un espace numérique, créant alors une illusion d'évasion de l'espace réel." Ainsi, dès lors que cette illusion serait brisée, le joueur serait amené à sortir de l'expérience. Or, une nouvelle fois, chaque

profil de joueur est singulièrement différent. Dans ce cas, un élément de gameplay (comme une mécanique de saut plus ou moins ample), un procédé narratif utilisé (comme retournement dans l'histoire), ou même l'utilisation de certains feedbacks sensoriels (comme les retour haptiques d'une manette) serait à même de briser l'intérêt qu'un joueur porte à une expérience. Si certains procédés d'immersion basiques, comme les retours haptiques ou le son, peuvent parfois s'ajuster via divers paramètres dans l'interface de jeu (permettant alors de régler l'intensité des vibrations, d'ajuster le mixage sonore du jeu, d'adapter le son au système d'écoute utilisé, etc.), d'autres sont autrement plus complexes à modifier selon les particularités d'un joueur, et parfois figées lors d'une expérience effectuée sur un produit fini. Via toutes ces problématiques liées à la création et à la rétention de l'état d'immersion chez le joueur, l'état de flow peut alors être utilisé comme outil par les développeurs afin de guider l'expérience, et retirer les frictions présentes sur le plan physique, intellectuel et/ou émotionnel.

Pour rendre ce concept plus clair, utilisons un exemple: un joueur entre dans une salle vide, et la porte d'entrée se referme derrière lui. Un message apparaît alors sur l'écran: "Pour sortir, vous devez résoudre une énigme, et resterez prisonnier de cette salle jusqu'à la résolution de l'énigme en question". Cette énigme consiste en un rubik's cube de 3 carrés par 3 carrés sur chaque face. Là où cette énigme plus ou moins complexe pourrait frustrer, voir susciter un désir d'abandon de l'expérience chez le joueur en fonction de ses capacités, le développeur pourrait choisir de diffuser des feedbacks sonores apaisants, comme des ambiances douces d'écoulement d'eau ou de bruits d'espace naturels, afin d'apaiser le joueur et faciliter sa concentration, ainsi qu'accroître sa capacité de résilience par rapport au challenge imposé. Également, à chaque manipulations du rubik's cube, des feedbacks sonores satisfaisants et impactants, ainsi qu'un motif musical entraînant apparaissant à la complétion de chaque faces de l'objet, permettraient de stimuler le joueur en amenant chez lui des pics de dopamines. De tels procédés, simples en apparence, permettraient ainsi de garder le joueur dans l'expérience, pouvant, alors, aller jusqu'à inscrire la répétitivité des manipulations effectuées comme seul plaisir de jeu, ce dernier devenant alors décorrélé de la réussite ou l'échec de l'énigme.

Via l'exemple utilisé, et les explications fournies dans notre première partie, on peut alors synthétiser une définition du flow comme suit: "Un état mental caractérisé par l'investissement psychique total d'une personne effectuant une action, survenant lorsque le sujet se retrouve entre ennui et anxiété. Maintenu par divers stimuli, cet état permet de maintenir l'engagement du sujet dans l'action en train d'être effectuée."

Ainsi, d'après une synthèse des dires de Christophe Héral lors de notre interview: "Si l'objectif d'une expérience est un phare, et que l'environnement à traverser pour l'atteindre est un océan, l'état de flow serait alors le gouvernail qui permettrait de nous y conduire, tout en nous aidant à naviguer entre les remous susceptibles de

faire chavirer l'embarcation. Or, pour maintenir le cap, les feedbacks sonores servent de balises, permettant de nous orienter afin d'arriver à bon port, et ceci sans nous décourager." Donc, si l'immersion et le flow sont deux états différents, le flow permettrait de maintenir l'immersion. Mais alors, la rupture de l'état de flow serait susceptible de sortir le joueur de l'état d'immersion, et vice versa: ces deux états seraient donc interdépendants, soit les deux faces d'une même pièce sans lesquelles le plaisir et l'intérêt d'un joueur vis-à-vis d'une expérience ne pourrait exister.

3. Le Son, pilier du flow

3-A - Les problématiques techniques et humaines liées à la réalisation de la conception sonore d'un jeu vidéo

L'on peut simplifier la définition de "game design" comme suit: "La définition d'un ensemble de règles et de parti pris artistiques afin de concevoir un jeu". Or, une multitude de jeux vidéo ont, pour des raisons techniques, de budgets et d'ambition, traités le son comme un élément à part, voir secondaire, que ce soit dans sa fonction ou les ressources y étant allouées. Sans rentrer dans une dénonciation de ces titres, bien trop nombreux pour en faire le compte, la raison fut d'abord technique puis marketing.

En effet, jusqu'aux années 90, la mémoire contenue dans les cartouches de jeu ne permettait pas de contenir ni restituer une qualité ou une quantité de matériaux sonores conséquente. Pour pallier cela, les sons étaient synthétisés, et les éléments musicaux étaient pauvres en nombre de voix de polyphonie ainsi qu'en termes de timbres, en plus de s'avérer parfois très répétitifs. De 1990 à l'apparition du CD, certains titres (comme les jeux d'aventures de Lucas Art, notamment Monkey Island) commencent à explorer sérieusement la notion d'interactivité sonore. En effet, la technologie derrière les puces sonores ayant évolué, le nombre de voix pouvant être jouées en même temps, ainsi que la puissance de calcul des machines, permettait d'allouer des systèmes plus robustes et complexes à la direction sonore. Également, les instruments étant contrôlés en MIDI, leur lignes mélodiques pouvait être contrôlées et remodelées en temps réel, en fonction des besoins: par exemple, dans Monkey Island, les programmeurs audio avait créé un système permettant de réajuster le midi des instruments joués de sortes à ce que les transitions musicales

d'un tableau à un autre s'effectuent de la manière la plus fluide possible. A cette époque, le son commençait à prendre une ampleur et une importance réelle. Mais, dès la sortie de la première Playstation, un changement drastique dans la conception des bandes son de jeu vidéo s'opéra: ces dernières étaient maintenant constituées de véritables fichiers audio lus en temps réel, et non plus des données MIDI interprétées par une puce sonore. Or, le coût d'une approche interactive de l'audio sous forme de samples s'avérait nettement supérieur (en termes de calculs et de mémoire nécessaire), vis-à-vis d'une approche au format de synthèse contrôlée en MIDI, et les consoles faisaient alors le choix de se concentrer sur le rendu graphique des jeux, afin de proposer le prétendu "photoréalisme" avancé dans les campagnes marketings d'époque. Également, la manipulation de fichiers audio en temps réel s'avérait alors bien moins flexible. Cela conduisit l'interactivité audio dans le jeu vidéo à prendre une place sur le siège passager durant plus d'une décennie, jusqu'à l'apparition et l'adoption de solutions middlewares telles que Wwise.

A compter de la date d'écriture de ce mémoire, les consoles de dernière génération ainsi que la puissance des composants PC permettent aux créateurs de jeux d'allouer bien plus de puissance au profit de la direction sonore. Néanmoins, les ressources matérielles et humaines allouées à la partie sonore des jeux reste très minime. Voici quelques statistiques sur des jeux AAA en monde ouvert sortis au cours de ces dix dernières années:

- Pour "Ghost of Tsushima", l'équipe audio représentait seulement 4% des effectifs du studio Sucker Punch au maximum (selon Rev. Dr. Bradley Meyer via son interview dans The Sound Architect Podcast, directeur audio chez Sucker Punch).
- Pour "Cyberpunk 2077", l'équipe audio représentait 3,5% des effectifs de CD Projekt Red (16 personnes sur environ 500 employés).
- Pour la trilogie Hitman (2016, 2018, 2021), l'équipe audio représentait 2,5% des effectifs de IO Interactive (sur un total d'environ 200 employés).
- Lors du développement de "The Elder Scrolls V: Skyrim", l'équipe audio était composée d'un seul membre, Mark Lempert (plus tard rejoint par Dave Shreiber), tandis que le total des effectifs travaillant sur le jeu s'élevait à environ 100 employés (soit environ 1 à 1,5% des effectifs de l'équipe travaillant sur la partie sonore).

On peut donc facilement mettre en corrélation ces statistiques avec un manque d'intérêt, de compréhension et d'ambition venant des chefs(fes) de studio et chefs(fes) de production vis-à-vis de l'importance de la direction sonore dans un jeu. Également, les jeux en monde ouvert étant ceux nécessitant le plus de systèmes et de ressources (au vu de leur envergure et complexité), les statistiques fournies plus haut n'en sont que plus révélatrices. Toutefois, selon Bjorn Jacobsen: "L'audio ne serait pas nécessairement mis au second plan de par le manque d'effectifs, mais plus par l'attitude générale du processus de développement envers la dimension sonore d'un

jeu, ce qui peut alors provoquer l'ignorance de l'équipe audio dans le processus décisionnel, et ce, quelque soit sa taille. Cela est généralement dû à un manque de compréhension vis-à-vis du temps nécessaire à l'élaboration d'une conception sonore de qualité, le montant de code et le nombre d'outils nécessaires à cette tâche, ainsi que le nombre de programmeurs son requis pour réaliser ces derniers."

Maintenant que nous avons pu constater et comprendre l'origine de la mauvaise représentation du son dans l'industrie du jeu vidéo, il est temps d'expliquer pourquoi la qualité d'une direction sonore constitue un élément vital d'une expérience de jeu, et comment la considération de la partie sonore comme partie intégrante du processus de game design résulte en un ensemble cohérent et impactant, permettant alors de faire apparaître et maintenir l'état de flow chez le joueur.

3-B - La direction sonore conçue comme partie intégrante du game design

"Eyes lie. But the ears don't.", nous dit Seth Horowitz, neuroscientifique spécialisé dans le milieu auditif. Selon lui: "Si une information visuelle prend environ 0,25 secondes à arriver au cerveau, une information sonore met quant à elle 0.05 secondes à être perçue. Également, notre cerveau est si habile à distinguer les différences entre les sons, qu'il peut percevoir des changements sonores se déroulant en moins d'un millionième de seconde. Cette acuité auditive est due à la formation évolutionniste du cerveau humain, permettant jadis à nos ancêtres de réagir instantanément à la moindre menace via l'ouïe, et ceux sans qu'un quelconque signal visuel ne soit nécessaire." A travers ces informations, l'on se rend compte que les informations sonores constituent un canal sensoriel beaucoup plus rapide et instantané que celui de la vision. Si ces deux sens possèdent chacun leurs avantages et angles morts vis-à-vis de la retranscription de l'environnement qui nous entoure, l'immédiateté de l'audition, sa capacité à nous alerter vis-à-vis d'une menace ainsi que sa faculté à nous représenter et diriger dans un espace ne peut être remise en cause.

Or, si le jeu vidéo est un médium large, dont chaque expérience peut être abordée sous une multitude d'angles différents, la plupart des titres sortis à ce jour se rejoignent en un point: la notion de challenge, souvent en corrélation avec la notion de danger. Le jeu vidéo fait ainsi appel à notre cortex préfrontal et l'amygdale, à savoir les parties du cerveau agencant nos instincts les plus primaires. Si certains jeux arrivent à proposer une expérience où cette notion de danger est pratiquement voir totalement absente (par exemple, le jeu de puzzle The Witness, ou encore

Flower) et où le son intervient dans d'autres buts, le game design s'efforce donc de proposer un challenge intéressant dans un contexte précis, où le danger (l'échec, la mort du personnage) est souvent le principal moteur. De là, on peut en déduire que la manière la plus efficace de représenter ce danger dans une expérience serait d'exploiter les nombreuses possibilités de suggestion offertes par le canal auditif. Malheureusement, contrairement au cinéma où le son est utilisé de façon fine et réfléchie (véhiculant un panel d'émotions et de sensations complexes, des éléments scénaristiques, etc.), beaucoup de titres actuels se cantonnent à une utilisation basique du son, sans chercher à fournir autre chose qu'un support sensoriel à l'aspect visuel d'une expérience: le son devient alors un élément qui vient se greffer en second plan au reste du jeu, et dont les systèmes principaux ont donc été élaborés sans réelle considération de l'un de nos 5 sens les plus importants. Alors, les concepteurs sonores n'ont que peu de marge de manœuvre afin d'utiliser de façon pertinente le son dans une proposition vidéoludique. C'est ce manque de considération ou de compréhension qui, naturellement, viendra provoquer une rupture régulière de l'état de flow dans l'expérience.

Mais alors, à quoi ressemble une production qui, dès le départ, intègre la partie sonore au centre de la conception du game design? Prenons le cas du jeu "Hellblade: Senua's Sacrifice", sorti en 2019 et ayant remporté une nomination pour sa direction sonore aux BAFTA Awards. Dans cette expérience d'action narrative, le joueur suit le destin torturé de Senua, une jeune femme sujette à des psychoses et déviances psychiatriques, harcelée par de nombreuses voix et personnalités dans sa tête. Afin de permettre à un joueur totalement inconnu à ces troubles psychologiques de mieux comprendre le calvaire de Senua et ainsi s'identifier à elle, l'équipe audio a décidé d'utiliser des micros d'enregistrement binaural. L'enregistrement binaural est une méthode de captation sonore obtenue en positionnant deux microphones de manière à ce qu'une personne entende le résultat réaliste (comme l'on entendrait un son dans la réalité, non diffusé par des hauts parleurs). Ce procédé fut utilisé pour capturer les voix de différents acteurs et actrices se déplaçant autour des microphones, afin de donner à l'écoute l'impression que ces voix tournent autour de la tête du joueur, lui faisant alors ressentir la folie du personnage de façon tangible. Dans Hellblade, l'interaction avec ces hallucinations auditives est permanente, et au centre même du game design (notamment au travers des divers challenges proposés). Le joueur est alors investi de manière personnelle, et les épreuves endurées par Senua sont directement perçues et ressenties par celui ou celle prenant part à l'expérience.

En considérant le son au même titre que les autres canaux d'information sensoriels, le studio Ninja Theory a su tirer le meilleur parti de son sujet. Ainsi, là où chaque jeu est différent dans son concept et ses besoins, une réflexion judicieuse de la conception sonore dès l'élaboration du game design et des mécaniques d'un jeu permettent de drastiquement renforcer (voir amener) l'implication du joueur au sein d'une expérience, permettant alors de faire apparaître l'état de flow.

3-C - L'utilisation des caractéristiques sonore comme vecteur de flow

Comme nous avons pu le voir plus haut, les différents types de son, de manipulation sonore et de diffusion utilisés dans une expérience possèdent la faculté d'en modifier drastiquement la perception chez le sujet. Selon les besoins d'évocation (si le développeur souhaite effrayer, surprendre, galvaniser ou calmer), différentes caractéristiques sonores peuvent être présentées et/ou altérées afin d'amener tout un panel d'émotions chez le joueur.

3-C-Partie 1: Les variations d'amplitude sonore

Les variations d'amplitude sonore (volume) abruptes ou graduelles sont souvent utilisées afin de varier la tension chez le joueur à un moment T. L'idée de tension graduelle véhiculée par le son, déjà présente dans la création musicale depuis les débuts de son apparition, permet de rythmer une expérience, s'articulant autour de "points chauds" (pics de tensions) et de "points froids" (retombées, calme avant la tempête). Dans les jeux et films d'horreur, les variations de volume abruptes (ou "screamer") sont également beaucoup employées: le joueur marche le long d'un couloir, les bruits extérieurs deviennent de plus en plus silencieux, et d'un coup, un monstre surgit du décor, amenant avec lui toute une panoplie de sons et d'éléments musicaux qui surprennent le sujet. Ainsi, les variations importantes de l'amplitude d'un son sont naturellement perçus comme un danger par le cerveau humain: ces variations servent alors à instaurer un rythme dans l'expérience, rythme que les concepteurs sonores s'emploient à instaurer et briser selon les situations. Dans DARQ, jeu de puzzle horrifique sorti en 2020 sur consoles et PC, le sound designer en charge de la direction sonore, Bjorn Jacobsen, utilise fréquemment ces procédés de cassures de rythme propre à l'horreur. La grande gamme dynamique du titre est composée de sons très faibles récurrents (comme les bruits de pas et ceux de divers mécanismes), mis en opposition avec des sons très forts et ponctuels (afin de souligner certains éléments de gameplay ou scénaristiques). Ce choix de conception sonore permet alors de garder l'intérêt de l'auditeur vis-à-vis de l'action en jeu, tout en le plaçant dans un état d'alerte faible mais constant, sans fatiguer le joueur outre mesure.

3-C-Partie 2: La gestion de la gamme dynamique

La dynamique globale d'une conception sonore peut ainsi servir d'outil afin de véhiculer des informations au joueur, et le placer dans un sentiment d'apaisement ou de tension constante. Dans la plupart des jeux vidéo traditionnels, les phases d'affrontement/de tension sont toujours représentées par une dynamique très réduite, de par l'ensemble que constitue la musique et les nombreux feedbacks sonores, venant alors apporter un côté stressant, presque suffocant. Si ces phases de tensions deviennent trop récurrentes, et si leur plage dynamique est trop réduite, cela peut d'ailleurs conduire à une rupture du flow dans l'expérience, le manque de variations d'amplitude devenant alors très fatiguant pour le cerveau humain. Dans le cas contraire, un jeu possédant une plage dynamique très grande, avec très peu d'informations sonores allant au-dessus des -30 à -25 dbfs, peut finir par créer un désintérêt chez le joueur sur une période prolongée. Pour pallier ces disparités et crêtes importantes, les concepteurs sonores se réfèrent, en jeu vidéo, à une norme de niveau, consistant à maintenir l'ensemble des informations sonores autour de -23 LUFS. Or, cette normalisation peut conduire à représenter certains types de sons (comme des ambiances naturelles) plus fortes qu'elles ne le sont dans la réalité, pouvant alors créer une dissonance cognitive ou plus simplement une fatigue auditive perceptible sur une durée relative à chaque individu, et pouvant donc entraîner la disparition de l'état de flow. L'intégration sonore et l'utilisation d'un mixage dynamique permettent d'éviter ces soucis, sujet que nous explorerons plus loin. Par conséquent, l'utilisation de variations d'amplitude, ainsi que la gestion de la gamme dynamique au sein d'une conception sonore, s'avèrent être un travail de minutie pouvant faciliter et maintenir l'état de flow chez le joueur, comme pouvant avoir des conséquences désastreuses sur l'expérience en cas de mauvais dosage.

3-C-Partie 3: Le domaine fréquentiel

L'utilisation du domaine fréquentiel au sein de la création sonore dans un but utilitaire plutôt qu'esthétique est un sujet dense et complexe, que nous ne pourrions que survoler ici. Cela étant dit: la perception auditive de chaque individu est différente et propre à ce dernier, dû à la courbe de réponse unique de chaque paire d'oreilles humaines. Cependant, on peut constater que l'appréciation d'un type de son, ou du moins d'une plage fréquentielle particulière, est aussi culturelle: par exemple, la langue japonaise comprenant beaucoup de sons aigus, la culture japonaise se montre généralement plus friande de sons "brillants" comportant une certaine quantité de haute fréquence en comparaison à des cultures plus occidentales, car habituée à ce type de sons. On peut le remarquer en examinant le

timbre des instruments japonais traditionnels, voir récents (comme les pianos et les synthétiseurs), ainsi qu'en observant la courbe de réponse des outils audio produits sur le territoire japonais, ou encore par l'analyse de la répartition du spectre fréquentiel dans leurs contenus audio.

Malgré des tendances plus ou moins ancrées culturellement, on peut tout de même dégager certains faits universels: la courbe isosonique (soit une moyenne de la courbe d'audition d'un individu de 18 ans) indiquant un net pic de 8 à 8.50khz, on sait que cette plage fréquentielle constitue le point le plus sensible de l'audition humaine. Les instincts de réaction au danger se manifestent donc plus vivement autour de cette fréquence, et les cris de bébés humains se plaçant également dans cette plage là, on peut en déduire que cette caractéristique-ci a influencé l'évolution humaine (les humains s'étant adaptés au fil du temps à percevoir plus vivement cette plage fréquentielle afin de protéger leur progéniture). Donc, on peut utiliser cette information afin d'orienter la place du son dans le spectre fréquentiel des informations sonores d'alerte (ou servant à indiquer un danger), dans le but de fournir des feedbacks plus efficaces et cohérents par rapport aux intentions de game design. Par conséquent, cela permet au joueur d'avoir une meilleure lisibilité sur son environnement et l'action en cours, lui donnant la possibilité de réagir plus instinctivement aux potentielles menaces: cela contribue alors à le maintenir dans un état de flow.

3-C-Partie 4 - La Localisation

Une multitude d'expériences scientifiques ont prouvé que le cerveau se base sur les différences plus ou moins subtiles de temps, d'intensité et de timbre entre le canal auditif gauche et le canal auditif droit afin de se localiser dans l'espace, ainsi que pour localiser l'ensemble des sources sonores qui l'entourent. La localisation de l'avatar du joueur dans un espace, ainsi que l'identification de la position et de la distance d'une source relative à la position de cet avatar, constituent des éléments essentiels afin de fluidifier une expérience vidéoludique. Plus les informations de localisation des sources sont précises, plus la vitesse à laquelle les différents feedbacks et informations sonores sont assimilées est accrue. De ce fait, l'effort mental à fournir par le joueur afin de réinterpréter les signaux qui l'entourent est moindre, permettant à ce dernier de réagir de manière plus instinctive, l'ancrant donc plus profondément dans l'expérience; l'état de flow est alors maintenu plus facilement. Ainsi, l'utilisation et la précision de simulation des phénomènes acoustiques et physiologiques présents dans la vie réelle peuvent permettre de faciliter la compréhension et l'immersion du joueur dans une scène. Sans rentrer dans une liste exhaustive, voici quelques propriétés, phénomènes acoustiques et

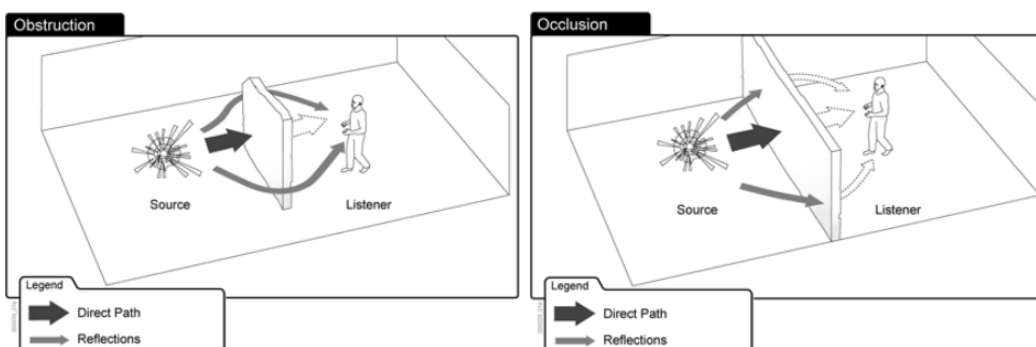
physiologiques dont la compréhension et la simulation peuvent simplifier les processus de localisation au sein d'une expérience vidéoludique.

3-C-Partie 4.1 - La convolution

La convolution est une technique qui consiste à reproduire algorithmiquement, à partir d'une réponse impulsionnelle (comme un tir de pistolet), les propriétés acoustiques d'un espace. Pendant longtemps, cette technologie fut trop coûteuse en puissance de calcul pour être utilisée en temps réel. Si l'utilisation de réverbérations à convolution, en particulier dans les jeux de tir à la première ou troisième personne, s'est démocratisée au fil du temps, on assiste depuis quelques années au développement croissant d'une forme de convolution plus dynamique dans le jeu vidéo, qui serait basée non plus uniquement sur la transition d'une réponse impulsionnelle à une autre, mais bien sur une réflexion des signaux sonores basée sur les changements en temps réel de la géométrie d'un espace en 3D. A l'heure d'écriture, le concept de ray-tracing appliqué à l'audio, soit la réflexion de signaux sonores basée sur le taux de réflexion et le niveau d'absorption des surfaces entourant la source, continue d'être explorée (notamment via la technologie VRWorks de NVidia). Un tel accroissement dans la précision de la représentation de l'espace sonore permet évidemment d'accentuer le rapport tangible que le joueur tisse entre l'expérience virtuelle et lui-même, contribuant à maintenir son état de flow.

3-C-Partie 4.2 - La simulation des effets naturels d'obstruction et d'occlusion

Essentiellement, l'effet d'obstruction consiste à altérer la composition fréquentielle et d'amplitude d'un signal, suivant les objets ou les surfaces qui viennent se mettre en travers d'une source sonore et de son auditeur. Cependant, cet effet ne prend pas en compte les réflexions de la source sonore, à l'inverse de l'effet d'occlusion.



La simulation de ces effets permet d'accroître le réalisme d'une scène, tout en donnant au joueur la possibilité de mieux se représenter l'espace, ainsi que les différents éléments qui le composent.

3-C-Partie 4.3 - L'utilisation du binaural et des fonctions de transfert

Comme nous l'avons vu plus haut, l'on peut définir le binaural de la manière suivante: "Tout système de prise de son capable de restituer un rendu spatial au casque, avec la sensation que les sources proviennent de l'extérieur de la tête" (comme pour la perception naturelle ou lorsque l'auditeur est placé au centre d'un système de diffusion multicanal). Cette forme de spatialisation se distingue de la stéréophonie au casque traditionnelle (laquelle donne une perception intracrânienne) en reproduisant les fonctions de transfert HRTF de notre tête. L'HRTF (Head Related Transfer Function, ou "fonction de transfert relative à la tête") caractérise, par une fonction de transfert mathématique, les déformations (diffractions, réflexions et atténuations) apportées aux ondes sonores par les épaules, la tête et les pavillons de chaque oreille de l'auditeur, lesquelles nous permettent de repérer l'origine d'un son, tant horizontalement que verticalement. Par conséquent, ces fonctions dépendent de la physionomie de chacun, et varient donc largement d'un individu à un autre. Si certaines entreprises du milieu du jeu vidéo (notamment Sony) réfléchissent à un moyen permettant à l'utilisateur de créer facilement son propre profil HRTF, afin d'obtenir une localisation plus réaliste des informations sonores dans l'espace de jeu, certains middlewares audio, comme Wwise, permettent d'encoder un signal en binaural. Il existe également certaines solutions logicielles permettant de synthétiser un rendu binaural à partir d'un fichier audio traditionnel, via diverses techniques de filtrage propres à chaque développeur.

3-C-Partie 4.4 - La simulation de l'effet Doppler

Ce phénomène se manifeste dans la vie réelle par une accentuation des hautes ou basses fréquences provenant d'une source en mouvement, en fonction de la proximité entre cette source et l'auditeur. Plus la source en mouvement est proche et plus les ondes perçues sont rapprochées (plus aiguës), plus la source en mouvement est lointaine et plus les ondes seront éloignées (plus graves). La simulation de l'effet Doppler permet de localiser et de se localiser plus facilement vis-à-vis des sources sonores qui composent une scène : dans des jeux de tir, de courses ou tout autre expérience nécessitant la réactivité totale et instantanée du joueur à des menaces pouvant provenir de toutes les directions, la reproduction convaincante de ce phénomène acoustique permet au joueur de se représenter plus facilement l'action, et par conséquent d'augmenter le contrôle sur ses mouvements et décisions, ce qui contribue à créer ou maintenir l'état de flow.

3-C-Partie 4.5: Les phénomènes d'hyper localisation, d'ubiquité et de délocalisation

Les phénomènes d'hyper-localisation, de délocalisation et d'ubiquité, relatifs au domaine de la psycho-acoustique, peuvent parfois être ignorés lors de la conception sonore d'un jeu. Cependant, leur compréhension permet d'accroître drastiquement l'habileté du joueur à repérer les sons qui l'entourent dans l'espace.

L'hyper localisation est un effet liée au caractère ponctuel d'une source sonore, focalisant irrésistiblement l'attention de l'auditeur sur le point d'émission. Lorsque la source se déplace, le son continue d'être suivi à la trace. Il est souvent consécutif à des transmissions solidiennes, comme par exemple une bille roulant à l'étage au-dessus. Si un objet en mouvement ou un son indésirable retient l'attention de l'auditeur, sa vision, mais surtout son audition, peuvent se désintéresser de l'action ou de la scène en cours, créant alors une nuisance pouvant rompre l'état de flow et d'immersion dans lequel le joueur était plongé jusqu'à lors. Le concepteur de jeu doit donc rester vigilant vis-à-vis de la représentation sonore des sources en mouvement, pouvant faciliter le processus de localisation comme déconstruire l'état de flow chez le joueur.

L'ubiquité est un phénomène qui se produit lorsque la source sonore ne peut être localisée, ou que le son semble provenir de toutes parts sans que l'auditeur puisse en détecter la provenance. Si cet événement sonore peut être utilisé comme moyen pour construire un sentiment de tension avant la révélation de sa localisation, il est aussi susceptible de perturber l'auditeur, et briser l'état d'immersion et de flow dans lequel il était plongé. Également, l'effet d'ubiquité peut être perçu comme une nuisance, venant déconcentrer le joueur de l'action en train d'être effectuée, ou l'éloigner de l'intention émotionnelle censée être véhiculée au joueur à un moment T. Si certains sons peuvent être considérés comme des couches sonores "globales", tel que des sons d'ambiance naturels ou urbains, leur source doit être révélée au bout d'un certain temps, afin de ne pas créer une dissonance image/son qui viendrait nuire à l'expérience. Par exemple, si le joueur se trouve dans une ville et entend des bruits de pots d'échappements suffisamment distincts, l'origine de cette source doit pouvoir se matérialiser visuellement devant lui, passé une certaine durée. Le jeu de puzzle à l'architecture escherienne "Manifold Garden" utilise notamment ce phénomène d'ubiquité dans sa conception sonore: en jeu, des ambiances naturelles d'arbres agités par le vent, ou de vie sauvage naturelle, sont régulièrement diffusées dans les oreilles du joueur. Cependant, le titre plongeant continuellement le joueur au cœur de thématiques telles que la multi-dimensionnalité et un renversement des lois de la physique humaine, l'utilisation de l'ubiquité vient ici renforcer le propos et l'atmosphère du jeu, plutôt que créer une nuisance sonore pour le joueur.

La délocalisation est une forme minorée, ainsi qu'une composante de l'ubiquité. S'il peut y avoir une délocalisation sans ubiquité, il ne peut y avoir d'ubiquité sans délocalisation. Cet effet implique la reconnaissance d'une erreur dans la localisation de la source sonore: on ne sait pas d'où provient le son, mais à la différence de l'ubiquité, on sait précisément d'où il paraît venir, tout en ayant conscience que cette localisation est factice et illusoire. Pour plus de contexte, dans le jeu de tir "Bioshock: Infinite", il est fréquent que lors d'une cinématique où le personnage principal ouvre lentement une porte, tandis qu'une conversation derrière cette porte peut être perçue, aucun effet de filtrage ne viennent nous indiquer que l'on perçoit une source à travers une surface (dans ce cas ci, une porte). Le cerveau comprend alors que le son devrait posséder certaines caractéristiques, ici absentes, afin de sonner de manière plausible. Cela vient donc augmenter le "score de nuisance sonore" ou "NUFS", concept théorisé par Bjorn Jacobsen et sur lequel nous reviendrons plus en détail par la suite.

3-C-Partie 5 - Le rythme de la conception sonore

L'on peut définir le "rythme" au sens large comme suit: "La caractéristique d'un phénomène périodique, induite par la perception d'une structure dans sa répétition." La question du rythme au sein d'une expérience vidéoludique est cruciale. Si une partie de ce rythme est conditionnée, voir définie par le game design, il est également agencé, maintenu ou brisé par la conception sonore, que ce soit au niveau de la partie musicale, ou de la représentation sonore des phénomènes apparaissant en jeu et des actions effectuées par le joueur.

Une constituante importante du rythme de la conception sonore est la répétition de sons. "Lorsqu'un schéma récurrent se dessine, soit dans un son qui boucle ou dans une multitude de sons joués à nouveau les uns après les autres, il peut y avoir une ligne très fine entre, d'une part, l'acceptation de ce schéma comme quelque chose que l'on reconnaît et qui nous fait nous sentir en sécurité, et d'autre part, quelque chose qui devient ennuyeux de par sa nature monotone et répétitive.", nous dit Bjorn Jacobsen dans sa thèse "The Tuning of the Game". Cela permet d'identifier la répétitivité sous l'angle de deux facettes distinctes: tantôt celle d'un outil, tantôt celle d'un obstacle vis-à-vis de l'apparition et du maintien du flow. On peut alors distinguer plusieurs composantes primaires du rythme sonore: la répétition d'un même type de son, et le rythme musical ou taux de déclenchement des sons.

3-C-Partie 5.1: La répétition d'un même type de son

Pour illustrer cet aspect, prenons l'exemple des bruits de pas. Le cycle de répétition des variantes de bruits de pas (induits par les déplacements du joueur), s'il peut être accepté comme un schéma rassurant et confortable, peut tout autant venir fatiguer et dégoûter ce dernier de l'expérience auditive, soit par les composantes fréquentielles de ces sons, soit par leur constance de volume. Alors, ces bruits de pas finissent par être mis de côté par le cerveau, grâce au célèbre effet "cocktail party" (permettant au cerveau d'occulter les informations provenant de bruits plus ou moins constants afin de se concentrer sur un ou plusieurs éléments précis). Or, cet effort mental demandé par l'effet cocktail party vient fatiguer le joueur, et accélérer, voire provoquer un sentiment de lassitude du joueur vis-à-vis de l'expérience.

Certains jeux développent des stratégies permettant d'éviter ce rejet de l'aspect répétitif des bruits de pas: "The Legend of Zelda: Breath of the Wild" possède un système de mixage de sons dynamiques très subtil, qui consiste notamment à réduire progressivement le volume des bruits de pas du joueur à mesure que celui se déplace de plus en plus longtemps sur la même surface. Cependant, lorsque le joueur vient à passer d'une surface à une autre, le volume se réhausse, permettant de fournir les informations sonores nécessaires au joueur afin de comprendre et de s'immerger dans le monde qui l'entoure. Ainsi, plus le cerveau s'habitue au déclenchement de ces sons et au cycle sonore qu'ils constituent, plus le mixage anticipe cette accoutumance, empêchant alors les sons de pas de se changer en bruit constant, ou nuisance sonore.

Un autre exemple de solution apportée serait celle adoptée dans "Grand Theft Auto 5", ainsi que le jeu atmosphérique "Inside". Ces deux jeux sont singulièrement différents en termes d'approche et de structure, l'un étant un jeu d'action violent en monde ouvert possédant une multitude de gameplays, l'autre étant un jeu de plateforme à la mise en scène figée et très cinématique avec un gameplay relativement statique. Pourtant, ils partagent un point commun: la décomposition des sons de bruits de pas du personnage incarné en une multitude de composantes, répondants à diverses variables en jeu, telles que le type de vêtement ou tissu porté, la surface rencontrée par les pieds, le niveau d'eau présent sur le sol, la démarche du personnage, la respiration de ce dernier, etc. Naturellement, chacune de ces différentes composantes possèdent elles même une multitude de variations, rendant alors impossible (ou beaucoup plus rare) l'apparition d'un sentiment négatif vis-à-vis d'une certaine répétitivité sonore.

3-C-Partie 5.2: Le rythme musical et le taux de déclenchement des sons

Une autre constituante majeure du rythme sonore serait le tempo, soit, d'une part, la période temporelle séparant deux occurrences d'un son, et d'autre part, le "rythme musical". Une étude scientifique réalisée par François Haas au Centre Médical Langone de New York, établit une corrélation entre le rythme musical perçu et le

comportement (rythme) de la respiration humaine. L'hypothèse résultante des tests effectués stipule que l'on peut provoquer une accélération ou une décélération du rythme respiratoire en fonction des variations de tempo présentées à un auditeur. Toutefois, cette idée est loin d'être nouvelle: en musique, puis au cinéma, un tempo plus ou moins élevé a toujours été utilisé afin d'accroître la tension d'un moment T. La gestion du tempo, notamment au sein de l'expérience vidéoludique, est donc cruciale; elle peut élever comme interrompre l'état de flow.

- Un tempo trop soutenu lors d'une période prolongée fatigue plus ou moins rapidement le joueur, comme cela peut être le cas avec des jeux comme "Thumper" (jeu musical à la musique industrielle très agressive et au rythme intense) ou encore avec les jeux de type "Shoot Them' Up" (comme "Ikaruga") où les nombreux sons de tirs et d'explosions finissent par former un bruit constant au point de nuire à l'expérience. Cela vient alors à l'encontre de la réception de feedbacks sonores importants, ces derniers finissant par être noyés dans une masse de sons.
- Un tempo trop lent, ou une absence de tempo est susceptible de venir lasser le sujet, ne procurant alors plus assez de stimuli sonores afin de diversifier l'expérience auditive et maintenir l'intérêt du joueur.

Si ces questions de rythme musical et de taux de déclenchement des sons sont souvent induites par le game design en lui-même, certains jeux mettent en place des procédés via leur direction sonore ou leur implémentation afin de contourner ces problèmes. "Doom Eternal", dernier jeu de tir en date de la célèbre licence d'ID Software, possède un gameplay effréné, extrêmement exigeant, brutal et dense. La conception sonore du titre reflète bien cela, que ce soit par les riffs de métal industriel et les nappes sonores électroniques saturées de Mick Gordon, ou encore par les sons de tir, d'explosions et de cris de monstres provenant de toute part. Pourtant, l'équipe audio d'ID Software, via l'implémentation d'un système de sidechain (pratique audionumérique qui consiste à réduire l'amplitude d'un signal sonore par le déclenchement d'un autre signal ponctuel, et donc la réduction est définie par l'amplitude du second signal), parvient à faire apparaître les feedbacks sonores cruciaux du jeu - lorsque le joueur tire, lorsqu'un ennemi dangereux meurt, etc. - au delà de la "masse sonore" présente en jeu. Également, le jeu d'exploration spatiale No Man's Sky parvient, malgré sa dynamique élevée, à maintenir une présence sonore plus ou moins constante en déclenchant de façon procédurale des couches musicales et d'ambiances, selon les actions effectuées par le joueur ainsi que la position de sa caméra. De cette façon, l'espace sonore est régulièrement occupé, sans être surchargé ou laisser de vides venant mettre fin à un sentiment de continuité, et donc interrompre l'état de flow du joueur.

3-C-Partie 6 - La gestion du rapport harmonique des éléments sonores

Le mouvement harmonique, à savoir une alternance de moments de dissonance (tension) et de moments de consonance (détente) est un procédé utilisé depuis les débuts de l'histoire de la musique, constituant le plus souvent le squelette d'une œuvre musicale. A l'intérieur de ce mouvement, on retrouve la "préparation", soit la façon d'amener une dissonance et d'introduire la tension, et la "résolution", soit la transformation de cette dissonance en une nouvelle consonance. Ces concepts sont aussi employés depuis les débuts du cinéma et communément adoptés dans le média du jeu vidéo, permettant de construire une anticipation chez le spectateur/joueur pour la relâcher ensuite, formant un cycle plus ou moins prédictible et créant de l'intérêt chez le sujet.

Le problème, comme indiqué dans la thèse de Bjorn Jacobsen citée plus haut, survient lorsque les éléments sonores d'une expérience vidéoludique entrent en dissonance de manière incontrôlée, soit les uns par rapport aux autres, ou par rapport aux éléments musicaux qui y sont superposés. Cela peut entraîner, dans bien des cas, une fatigue auditive et mentale, suivi par la rupture de l'état de flow. Pour faire entrer en consonance ou en dissonance les effets sonores avec la musique de façon contrôlée, le concepteur sonore a alors trois options:

- Soit ajuster la tonalité des effets sonores et des éléments musicaux de manière à les faire entrer en consonance entre eux, ce qui demande de prévoir toutes les éventualité de combinaisons de sons possible, rendant le processus fastidieux, voire impossible selon l'échelle du projet.
- Soit mettre au point un ou plusieurs systèmes lors de l'implémentation sonore, afin d'arranger automatiquement la composition harmonique des effets sonores et musicaux entre eux, pour éviter toute dissonance involontaire.
- Soit briser la ligne séparant les effets sonores des éléments musicaux, afin de les inscrire dans un ensemble cohérent et doublement immersif. Ce procédé, notamment utilisé au cinéma avec "Blade Runner", est rendu autrement plus compliqué dans un médium interactif comme celui du jeu vidéo. Cependant, le compositeur Olivier Derivière réussit à inscrire ce concept dans la conception sonore du jeu "Get Even", notamment via une bande son générée en temps réelle selon les actions du joueur: en effet, certains des effets sonores (sons de lampes vacillantes, frappements de porte) ainsi que certains instruments musicaux (piano, violons) et percussifs sont joués via un fichier midi, afin d'être synchronisés en temps réel aux battements de coeur du personnage principal. Aussi, certains effets sonores (comme celui des sirènes de police) sont d'abord diffusés et perçus de manière classique et diégétique,

puis, à mesure que le joueur s'éloigne de la source de ces sons, effectuent une transition vers des drones de musique orchestrale étant inscrits dans la même tonalité. A nouveau, en inscrivant dès le départ la conception sonore dans le processus de game design, Derivière et le studio "The Farm 51" parviennent à un résultat homogène, cohérent et impactant, ce qui permet de faire apparaître et maintenir l'état de flow dans l'expérience proposée.

Toutefois, comme l'indique Joonas Turner, la dissonance peut également être utilisée comme un outil, afin d'amplifier un son d'alerte ou un sentiment de danger, ce qui permet de faire plus facilement ressortir un feedback de l'ensemble sonore, et donc mieux véhiculer l'information. Il convient alors au concepteur de trouver un juste milieu, et d'utiliser ce genre de procédés à bon escient, sans mettre en danger l'état de flow.

Au cours de cette partie sur la relation entre le flow et le son en lui même, nous avons pu voir que l'approche de l'état de flow par la conception sonore variait selon chaque projet, que ce soit dans la (ou les) méthode(s) utilisées afin de l'amener et le maintenir, ou encore selon l'échelle de l'expérience, ainsi que des différentes situations qui la compose. Le concepteur sonore possède de nombreux moyens d'agir sur le son, incluant une infinité de solutions, afin de pouvoir amener l'état de flow et supprimer la plupart des éléments risquant d'entraver son maintien lors d'une partie. Nous allons maintenant voir comment la compréhension du concept de "game feel", permet de lier encore plus le joueur à l'expérience au travers de l'état de flow.

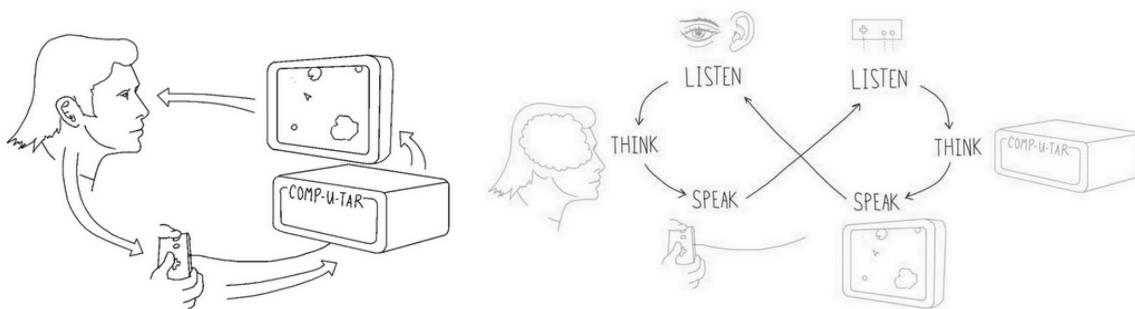
4. Game Feel, Flow et Son: une relation intrinsèque

4-A - Définition du “Game Feel”

L'on pourrait définir le game feel comme “la sensation intangible, tactile ressentie lors de l'interaction avec l'expérience vidéoludique”. Résultante d'un ensemble de mécaniques et systèmes, l'on peut séparer ce phénomène en 3 composantes: celle du contrôle en temps réel, de la simulation de l'espace de jeu, et enfin du “polish” (soit le niveau de finition de l'expérience).

4-A - Partie 1 - Le contrôle en temps réel

Le contrôle en temps réel est une forme d'interaction spécifique, incluant deux participants - ici, l'ordinateur et l'utilisateur - , formant une boucle fermée comme illustrée ci-dessous. Chris Crawford, dans son livre “Chris Crawford on Game Design”, décrit cette boucle fermée comme une conversation, soit un processus cyclique dans lequel deux agents actifs écoutent, pensent et parlent à tour de rôle.



Cependant, Steve Swink (auteur du livre “Game Feel”) rétorque que cette métaphore ne peut s'appliquer à toutes les situations, et que le game feel se rapproche plus de la conduite d'une voiture que d'une réelle conversation: “Si le conducteur veut tourner à gauche, il agit plus qu'il ne pense. Il tourne le volant correspondant à la direction qu'il veut prendre, au moyen de ce qu'il voit, entend et ressent jusqu'à ce que le tour soit complet. Ce processus est presque instantané. Or, une conversation prend place sur une temporalité beaucoup plus longue, en dessous du niveau de

conscience, dans un flot ininterrompu d'instructions, tandis que le résultat d'un "input" (contrôle) est perçu au même moment qu'il est exprimé. Cela constitue alors la base du game feel: un contrôle précis et continu exercé par le joueur sur un avatar ou un objet en mouvement.

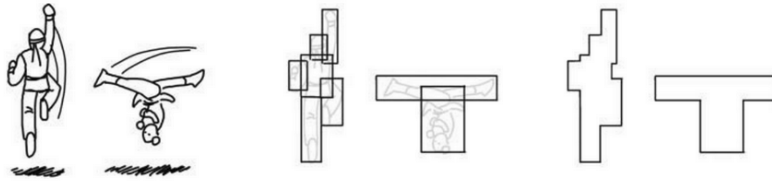
4-A - Partie 1 - La simulation de l'espace de jeu

La simulation de l'espace de jeu fait référence aux interactions physiques effectuées dans un espace virtuel, perçu activement par le joueur. Cela implique donc la détection de collisions et la réponse entre l'avatar et les objets qui composent le monde du jeu. Ces interactions donnent un sens au déplacement de l'avatar, en lui fournissant la possibilité de bouger autour, entre, contre les objets du jeu, tout en lui servant de référentiel vis-à-vis d'une impression de vitesse de déplacement. Cela permet de simuler la sensation tactile, physique d'interactivité que l'on expérimente au travers des espaces physiques qui constituent notre réalité de tous les jours. En utilisant l'avatar du joueur comme un canal de l'expression et de la perception, le joueur expérimente l'espace virtuel sur un plan physique et tactile, lui permettant de faire corps avec l'action et les éléments qui l'entourent en jeu. Une autre composante nécessaire à la simulation de l'espace de jeu, est que ce dernier doit être perçu de façon active, au contraire du format télévisuel qui est perçu de façon passive. Le game feel s'articule donc autour de la perception active et littérale de l'espace, au travers des moyens d'interactions qui sont proposés au joueur par les concepteurs de jeu. Cependant, de nombreux jeux possèdent des systèmes de collision/réponse détaillés, mais le joueur ne les expérimente que de manière indirecte, puisque décorrés de l'action du joueur. Alors, la conception des animations, d'effets visuels et de réponses sonores permettant de fournir des indices relatifs à leur composition physique (leur poids, leur texture, leur élasticité, ex.). Ces "indices" sont regroupés sous l'appellation du "polish", soit le processus de finition et d'affinage de l'expérience vidéoludique nécessaire à son appréciation.

4-A - Partie 1 - Le polish

Le polish fait référence à tout effet venant artificiellement améliorer l'interaction, sans altérer la simulation primaire. Par exemple, dans "Shadow of the Colossus", à mesure qu'un titan se déplace, des particules de poussière se soulèvent relativement à l'impact de ses pas, et le bruit de ces derniers est perçu plus ou moins fort à mesure que le joueur s'approche ou s'éloigne de la source. Le polish permet donc d'ajouter un intérêt et de mettre en emphase la nature physique des interactions, aidant les concepteurs de jeu à "vendre l'illusion" de ces objets au joueur, de façon plus ou moins réaliste. Comme on peut le voir avec l'exemple

ci-dessous, un personnage d'un jeu de combat, sans modèles et animations, revient à contempler d'étranges boîtes s'affronter.



“En assemblant ces trois éléments que sont le contrôle en temps réel, la simulation de l'espace de jeu et le polish, l'on arrive à la définition suivante: le contrôle en temps réel d'objets virtuels dans un espace simulé, au travers d'interactions soulignées par le niveau de polish.”, nous dit Steve Swink. Mais alors, comment cette notion de game feel se rattache-t-elle à l'expérience de l'état de flow en cours de partie?

4-B - L'identification du joueur à son avatar au travers de la conception sonore

“Un effet sonore peut complètement changer la perception du joueur vis-à-vis d'un objet dans un jeu vidéo. [...] Par exemple, l'un des animateurs de God of War avait créé une animation. Malgré la précision et la qualité de cette dernière, quelque chose manquait; l'animateur était sur le point de recommencer son travail, lorsque Derek Daniel, l'un des game designer du jeu, proposa d'ajouter un son à l'animation. Évidemment, c'était la pièce manquante du puzzle: soudain, le mouvement de Kratos devenait puissant, satisfaisant et violent.”, raconte Steve Swink. Comme l'on peut s'en douter, la conception sonore constitue la clé de voûte du game feel: c'est elle qui va permettre d'apporter une réponse sensorielle instinctive aux interactions du joueur, vis-à-vis de son avatar et des objets composant son environnement. “L'un des éléments qui semble être un aspect central de la rétention du flow en jeu, est quand le joueur reçoit des informations relatives à ses actions ou au contrôle de son avatar par le canal auditif, informations qui seraient alors perdues ou moins bien perçues si elles étaient uniquement diffusés par le canal visuel (les animations, l'interface, etc.). Par exemple, lorsque le joueur est sur le point de se retrouver à court de munitions, un feedback sonore tel qu'une série de cliquetis métalliques exagérés qui viendraient indiquer au joueur que son chargeur est presque vide, permet alors au joueur de recevoir l'information sans que ce dernier ai besoin de

détacher son regard de l'action pour examiner son compteur de munitions sur l'interface.", nous dit Jonas Turner, sound designer et compositeur sur bon nombre de jeux à l'aspect très "arcade" ("Nuclear Throne", "Scourge Bringer", "Downwell"...). Un tel procédé permettrait donc de fluidifier le "dialogue homme-machine", donnant un aspect plus instinctif à l'expérience de jeu et maintenant le joueur dans un rapport d'identification continu vis-à-vis de son avatar.

"Pour que le joueur ait le sentiment de réellement effectuer une action et contrôler son personnage, la question de la latence entre la pression d'un bouton et sa réponse sonore est cruciale. Si le moindre délai (réel ou perçu) vient s'interposer entre ces deux étapes, le joueur n'a alors plus la sensation qu'il est directement en contrôle de l'action", poursuit Turner, ce qui résulte en une dissonance sur le plan cognitif, susceptible d'interrompre l'état de flow. Or, même si aucun délai perceptible ne sépare l'événement de sa réponse sonore, il arrive souvent que l'effet sonore en lui-même comporte une certaine durée entre le début du fichier et la transitoire synonyme de l'impact de l'action effectuée. Cette erreur de conception, relativement commune, vient se mettre en travers du game feel, et donc de l'état de flow. Toutefois, la durée séparant l'input de la transitoire d'une action peut être utilisée afin de véhiculer certaines sensations au joueur: par exemple, si le personnage est pataud, ou soudainement trop lourd dû à un inventaire rempli, une telle latence permet de véhiculer cette sensation de lourdeur par le son. La série de jeux "Dark Souls" utilise ce procédé à de nombreuses reprises: par exemple, lors d'une roulade, le son sera déclenché avec plus ou moins de latence, afin de suggérer une lourdeur relative aux caractéristiques physiques de l'avatar du joueur (poids, nombre d'objets dans l'inventaire, niveau d'agilité, etc.). Écourter au maximum la durée d'un son au déclenchement relativement fréquent peut aussi permettre de favoriser la sensation de réponse et d'impact, tout en venant réduire la sensation de frustration pouvant être apportée par la répétition d'un tel son.

4-B - Les sons de l'interface comme substitution au besoin d'identification vis-à-vis d'un avatar

Selon Arnaud Roy, compositeur du studio français Amplitude, "le challenge des jeux de stratégie, c'est d'arriver à garder le joueur captif de l'expérience, en le poussant à jouer un tour supplémentaire. Au niveau de la conception sonore, cela passe principalement par l'interface utilisateur. Ainsi, toute notion d'avatar unique étant souvent absente dans les jeux de stratégie, ces derniers mettent en scène le contrôle d'un ensemble de personnages (troupes , armées, civilisation, etc.) au

travers des ordres qu'on leur donne. Pour ce faire, cet ordre, ou prise de décision, doit fournir suffisamment de feedbacks sonores afin de maintenir le joueur dans une relation tactile et physique vis-à-vis de l'expérience. "Ainsi, quand le joueur clique sur un élément, il reçoit un feedback sonore. Cela permet d'immerger ce dernier dans l'espace virtuel et de lui donner l'impression qu'il touche quelque chose de concret. Lorsque je clique sur une troupe, et que le général à sa tête se met à crier "A vos ordres", le joueur va avoir l'impression que cette troupe obéit au joueur, venant alors lui procurer un sentiment de puissance et de contrôle supplémentaire sur l'action. Aussi, pour tous les sons abstraits relatifs au divers menus d'un jeu, c'est la suite de sons d'interface qui va amener et maintenir le joueur dans un état de flow: ces sons possèdent chacun une texture, relative au contexte de l'histoire (période historique, univers réaliste ou de science fiction, etc.) et à un ensemble de propriétés physiques en résultant (organique, métallique, soyeux, brut, etc.).

4-C - La conception sonore au centre du game feel

Une autre problématique qui se pose est celle du lien qui rattache les éléments sonores au monde virtuel. Par exemple, si l'identité sonore d'un bruit de saut du personnage est perçue comme différente de celle de ses bruits de pas, une autre forme de dissonance peut alors se manifester, venant se mettre en travers de la perception des feedbacks sonores comme un ensemble cohérent par rapport au monde dans lequel ils surviennent. Toujours selon Joonas Turner, un moyen efficace afin d'éviter cette dissonance d'apparaître consiste à préserver une cohérence dans la chaîne analogique et audionumérique du processus d'enregistrement et de création sonore. "Au début de la production, je décide quel microphone, quel préamplificateur sera utilisé afin d'enregistrer une certaine catégorie de sons, et je maintiens ce choix pour l'ensemble des sons constituant cette catégorie tout au long de la création sonore. Ensuite, j'applique à tous les sons de cette catégorie la même chaîne de traitement, de façon à les "coller" ensemble. Enfin, je passe l'ensemble de toutes les catégories de sons au travers d'une même chaîne finale, venant alors unifier toute la conception sonore en la rattachant à une même identité. Cette identité sonore permet au joueur de ressentir les sons comme appartenant au même espace sonore, créant alors une consistance et un sentiment de cohérence qui rend l'ensemble des feedbacks plus fidèle par rapport au monde virtuel dans lequel ils interviennent. Ce processus permet ainsi de renforcer la sensation de game feel dans l'expérience du joueur".

Une autre approche intéressante de Turner consiste à utiliser des défauts techniques d'une implémentation, comme une distorsion causée par la superposition

importante d'un même type de son ou de plusieurs sons, afin d'accentuer la tension d'une situation précise en cours de partie. "Par exemple, si le joueur se retrouve dans une salle où des chevaliers en armures se mettent à attaquer en même temps le joueur, l'accumulation de feedbacks est susceptible de créer une masse sonore distordue, dûe au niveau globale de cette masse se heurtant au processeur dynamique en fin de chaîne servant à limiter le dépassement du seuil de distorsion. Mais, une fois l'affrontement terminé, tout redevient calme et silencieux." Utilisé judicieusement, ce genre d'approche permet de mettre en exergue les moments de tensions, tout en induisant une sorte de "réaction logique" des différents éléments composant le monde virtuel, ce qui permet au joueur de se le représenter de façon plus tangible.

Également, toujours selon l'interviewé, certaines techniques d'implémentation, comme celle du ducking (mentionné plus haut avec Doom Eternal), peuvent être utilisées comme outil créatif afin de renforcer la sensation de game feel, plutôt que d'être seulement employées afin de contourner un problème. Turner explique donc comment l'utilisation de filtres, contrôlés en sidechain par le déclenchement de certains sons, permet de renforcer la sensation d'impact de ces derniers. Par exemple, si une explosion ou un son au volume élevé retentit, un lowpass filter viendra se fermer puis se rouvrir pour souligner la puissance de l'action causée par le joueur. Cela donne alors à ce dernier une sensation de puissance et de plaisir de jeu, contribuant donc à maintenir l'engagement du joueur lors d'une partie.

Pour finir, Turner dit employer la technique suivante sur chacun de ses projets: tout d'abord, il découpe un son extrêmement court au milieu de la phase de sa forme d'onde, afin de créer un élément similaire un "clic" (son uniquement composé d'une transitoire, sans retombée aucune). Puis, il superpose cet élément au début de chacun des sons relatifs aux actions du joueur. Cela crée alors un son dont l'impact est ressenti instantanément à la lecture, permettant d'accentuer la sensation de puissance et de réponse des feedbacks sonores relatifs aux actions du joueur, tout en éliminant tout problème éventuel de latence perceptible.

Toutes ces techniques et procédés de créations permettent donc de créer un ensemble sonore dynamique, renforçant le sentiment de contrôle physique du joueur sur son expérience, et venant ainsi faciliter l'apparition et le maintien de l'état de flow.

5. La conception et implémentation de l'interactivité sonore, ciment du Flow

Précédemment, nous avons pu voir que la compréhension et l'utilisation pertinente des caractéristiques sonores composant une bande sonore, sont généralement liées à la création et l'implémentation de systèmes interactifs lors de la création de l'expérience vidéoludique. Maintenant, nous allons explorer, au travers de diverses sous parties, comment l'intégration des différents éléments sonores constitue un facteur primordial de l'apparition et de la préservation de l'état de flow.

5-A - L'implémentation de transitions fluides et naturelles

Contrairement au cinéma où les transitions sont fixes et souvent utilisées comme effet de style, les transitions dans une expérience vidéoludique sont définies par plusieurs variables, liées soit à un état (victoire, défaite, mort, réapparition...), soit à des raisons techniques (économie de mémoire, chargement d'éléments visuels, etc.), soit à une évolution narrative (cinématiques). Or, ces transitions sont souvent sujettes à une rupture drastique et brutale de l'état de flow, brisant alors le sentiment de continuité instauré jusqu'à lors.

Ce faisant, certains tentent, et ceux depuis les années 90, de faciliter ce processus de transition d'une scène vers une autre afin de préserver l'état de flow. D'abord, par l'élaboration de divers artifices visuels ou de game design: on peut naturellement citer l'exemple de "Resident Evil" premier du nom qui, afin de maintenir la tension en jeu lors des temps de chargement, dissimulait ces derniers par une animation d'ouverture de porte sur un fond noir. Cela créait un sentiment d'anticipation, de mystère, qui maintenait le joueur en haleine. Or, ce type de solution est rudimentaire: répétée un certain nombre de fois, il finit tout de même par rompre le flow et frustrer le joueur, parfois même plus que si le jeu en question s'était contenté d'un simple temps de chargement avec une image fixe.

Pourtant, 10 ans en arrière, l'équipe audio de LucasArt (composée de Michael Land et Peter McConnell) avait réfléchi et réussi à habilement contourner cette problématique via l'utilisation du moteur musical "iMUSE" (pour "Interactive Music Streaming Engine"), un système de musique interactive consistant à créer des transitions d'un thème musical à un autre de façon fluide et naturelle, permettant alors le changement d'un tableau vers un autre sans que cela vienne interrompre l'état de flow et l'immersion du joueur. McConnell décrivait iMUSE comme un système agissant à la manière d'un orchestre de fosse (type d'orchestre

accompagnant les acteurs dans les comédies musicales, opéras et ballets), déclenchant des séquences musicales plus ou moins longues suivant la vitesse à laquelle le moteur de rendu préparait la scène suivante.

Cependant, si ce type de système demandait à l'époque un coût en ressources important, les outils (Wwise, FMod) aujourd'hui mis à disposition permettent d'obtenir un résultat tout aussi probant, au moyen d'une rapidité de production et d'implémentation bien supérieure. Au travers de la conception sonore du jeu "INSIDE", créé par studio Playdead Games et mentionné plus haut, le directeur audio Martin Stig Andersen adopte une approche intéressante: afin de conserver une continuité absolue, le moteur sonore (ici, Wwise) est toujours actif, de l'écran de menu au jeu en lui même. Ainsi, chacun des sons sont divisés en deux catégories: les sons redémarrés lors de la réapparition du joueur, et ceux n'étant jamais interrompus (comme la musique, les sons d'ambiances, etc.). Un tel choix permet alors d'éviter toute transition brusque, et de garder le joueur dans une expérience homogène sans interrompre l'état de flow.

5-B : L'utilisation de la musique interactive comme vecteur de flow

"L'augmentation de la qualité du son dans le jeu vidéo au fil des générations et des avancées technologiques a conduit à un désir de maîtriser la conception sonore de ce média - notamment via la musique interactive -, de manière à pouvoir plus facilement introduire cet état de flow, ou du moins les composantes qui le constituent (la concentration intense, l'engagement, l'altération temporelle, etc.).", nous dit Masami Komuro. "Déjà, continue-t-il, au travers de la conception sonore du jeu de stratégie "Z", sorti en 1996 par le studio The Bitmap Brothers, les développeurs de jeu vidéo commençaient à expérimenter la notion de musique interactive en tant que moyen de communication, afin de transmettre au joueur de manière plus fluide et instinctive les informations nécessaires à la compréhension de l'action (et ainsi faire apparaître et maintenir l'état de flow), notamment par la lecture séquentielle de segments MIDI suivant que le joueur construise de nouveaux bâtiments, de nouvelles troupes, etc."

Un autre aspect de la musique interactive, consiste en sa capacité à pouvoir apporter une forme de variété, de continuité à un cœur de gameplay relativement

répétitif. “Le plus gros challenge de la musique interactive, c’est la répétition.”, me répond Olivier Derivière, suite à une question posée lors d’une masterclass organisée par “Jeux Vidéo Magazine” à la Cité des Sciences de Paris. “Un jeu vidéo est généralement constitué d’une boucle de gameplay principale, qui va se répéter au fil de l’expérience. Si l’aspect du monde dans lequel on se retrouve vient à changer (monde du feu, monde de la glace...), la boucle de gameplay, elle, n’évolue que très peu, voire pas du tout. Alors, tous les corps de métier du développement vidéoludique tentent de briser cette redondance. Or, la musique possède une grande capacité à pouvoir casser le sentiment de répétition.”, continue ce dernier. Ainsi, le compositeur de jeu vidéo doit, en amont du processus de création, réfléchir à l’agencement de sa ou ses compositions, tout d’abord afin d’assurer des transitions fluides entre chacun des éléments sonores, et ensuite, afin de prévenir tout sentiment de frustration lié à la répétition abusive de ces éléments. Pour cela, trois approches sont généralement employées:

- L’approche horizontale, pour reprendre les mots d’un article de Gamekult sur la musique adaptative (rédigé par Valentin Cébo, alias Noddus), “concerne de larges sections de musique, dans lesquels le moteur audio va pouvoir piocher afin d’habiller une séquence de jeu en s’appuyant sur “l’état de jeu” qu’il identifie. Ainsi, en conservant le même BPM, parfois la même mélodie, le jeu fera passer la musique d’un état à l’autre, en accentuant parfois les 4 à 8 derniers temps d’un morceau, et créant ainsi une transition fluide vers le suivant.”
- L’approche verticale, consistant à empiler différentes couches musicales, afin de les activer ou désactiver suivant certaines variables et changements dans le gameplay. Pour le projet “Dark Sense” sur lequel je travaille, j’utilise notamment cette méthode afin de “fade in” ou “fade out” certaines pistes musicales et rythmiques, suivant l’évolution d’une variable faisant référence à la vitesse de déplacement du joueur.
- L’approche générative, qui consiste à générer en temps réel une partition musicale jouée par le moteur audio, suivant les différentes actions et choix effectués par le joueur. C’est notamment via cette méthode que Richard Vreeland, alias “Disasterpeace”, conçoit la bande son du jeu “Mini Metro”: dans ce jeu de puzzle contemplatif consistant à créer et agrandir un réseau métropolitain, chacune des actions du joueur vient amener un motif musical, permettant alors de créer un ensemble qui fait office de feedback direct aux choix du joueur.

Si ces 3 approches relativement distinctes représentent à elles seules une philosophie bien particulière, il n’est pas rare qu’une conception sonore utilise plusieurs de ces approches avant de séquencer ses éléments musicaux. Toutefois, une distinction importante doit être faite entre “musique adaptative” et “musique interactive”. Selon moi, la musique adaptative est un système de musique qui vient s’adapter aux évènements en jeu, corrélés ou non aux actions du joueur, afin d’offrir

des transitions et une synchronisation image/son similaire à celle proposée par le cinéma. Or, la musique interactive est pour moi un système venant principalement se baser sur les actions et contrôles directs du joueur, afin de séquencer les différents éléments musicaux composant la bande son. On aurait donc d'un côté, une forme de systèmes venant s'adapter, anticiper les événements en jeu, et de l'autre des systèmes interactifs, créant un dialogue alimenté par le comportement du joueur. Certains compositeurs, comme Olivier Derivière, mettent un point d'honneur à séparer ces deux terminologies, ayant selon eux une définition propre à chacune. Dans la ludographie de Derivière, on retrouve notamment cette approche avec *Get Even*, qui vient générer et séquencer la bande son en temps réel en fonction des variables relatives au joueur (sa proximité vis-à-vis de certains éléments, les battements de cœur du personnage, etc.). Or, un autre jeu de sa ludographie, vient explorer de plus belle ce concept de "musique interactive": "*Remember Me*", jeu d'action/aventure de science-fiction, sorti par Dontnod en 2013. Lors des phases de combat qui rythment la progression, la musique est séquencée de manière horizontale suivant les divers combos du joueur, selon qu'il s'éloigne des ennemis ou se rapproche du combat, qui se fasse attaquer... Cela crée un résultat dynamique qui vient supporter l'état de flow du joueur à chaque instant, le plaçant dans une position de chef d'orchestre. Cependant, selon Derivière, "la question reste d'arriver à conserver et maintenir une cohérence derrière les systèmes créés. Dans *Remember Me*, les différents combos correspondent tous à des phrases de 8 mesures, se terminant par le même motif rythmique." Ainsi, tout en cassant la répétition et en venant créer un dialogue entre les actions du joueur et la musique en jeu, la conception sonore de *Remember Me* parvient à développer des systèmes efficaces afin d'amener et maintenir l'état de flow du joueur dans l'expérience. Cependant, une telle symbiose entre musique et gameplay n'est rendue possible que si, selon Olivier Derivière, "le compositeur comprend le langage des développeurs, et assimile le gameplay comme pièce centrale de l'expérience". Or, toujours selon l'intéressé, un tel travail est souvent ignoré: "Malheureusement, la plupart des compositeurs de jeux vidéo ne jouent pas ou très peu au jeu pour lequel ils composent." Enfin, il conclut avec l'affirmation suivante: "L'on dit souvent que la conception sonore constitue le parent pauvre du jeu vidéo, mais selon moi, ce n'est pas vrai: c'est ceux qui corroborent ce mythe qui contribuent à le rendre tel quel."

5-C : L'interprétation des données utilisateur comme moteur de l'interactivité

“En tant que concepteur sonore, j’essaie d’amener les différents éléments musicaux et de conception sonore de sorte à ce que ces derniers semblent réagir ou être la résultante des actions effectuées par le joueur. Parfois, c’est le cas, et d’autres fois, c’est une illusion. L’idée est de faire en sorte que le joueur se sente en symbiose avec le monde dans lequel il se retrouve plongé. De cette façon, ce dernier perd la notion du temps et entre dans un état de flow.”, nous dit Paul Weir, directeur audio chez Hello Games (“The Last Campfire”, “No Man’s Sky”). La musique de No Man’s Sky est, au même titre que le monde du jeu, générée (ou plutôt déclenchée) de façon procédurale. En découpant et divisant l’ensemble de la bande sonore (fournie par le groupe de post-rock instrumental “60daysofstatic”) en une multitude de nappes, Weir construit au fur et à mesure d’une partie un ensemble musical, venant accompagner le joueur lors de son exploration. Or, la problématique suivante apparaît: “En jeu, l’on peut agencer un ensemble sonore qui varie constamment, sans que cette variété soit perçue par le joueur moyen. Ainsi, malgré la complexité et le nombre de systèmes permettant de générer cet ensemble musical en perpétuel changement, l’on installe une forme de redondance auditive dans l’expérience du joueur. Pour empêcher cela, des systèmes viennent analyser en temps réel les actions du joueur (déplacements, angles de la caméra, distance séparant le joueur de certains éléments, etc.), et renvoient ces informations au moteur sonore sous forme de variables, qui vient ensuite diriger la création et l’agencement des éléments musicaux en temps réel en fonction de celles-ci. “En un sens, les entrées de ces variables n’importe que peu, tant qu’elles prennent leur source dans le comportement du joueur”, poursuit-il. Ainsi, en créant une variété sonore qui puise dans les actions du joueur plutôt que sur l’aléatoire, l’on crée une conception sonore qui vient presque effacer toute notion de répétitivité, tout en instaurant un dialogue permanent entre le joueur et le monde qu’il vient habiter. Ce procédé permet alors de ne jamais mettre la conception sonore en travers de l’expérience de flow du joueur. Cependant, cela n’est rendu possible que si, selon Olivier Derivière, “le compositeur comprend le langage des développeurs, et assimile le gameplay”. Or, toujours selon l’intéressé, un tel travail est souvent ignoré: “Malheureusement, la plupart des compositeurs de jeux vidéo ne jouent pas ou très peu au jeu pour lequel ils composent”.

5-D : La nécessité de concevoir une implémentation modulaire et flexible

“Dans No Man’s Sky, le monde est généré de façon procédurale. Cela implique que, contrairement à un jeu possédant une structure plus traditionnelle, il nous est impossible de prévoir les actions du joueur à l’avance, ainsi que le lieu et le contexte dans lequel il les effectue. Cela m’a amené à concevoir une implémentation sonore agencée autour de systèmes modulaires, plutôt que sur une implémentation axée sur un ensemble d’événements. “, nous explique Paul Weir, dans une conférence donnée à la GDC sur la conception sonore de No Man’s Sky.

- Une implémentation axée autour d’événements est conçue en relation directe avec un événement précis: le joueur effectue une action, ce qui vient déclencher un ou plusieurs sons. Un tel système est relativement figé: à mesure que la taille du projet augmente, l’on doit continuer à apporter de façon cumulative et basique une “solution” pour chaque situation envisageable, rendant l’organisation de l’implémentation complexe et étriquée, en plus de constituer une perte importante en temps et en ressources.
- En revanche, une implémentation agencée autour de systèmes modulaires, se basant sur la connaissance des différents états du joueur et des variables qui lui sont associées, est plus facilement expansible, car en “relation indirecte” avec un événement précis.

“Si, continue-t-il, j’utilise dans certaines situations une approche événementielle, comme pour les sons d’interface et certaines animations du joueur, je m’applique à adopter cette approche centrée autour de systèmes pour le reste des éléments sonores, notamment pour les sons d’ambiances: dans No Man’s Sky, chaque planète possède un type défini (volcanique, océanique, jungle, désertique, glaciaire, etc.), ainsi qu’un ensemble de caractéristiques (température, composition élémentaire) qui lui sont propres. Avec une approche systémique, si les game designers décident d’implémenter de nouveaux types de planètes, je n’ai plus qu’à créer un nouveau label dans Wwise (appelé “State”), à partir duquel je peux allouer de nouveaux sons, variables ou conditions, tout cela sans avoir besoin de créer de lignes de code supplémentaires dans le moteur de jeu. Une telle flexibilité au niveau de l’intégration me permet d’atteindre, en tant que concepteur sonore, les objectifs suivants: en premier lieu, de créer un environnement qui réagit de manière dynamique et convaincante; en second lieu, d’inclure une multitude de couches de détails afin d’enrichir l’expérience du joueur; en dernier lieu, de conduire et agencer la conception sonore à partir de données de jeu, permettant alors de simplifier le dialogue entre la conception sonore et le comportement du joueur lors de sa partie.”

“De tels systèmes permettent alors de faire apparaître des moments clés chargés d’émotion dans l’expérience du joueur, sans que cela soit planifié, ou pire, lui soit imposé à l’avance: un joueur m’a un jour contacté pour me raconter comment, après s’être mis à l’abri dans le cockpit de son vaisseau lors d’une tempête, il entendait le bruit de la pluie s’écraser sur la paroi de son engin, et la manière dont cela lui rappelait certains moments de son enfance, à attendre dans le camion de son père que la tempête se dissipe.”. Ainsi, en basant la majorité de l’implémentation sonore de No Man’s Sky autour de systèmes modulaires et complémentaires, Paul Weir arrive à maintenir l’état de flow du joueur, tout en créant des moments de jeu et souvenirs d’une rare pureté.

5-E - L’optimisation de l’implémentation sonore et de ses systèmes

L’optimisation de l’implémentation sonore et des systèmes qui la constitue, sujet en apparence très technique, est également liée à l’apparition et au maintien de l’état de flow. Ce processus permet de prévenir et d’empêcher tout heurt technique (problèmes de déclenchement, de mixage des sons, etc.), susceptible de se mettre en travers de l’expérience du joueur. Si, dans la partie précédente, nous avons pu voir comment l’implémentation sous forme de systèmes permettait d’obtenir une conception sonore réactive et polyvalente, il convient maintenant de passer en revue les problématiques liées à l’optimisation en temps réel de ces systèmes.

5-E - Partie 1 - Le processus de virtualisation

A mesure que les environnements de jeu deviennent de plus en plus gigantesques, notamment via la démocratisation des mondes ouverts depuis une quinzaine d’année, les développeurs de jeu peuvent de moins en moins se contenter de charger un même grand niveau et tout ses composants, avant de laisser le joueur s’y mouvoir (pour des raisons évidentes de performance et de mémoire). Ce faisant, les titres actuels se concentrent de plus en plus sur les problématiques liées au “streaming d’assets”, c’est à dire le fait de compresser et décompresser rapidement les divers éléments nécessaires à l’expérience de jeu (les modèles et objets graphiques, le son, les effets visuels, d’ombrage, de texture, etc.), selon la nécessité de leur présence à un moment T. Pour donner un exemple, dans GTA V, lorsque le

joueur est trop loin pour voir distinctement un élément visuel (comme une poubelle, une voiture, un panneau d'affichage, etc.), cet élément est momentanément compressé ou chargé dans la mémoire RAM, afin de pouvoir être rapidement décompressé et chargé si le joueur se rapproche à nouveau de cet élément. Ce concept s'applique également aux éléments sonores: afin d'économiser de la mémoire et de la puissance de calcul, le nombre de voix (c'est-à-dire de canaux sonores) en simultané est limité. Alors, il convient de créer une hiérarchie claire et organisée de tous les éléments sonores, relative à leur niveau d'importance et d'occurrence dans la scène, afin de décider en temps réel et selon une multitude de variables, lesquels sont lus, temporairement chargés dans la mémoire RAM afin d'être accessible presque instantanément, ou désactivés. Ce procédé est appelé "virtualisation", et est utilisé dans la plupart des productions vidéoludiques de moyenne et grande taille, afin d'optimiser la performance de l'expérience. Au fil de l'évolution des composants d'ordinateur et des générations de console, le nombre de voix pouvant être jouées en simultané a considérablement augmenté, permettant de créer des environnements sonores d'une qualité en constante expansion. Cela dit, la priorité de puissance de calcul est toujours accordée en grande majorité à la partie visuelle de l'expérience, induisant alors une certaine limitation dans le nombre d'éléments sonores pouvant être actifs au même moment; cela peut alors devenir un problème lorsque l'on souhaite implémenter un certain nombre de variétés sonores pour un même type de son. L'organisation et la hiérarchisation des voix deviennent alors cruciales, afin de pouvoir créer un environnement sonore aussi réactif que complet, paré à tout type de comportement potentiel de la part du joueur. Cette hiérarchisation est généralement effectuée via le traitement des métadonnées inscrites dans chacun des fichiers audio et midi, permettant alors de fournir une organisation modulaire et extensible pour la totalité du contenu relatif à la conception sonore.

5-E - Partie 2 - Le mixage des sources sonores en temps réel

Une autre problématique, plus ou moins abordée dans les parties précédentes, est celle du mixage dynamique des différents éléments entre eux. Dans Grand Theft Auto V, l'équipe audio, via le moteur de jeu maison des studios Rockstar intitulé "RAGE" (Rockstar Advanced Game Engine), crée un ensemble d'outils et de processus afin de fournir un impact cinématographique à la direction sonore, notamment par l'équilibre des différentes sources qui la compose:

- Tout d'abord, la partie "statique" du mixage est agencée selon une hiérarchie regroupant plusieurs catégories, ou groupes (ambiances, bruits de pas, actions diverses, bruits de tirs, bruits de voiture, etc.), eux-mêmes définis par leur niveau d'occurrence, leur intensité de volume...
- Une fois que cette balance statique est effectuée, ces groupes sont ensuite assignés à des altérations spécifiques (dynamiques, fréquentielles, etc.), le

plus souvent contrôlées par l'évolution d'une ou plusieurs variables pour chacun de ces changements. Ces altérations peuvent également être appliquées à l'ensemble des catégories du jeu, comme par exemple lorsque l'avatar du joueur meurt.

- Puis, différents réglages de ces altérations sont assignés aux différents types de scènes rencontrées en jeu (dans la voiture, en vue cockpit, en vue à la 3e personne, à pied sur l'autoroute, à pied dans la nature, etc.), afin d'asseoir le réalisme des différents espaces sonores au joueur, ainsi qu'à faire s'alterner ces derniers de façon convaincante, afin de créer un monde persistant et crédible.
- "La série de jeux GTA a toujours utilisé les radios de voitures comme principale source de musique. Dans GTA V, nous avons fait le choix d'ajouter à cela une bande originale venant accompagner le joueur même hors de contexte, afin d'opter pour une approche plus cinématique. Nous devons donc trouver un moyen d'effectuer les transitions entre la musique provenant des radios et celle provenant de la bande originale d'une façon agréable et naturelle pour le joueur." nous dit Alastair Mc Gregor, directeur audio chez Rockstar Games. "Notre approche vis-à-vis de la musique, est que le joueur ne devrait jamais percevoir le démarrage d'un morceau, ce dernier devant s'inscrire dans une continuité par rapport à la bande sonore globale afin d'éviter toute cassure.", continue-t-il. Pour ce faire, l'équipe audio sépare chacun des morceaux de la bande originale en 8 pistes stéréo. Des paramètres de mixage prédéfinis et des transitions précises sont ensuite appelées par le moteur audio, selon l'ambiance souhaitant être véhiculée à un moment (calme, énergique, tendue, etc.). Ces transitions d'ambiances sont appelées par un système d'événements, et certains motifs musicaux, ou "stingers" (soit des motifs synchronisés au tempo actuel du morceau en train d'être lu) sont parfois déclenchés afin d'appuyer une transition ou une action.

Le but de ces manœuvres en temps réel est de faire en sorte qu'elles soient suffisamment marquées pour avoir un effet positif dans l'expérience, mais qu'elles demeurent suffisamment implicites et subtiles de sorte à ce que le joueur ne les remarque pas, afin d'installer l'immersion et renforcer l'impact émotionnelle d'une scène. Ces systèmes de mixage dynamique viennent alors apporter un sentiment de cohésion et de continuité à l'expérience sonore, ce qui participe naturellement au maintien de l'état de flow lors d'une partie. Toutefois, comme le dit Paul Weir: "Les systèmes relatifs au mixage dynamique de la conception sonore dépendent vraiment du game design et du jeu en lui-même; il est impossible d'offrir une réponse ou d'élaborer une structure pouvant s'appliquer à tous les différents cas de figure, leur nombre étant presque infini. Il convient alors au concepteur sonore d'évaluer les besoins et particularités qui constituent l'expérience de jeu, afin de pouvoir élaborer et mettre en place les systèmes nécessaires au maintien de l'équilibre sonore en cours de partie."

5-F - Le budget, obstacle à une intégration sonore complexe selon l'échelle des productions

“Je pense que c’est possible de faire de la jolie musique interactive, mais que c’est beaucoup plus long, coûteux et compliqué. Si l’on veut créer des compositions musicales qui arrivent à dépasser leur structure primaire de boucle afin de procurer une émotion forte au joueur, au delà de leur fonctions utilitaires (c'est-à-dire, procurer une tension ou un sentiment d’apaisement selon le contexte), c’est assez difficile, et toutes les productions ne peuvent se le permettre.” nous dit Arnaud Roy. Selon lui, à budget équivalent, l’on peut avoir “une heure de musique à la structure standard”, et “vingt minutes de musique interactive”, ce qui revient donc à coûter trois fois plus cher. C’est pour cela que la plupart des studios de développement préfèrent une approche quantitative à celle d’une approche qualitative, afin de briser le sentiment de répétition avec le moins de moyens possible. Or, ce postulat peut être contesté: de la même manière qu’une implémentation de systèmes modulaires, plus complexe en amont, permet par la suite une croissance de la variété sonore et du nombre d’assets plus optimale, un ensemble musical moins fourni en nombre d’assets mais offrant alors une variété bien supérieure en terme de combinaisons et d’interactions permet de fournir une expérience plus réactive, agréable et impactante au joueur, en venant prévenir plus efficacement tout sentiment de répétition. “Cela dit, le problème de l’élaboration de ce genre de systèmes est qu’ils se basent le plus souvent sur une importante masse de lignes de code additionnelles, nécessitant des programmeurs spécialisés dans l’audio, ce qui coûte relativement cher. Au-delà d’une certaine échelle de production, le compositeur seul ne peut assurer à la fois la création de la conception sonore et celle des systèmes qui l’agencent, et en même temps se charger de l’implémentation de ces derniers.”, conclut-t-il. Ainsi, la question budgétaire serait directement corrélée au potentiel de flow d’une expérience vidéoludique.

Dans un article de Gamasutra (reprenant les mots d’un numéro du magazine “Game Developer” publié en septembre 2005), Alexander Brandon, compositeur de jeu vidéo ayant notamment travaillé sur des titres comme “Unreal Tournament” ou “Deus Ex”, essaye de donner une idée de la grille tarifaire concernant la création d’effets sonores et la composition musicale. Selon Brandon, le taux standard de l’industrie par minute de musiques approuvées correspondrait à un nombre allant de 1000 à 1200 dollars. “Lorsque vous soumettez un budget pour un travail de composition, vous devez estimer le nombre de minutes requises pour de la musique originale en jeu. Le nombre d’MR (“Minutes Required”) est actuellement autour de 2 à 3 minutes de musique originale par niveaux, mais la conception d’une bande son adaptative permet souvent de réduire ce nombre et donc diminuer les coûts”. Même si l’article date d’une bonne quinzaine d’années, l’on peut supposer que cette affirmation reste encore vraie et que dans certaines situations, la conception d’une bande son

adaptative peut figurer en tant qu'alternative économique plutôt que coûteuse. Brandon continue, en décrivant cette fois les tarifs moyens relatifs à la conception d'effets sonores: "Cela varie d'environ 5 dollars par son relativement simple (comme des bruits de pas), à 50 dollars par son plus ou moins complexe (comme un son de machine gun nécessitant l'élaboration d'une ou plusieurs boucles lu selon un système plus ou moins coûteux en temps de production). La formule que j'utilise pour estimer le coût d'un effet sonore est la suivante: $(FX\ Rate \times Production\ Rate\ (Simple\ Effect)) + (FX\ Rate \times Production\ Rate\ (Complex\ Effect)) =$ le montant estimé pour la production et la livraison d'un effet sonore."

Pour finir cette partie, l'on peut à nouveau citer Paul Weir: "Le problème avec l'implémentation de systèmes de plus en plus complexes et réalistes, est que passé un certain cap de détail, le joueur moyen n'est plus à même de percevoir la différence vis-à-vis d'un système plus simple. Il devient alors difficile, pour le concepteur sonore comme le ou les responsables budgétaires, de justifier l'implémentation de tels systèmes. Il s'agit donc pour le concepteur sonore de questionner chaque système en fonction de leur ratio coût/accroissement de la qualité sonore perçue, afin de pouvoir correctement attribuer les ressources et les effectifs nécessaires à l'élaboration de la conception sonore d'un jeu."

5-G - La dé-corrélation entre complexité et efficacité d'une implémentation sonore

Comme nous avons pu le voir jusqu'à présent, l'élaboration de systèmes complexes permettant d'agencer les différents éléments d'une conception sonore comporte ses avantages en termes de potentiel de flow, comme ses désavantages en termes de coût. Cependant, suivant le type de jeu, il n'est pas forcément nécessaire de concevoir une multitude de systèmes complexes afin de plonger et maintenir le joueur dans un état de flow. Prenons pour exemple le cas du jeu "Hotline Miami", sorti en 2012 et ayant marqué plus d'un joueur par sa conception sonore brutale et galvanisante, agrémenté de visuels d'une rare violence, bien que très minimaliste visuellement parlant. Ici, la musique du jeu est composée de morceaux fixes, jamais séquencés en fonction de l'action, et les différentes actions du joueur et des ennemis qui l'entourent (bruits de tir, de portes qui s'ouvrent et se ferment, sons de mort des ennemis, etc.) se basent uniquement sur une approche événementielle (une action/un événement = un son). Pourtant, le choix des musiques et la conception même des feedbacks sonores est telle, que le joueur reste scotché à sa chaise du début à la fin d'un niveau, et quelques soient le nombre de tentatives pour

y parvenir. Cela est dû à deux choses: premièrement, les différentes musiques qui composent la bande son, choisies avec soin par les développeurs et regroupant une multitude d'artistes électro, comportent toutes un rythme et une structure très entraînante, bien que relativement répétitives. La partie musicale vient alors entrer en symbiose avec le rythme du jeu, permettant au joueur de se retrouver dans un clip musical grandeur nature dont il serait le principal sujet. Deuxièmement, l'impact et la conception sonore maîtrisée des armes à feu et des actions du joueur, viennent "couper" au travers de l'ensemble musical, permettant de donner au joueur tantôt un sentiment de puissance extrême relative à ses actions et ses enchaînements de tueries, tantôt un sentiment d'adrénaline permanent, décuplé par le côté "die and retry" du game design (c'est à dire que le joueur meurt à l'impact du moindre coup lui étant porté, le forçant alors à recommencer au dernier point de sauvegarde). Ainsi, comme le dit Masami Komuro, "l'interactivité sonore permet de faciliter et fluidifier le dialogue entre le joueur et les conséquences de ses actions, cependant, selon le game design et le concept de jeu, je ne dirais pas qu'elle est nécessaire. Hotline Miami utilise d'autres techniques pour amener l'état de flow, notamment l'application d'un contraste fort entre, d'une part, la zone visuelle de l'écran et les espaces très réduits du level design, et d'autre part, les mouvements et actions du joueur très amples servis par une bande son très dense. Cependant, si la conception de systèmes complexes et interactifs peut être, suivant les différents cas de figures, dé-corrélés de l'efficacité de l'implémentation, plus l'espace de jeu est grand et libre (comme dans un monde ouvert), plus le nombre de problématiques émergentes viennent nécessiter une réflexion poussée et ambitieuse de l'interactivité sonore et des systèmes qui la constituent.

Dans cette partie, nous avons pu voir comment la conception de l'interactivité sonore pouvait constituer le ciment de l'apparition et du maintien de l'état de flow. Nous allons maintenant explorer plus en profondeur le sujet des dangers de ruptures du flow relatifs à la conception sonore.

6. Les dangers de rupture du Flow liés à la direction sonore

6-A - L'abus de stimuli sonores, ou la peur du vide

“Récemment, j’ai pu regarder plusieurs heures de gameplay du jeu Far Cry 5, et j’ai eu l’impression de contempler un enfant hyperactif gesticuler dans tous les sens. On aurait dit que le titre refusait de fournir la moindre pause au joueur, l’assaillant constamment d’activités ou d’altercations avec divers ennemis, venant se mettre en travers de ses déplacements. Ce n’est pas la première fois que je ressens cette sensation d’étouffement dans un jeu vidéo, mais c’est pourtant la première fois qu’elle m’apparaît de manière aussi nette, alors que je n’ai même pas la manette en main.” nous raconte le YouTuber “Pseudoless”. “Pourquoi le calme semble être devenu un problème pour le jeu AAA grand public? La plupart des jeux en monde ouvert à gros budget partagent fréquemment cette peur liée à l’ennui ou au désintérêt du joueur. Seulement, la réponse de surface choisie afin de contrecarrer ce phénomène est disproportionnée. C’est comme si les concepteurs de jeux se mettaient à agiter un trousseau de clé près des oreilles du joueur, dans le but de le stimuler constamment.”, continue-t-il. Cette sensation de frustration liée au sur-remplissage de l’espace dû à une peur du vide ne date malheureusement pas d’hier. Or, la direction sonore, victime de ces choix malheureux de game design, ne vient qu’accentuer cette frustration. “Parfois, préférer une approche qualitative à une approche quantitative permet de renforcer l’expérience de jeu.” Ce postulat, naturellement applicable à la direction sonore d’un jeu vidéo, est notamment celui adopté par la série des jeux de thriller horrifiques “Silent Hill”. “Silence is also a sound.”, comme aime le rappeler le compositeur phare de la série, Akira Yamaoka. “In Silent Hill 2, we tried to create a sense of fear sustained by silence. That silence lying within each individual, which is progressively transformed into anxiety. It was a kind of fear that built up little by little. Gradually, it escalated into horror.” A travers ces dires, on comprend que la force de la conception sonore de Silent Hill 2 ne réside pas tant dans l’agencement de systèmes complexes (de toute façon forcément rudimentaires vu le contexte d’époque des jeux en 3D), mais plus dans l’introduction graduelle du joueur à l’effroi. Utilisant la règle sacrée du “show, don’t tell”, la conception sonore sait se faire très légère et subtile, au fur et à mesure que le personnage de James Sunderland s’enfonce dans la ville fantôme de Silent Hill, sorte de songe comme tout droit sorti de l’imaginaire torturé de David Lynch. Par exemple, lors de l’arrivée du joueur dans la ville, le joueur est assailli de toute part par des ennemis émergeant de la brume suffocante, dont la présence et le bruit de

radio venant alerter James de leur présence remplit l'espace sonore. Mais soudain, lorsque le joueur s'aventure dans le bâtiment résidentiel qui constitue le premier niveau du jeu, le silence vient engloutir le joueur, ne laissant audible que le craquement des pas de son avatar sur le sol. "I think that selecting moments of silence is another way of producing sound.", nous dit Yamaoka, concept qu'il applique pour la totalité de la conception sonore de la série d'épouvante. Or, l'approche de la peur par l'utilisation du silence et de la suggestion est ici double. En effet, le joueur moyen ayant été déshabitué à la présence du silence, au travers des diverses œuvres audiovisuelles rencontrées (jeux, cinéma), le forcer d'un coup à se confronter à sa peur interne du vide via la conception sonore vient friser le coup de génie. C'est ce qui démarque encore aujourd'hui la licence Silent Hill des autres jeux d'horreur sur le marché: cette capacité à jouer avec les attentes et la psychologie des joueurs, afin de décupler le potentiel d'effroi de l'œuvre. C'est également cette utilisation de la conception sonore qui vient inscrire Silent Hill 2 en tant que référence intemporelle de la peur, même 15 ans plus tard. Toutefois, le remake sorti en 2019 du célèbre jeu d'horreur "Resident Evil 2" adopte une approche à la fois similaire (notamment sur sa gestion du rythme ponctué par des silences), et à la fois opposé: en effet, dans le jeu, notre avatar est traqué en permanence par une entité monstrueuse aussi imposante que menaçante, nommée "Mr. X". Afin d'asseoir la menace constante qu'est censé représenter cet antagoniste, les concepteurs sonores ont eu l'idée de faire entendre de manière continue les lourds bruits de pas de notre opposant, qui se rapproche constamment du joueur où qu'il se trouve. Ces sons de pas sont si bien localisés, (notamment via la simulation d'effets d'obstruction et d'occlusion), qu'ils permettent de plonger le joueur dans une pression et une vigilance constante vis-à-vis du danger. Ainsi, notamment via sa conception sonore judicieuse, ce remake de Resident Evil 2 parvient à véhiculer les émotions et moments de tensions induits par son game design, sans venir lasser ou frustrer de façon involontaire le joueur.

Cependant, si des jeux d'horreurs ou assez lents en terme de rythmes peuvent se permettre d'utiliser une telle conception sonore afin d'asseoir leur ambiance, ce n'est pas forcément le cas d'autres titres, comme ceux d'action, qui, de par leur game design, peinent à échapper à la malédiction du remplissage sonore. Or, certains titres, comme le jeu compétitif de Blizzard Entertainment "Overwatch", parvient à implémenter au travers de divers systèmes de récolte de données utilisateur un mixage dynamique via une hiérarchisation de l'importance des événements entourant le joueur, ce qui permet d'éviter l'abus de stimuli sonore, tout en venant accentuer la lisibilité de l'action.

Un de ces systèmes consiste, via l'approche "dense clarity, clear density" (mantra de Walter Murch, monteur son sur Apocalypse Now), à supprimer des éléments du mixage afin de laisser plus de place aux éléments principaux de la conception sonore sans que cette suppression soit remarquée par l'auditeur. Pour cela, l'équipe audio d'Overwatch attribue chacun des événements sonores à une hiérarchie de priorités, basée sur un "niveau de menace" calculée en temps réel: en résumé, les

éléments comportant le plus grand niveau de menace seront mis en avant dans le mixage, tandis que ceux possédant un niveau moindre seront mis en retrait, voir désactivés. Un tel système permet au joueur d'identifier et de se concentrer plus facilement sur les feedbacks nécessaires à sa survie, ce qui vient clarifier l'action et lui donner un sentiment de contrôle et de maîtrise supplémentaire vis-à-vis de son environnement. Ainsi, en orientant la conception sonore autour d'une représentation claire de l'information, plutôt qu'en une sur-représentation basique et abusive des événements en jeu par le son, l'équipe audio d'Overwatch parvient à inciter le joueur à se référer à son audition, plutôt qu'uniquement aux feedbacks visuels qui lui sont fournis. Une telle conception vient alors supporter l'état de flow, gardant le joueur plongé et engagé dans un sentiment de vigilance permanent sans le fatiguer.

6-B - La question de la latence numérique, séparant l'input de sa réponse sonore

Comme nous avons pu le voir dans la partie sur la relation entre la création sonore et l'apparition du game feel ainsi que du flow, il est possible d'introduire une latence perceptible relative à l'espace laissé entre le début d'un fichier sonore et le pic de sa transitoire. Or, la création sonore n'est pas le seul élément de la chaîne pouvant introduire ce phénomène de latence: le code, ou encore l'interaction du moteur audio avec le moteur de jeu est susceptible de créer une latence numérique, venant affecter de façon négative l'expérience du joueur, et s'inscrire dans la liste des obstacles au maintien de l'état de flow. L'on peut définir la latence numérique comme suit: "Le délai nécessaire à la conversion d'un signal analogique vers sa représentation digitale". En fonction de diverses situations (l'implémentation, le support de jeu, le code, etc.), le moteur de jeu peut mettre un certain temps à envoyer une information au moteur sonore, qui à son tour peut mettre un certain temps afin de répondre au moteur de jeu. Également, une problématique plus commune peut être le temps nécessaire au moteur sonore afin de calibrer le déclenchement d'un événement par rapport à l'image, ce qui est souvent le cas lorsque le moteur sonore fonctionne de manière asynchrone avec le moteur vidéo (c'est à dire, indépendamment du nombre d'images par secondes du joueur). Cela vient alors entraîner une désynchronisation entre le moteur vidéo et le moteur sonore, qui opèrent chacun sur des "thread" différents. Si dans certains cas, la latence et donc la désynchronisation introduite peut être imperceptible, dans d'autres, elle peut nuire à l'instantanéité des feedbacks, et donc, inévitablement, à la sensation de contrôle du joueur vis-à-vis de son expérience.

Pour contourner ce problème, Jakob Schmid, le concepteur et programmeur audio sur le jeu "140" (mais aussi précédemment sur le célèbre INSIDE), parvient à créer

une synchronisation parfaite entre la conception sonore et les animations, en lançant ces dernières directement via la partie du code relative au son. Aussi, toutes les boucles musicales et éléments sonores sont déclenchés au même moment, c'est-à-dire au lancement de la scène, pour ne pas avoir à devoir re-déclencher ces boucles et ainsi causer une désynchronisation entre le visuel et l'audio. Les fade-ins et fade-outs sont alors effectués au travers du code, ce qui permet d'occulter toute notion de latence relative au déclenchement d'un événement sonore, et ainsi maximiser le potentiel de flow de l'expérience au travers d'une réactivité instantanée.

6-C - L'ignorance des phénomènes psychoacoustiques

Comme nous avons pu le voir dans notre partie sur la localisation, les phénomènes psychoacoustiques constituent une part habituellement sous-estimée, mais pourtant bien réelle, de la façon dont le joueur perçoit et ressent les sons constituant son environnement. L'ignorance de ces phénomènes lors du processus de conception sonore peut conduire à égarer l'attention du joueur vis-à-vis d'un événement ou d'une action en train d'être effectuée, et ainsi interrompre l'état de flow du joueur. Passons en revue quelques-uns de ces phénomènes.

6-C - Partie 1: Le phénomène de bourdon

Ce phénomène est caractérisé par la présence d'une strate au sein d'un ensemble sonore, strate de hauteur stable et sans variation notable d'intensité. Liée à la musique dans sa désignation, (le bourdon est un son permanent grave sur lequel reposent certains morceaux), il s'observe également dans les paysages sonores industriels et urbains. De nombreux systèmes techniques engendrent des constances sonores qui se rapprochent de cet effet (ventilations mécaniques, tubes fluorescents, ronronnement des moteurs, canalisations...), même si les fréquences concernées ne se limitent pas aux sons graves qui le caractérisaient à l'origine. Dans une scène de jeu vidéo, cet effet se manifeste la plupart du temps par la présence de bruits d'ambiance de pièce, ou "room tones" (généralement constitués d'un ensemble de bruits prenant leurs sources dans les exemples mentionnés plus haut). La représentation trop intense de ce type de sons dans le mixage sonore, ou une exposition prolongée à ces derniers, peut rapidement entraîner une fatigue auditive et mentale susceptible de mettre un terme à l'état de flow.

6-C - Partie 2: Le phénomène d'anticipation

Dans l'attente d'une situation à venir, une personne "pré-entend" - c'est-à-dire croit entendre effectivement - le signal attendu, alors qu'aucun son n'a encore été émis. Cet effet s'observe aussi bien dans des situations d'attente de sons inconnus, où le moindre bruissement devient indice, que dans des circonstances familières où l'auditeur anticipe un contexte sonore prévisible (pré-audible) dans sa mémoire. C'est notamment avec ce phénomène que joue la conception sonore de Silent Hill 2, afin de maintenir le joueur dans une vigilance et une tension continue. Cependant, si ce phénomène peut participer au maintien de l'état de flow dans l'expérience, il peut aussi venir l'interrompre: si un personnage est en mouvement et que le joueur s'attend à entendre ses bruits de pas, mais que ces derniers n'apparaissent pas dans son environnement sonore (par exemple, si le personnage est trop loin du joueur et que les paramètres d'atténuation de ses bruits de pas rendent le volume de la source trop bas), un phénomène similaire à celui de "l'uncanny valley" se produira, pouvant alors venir distraire et rompre l'immersion du joueur, entraînant dans ce cas une perte de l'état de flow.

6-C - Partie 3: Le phénomène d'irruption

Le phénomène d'irruption caractérise un événement sonore imprévu, venant modifier le climat du moment et le comportement de l'auditeur de manière caractérisée. Cet effet est au temps ce que l'effet d'intrusion est à l'espace. Par exemple, malgré la généralisation de son usage, la sonnerie du téléphone demeure pour beaucoup de personnes un événement sonore agressif, moins par son timbre qui a su s'adoucir, que par son caractère imprévu et impérieux: non seulement un appel interrompt l'état présent, mais il dicte un nouveau comportement pendant un temps donné. Ce phénomène survient notamment lors de l'exploration de l'espace de jeu de certains mondes ouverts, où le joueur, apaisé par les sons et les ambiances de la nature, se retrouve extirpé de son état de plénitude par le bruit d'une patrouille d'ennemis passant près de lui à toute vitesse.

6-C - Partie 4: Le phénomène de rémanence

Ce phénomène est caractérisé par la sensation de perdurance d'un son qui n'est pourtant plus entendu. Après extinction de l'émission et de la propagation, le son donne l'impression d'être encore « dans l'oreille ». Ce phénomène intervient notamment lorsque le joueur passe d'un environnement sonore bruyant et riche à un environnement beaucoup plus silencieux. Si cette alternance peut être utilisée comme effet de style, elle peut aussi contribuer à amplifier le souvenir de l'intensité

de l'environnement précédent, ce qui peut contribuer à provoquer un effet de rupture, venant alors interrompre l'état de flow du joueur. Le phénomène de rémanence intervient notamment lors d'un temps de chargement qui se succède à une scène d'action, où la différence de dynamique ou d'intensité en volume crée une sensation psychologique désagréable chez le joueur.

6-C - Partie 5: Le phénomène d'anamnèse

Aussi appelé "effet de réminiscence", le phénomène d'anamnèse caractérise un signal ou un contexte sonore provoquant chez l'auditeur le retour à la conscience d'une situation ou d'une atmosphère passée. Effet de sens, il caractérise le déclenchement, le plus souvent involontaire, de la mémoire par l'écoute, et le pouvoir d'évocation des sons. Ce phénomène est souvent créé de façon volontaire par les concepteurs sonores, notamment via l'utilisation de certains sons dans les jeux d'horreur comme celui de la boîte à musique, qui viennent évoquer au joueur l'enfance, et la vulnérabilité ressentie durant cette période. Dans ce cas présent, un effet de surprise survient généralement peu après, afin de brusquer le joueur une fois sa garde baissée. A l'instar du phénomène d'anticipation, le phénomène d'anamnèse peut servir d'outil comme apparaître en tant que nuisance: par exemple, si le joueur se retrouve immergé dans un ensemble sonore, par exemple une ville, où certains sons viennent le distraire de l'action en cours, alors l'état de flow est susceptible d'être brisé. Afin d'éviter ce phénomène, il est donc important d'établir une hiérarchie dans le mixage dynamique des sources (comme celui d'Overwatch), afin de pouvoir plus ou moins contrôler et maintenir le joueur dans le fil de l'action. Aussi, si l'utilisation de certains sons peut servir à établir une référence souhaitée (comme par exemple via le fameux "cri de Wilhelm"), ce procédé a plus tendance à desservir l'expérience, plutôt qu'à la renforcer.

6-D - Le concept de score de nuisance et de l'unité de mesure "NUFS"

Selon Bjorn Jacobsen, le "score de nuisance" caractérise une mesure relative au niveau de désagrément produit par un ou plusieurs sons. Chacun de ces sons ou ensemble de sons sont alors labellisés en fonction de leur score de nuisance et leur potentiel d'occurrence, afin de définir lesquels de ces éléments doivent se démarquer des autres au sein du mixage. Jacobsen décrit notamment un cas de figure, où une séquence de bruits de pas constituée de 5 fichiers sonores, chacun

altérés en pitch et en volume de façon aléatoire pour chaque répétition et procurant donc une infinité théorique de variations, venait se démarquer du reste de l'environnement sonore de façon nuisible. Cela était dû à un des fichiers de bruits de pas de la séquence, qui sonnait "de façon trop différente" par rapport aux autres sons compris dans le cycle. Avec le temps, l'attention du joueur finissait par se porter de plus en plus sur cet élément trop singulier, résultant en un schéma de répétition aussi prédictible qu'imprédictible, et venant alors créer une importante nuisance sonore dans l'expérience du joueur. Ce faisant, Jacobsen attribua un score de nuisance de 8 aux bruits de pas, ainsi qu'aux sons de portes et d'explosions. Alors, ces événements sonores venaient à être considérés comme "bugs sonores brisant l'immersion". Un tel score de nuisance pouvait également être conséquent selon certains espaces de jeu ou niveaux, si les harmoniques des ambiances sonores venait entrer en dissonance par rapport au contenu harmonique de la musique jouée au même moment.

Le concept de "NUFS" (Nuisance Unit Full Scale), théorisé dans l'ouvrage "Tuning of a Sound" de Bjorn Jacobsen, consiste donc à représenter le "score de nuisance" d'une conception sonore relatif à une certaine durée. Ce procédé, bien que relativement subjectif, revient à mesurer le score de nuisance des différentes sections du jeu et les catégoriser par une mesure en NUFS, afin de pouvoir plus facilement cibler les diverses problématiques liées à la conception sonore, et orienter la création de systèmes en fonction.

6-E - Les problématiques de décorrélation entre un événement sonore et son intention

Le phénomène de décorrélation entre une action et l'événement sonore lui étant associé est relativement courant. Par exemple, si l'impact sonore d'un coup d'épée est inversement proportionnel à celui suggéré par l'animation ou la situation en jeu, cela peut venir créer une forme de dissonance cognitive. Cela peut notamment se produire lorsque le concepteur sonore se contente de créer et livrer un fichier sonore, sans l'avoir implémenté et testé par lui-même auparavant. Ce type de négligence peut amener bon nombre d'incohérences dans la conception sonore d'un jeu, devenant alors nuisible à l'apparition et au maintien de l'état de flow. Comme le dit Paul Weir: "Je conçois que selon l'échelle des productions et l'organisation de leur hiérarchie, il est parfois compliqué de pouvoir implémenter et vérifier soi même son travail. Cependant, il y a toujours la possibilité de demander des vérifications journalières, afin de pouvoir prévenir toute négligence au niveau de la conception des sons ou des systèmes qui les agencent".

Il arrive également que, suivant le niveau de répétition ou le faible niveau de variété d'un son, ce dernier viennent perdre en importance et en impact dans l'expérience, voire se montrer frustrant et venant alors à l'encontre de l'intention de départ. Il est alors préférable de prioriser une approche qualitative à une approche quantitative: plus l'agencement des systèmes et le niveau de variété perceptible est élevé, moins il est nécessaire d'ajouter du contenu afin de contourner la redondance d'un même son.

7. Sans l'image, le son suffit-il à atteindre et maintenir l'état de Flow?

Selon Masami Komuro, "il est plus facile d'amener l'état de flow dans un jeu uniquement composé d'éléments et d'interactions sonore (par exemple répondant à la voix du joueur), plutôt que d'amener l'état de flow dans un jeu où tout aspect sonore est absent." L'augmentation de la qualité du son dans le jeu vidéo au fil des générations et des avancées technologiques a conduit à un désir de maîtriser la conception sonore de ce média - notamment via la musique interactive -, de manière à pouvoir plus facilement introduire cet état de flow, ou du moins les composantes qui le constituent (la concentration intense, l'engagement, l'altération temporelle, etc.). Déjà, au travers de la conception sonore du jeu de stratégie "Z", sorti en 1996 par le studio The Bitmap Brothers, les développeurs commençait à expérimenter la notion de musique interactive en tant que moyen de communication, afin de transmettre au joueur de manière plus fluide et instinctive les informations nécessaires à la compréhension de l'action (et ainsi faire apparaître et maintenir l'état de flow), notamment par la lecture séquentielle de segments MIDI suivant que le joueur construise de nouveaux bâtiments, de nouvelles troupes, etc.". La question suivante peut alors se poser: sans feedbacks visuels, la conception sonore peut-elle suffire à l'atteinte et au maintien de l'état de flow?

Si cette interrogation peut paraître secondaire aux yeux d'un individu en pleine possession de ses capacités visuelles, il en est tout autrement pour les personnes malvoyantes et aveugles. On ne peut alors que saluer une approche comme celle de l'équipe audio d'Overwatch, qui vient encourager le fait de jouer en s'aidant presque uniquement des informations sonores véhiculées par le jeu. Une vidéo sur la chaîne YouTube "Vice News" raconte le parcours de joueur de Michael Espinoza, joueur aveugle depuis sa naissance, et comment ce dernier arrive à jouer à des titres grand publics tels que Mortal Kombat, pourtant non destinés ou orientés autour de

l'accessibilité aux personnes malvoyantes et aveugles. Si le travail d'apprentissage et de mémorisation des patterns de coup, des temps de délai entre ces derniers, ainsi que l'identification des coups par rapport à leur réponse sonore est naturellement plus laborieux que celui effectué par un joueur en pleine possession de ses capacités visuelles, ces jeux sont néanmoins rendus accessibles grâce à leur conception sonore. On peut donc dire que sans l'image, le son peut suffire à faire atteindre l'état de flow à un joueur, même si la symbiose de l'expérience visuelle et sonore est toujours plus efficace pour un joueur lambda. Une anecdote amusante de Christophe Héral vient confirmer cela: "Lors d'une conférence donnée à une édition du Toulouse Game Show, et portant sur la musique de jeu vidéo, un homme s'était plaint du fait que dans Rayman Legends, le son ne venait pas renseigner le fait que l'avatar du joueur se déplaçait contre un mur. Si cette remarque m'avait au début surpris, je me rendis compte par la suite que l'homme en question était aveugle.", dit-il. "C'est intéressant, parce que via les feedbacks sonores procurés par le jeu, comme les bruits de la lave ou d'un danger qui s'amplifient à mesure que l'on s'en approche, même un aveugle peut jouer à Rayman Legends! Cela se confirme d'autant plus avec les niveaux musicaux du jeu, où via la mémorisation de patterns et surtout la concentration sur la musique, un joueur peut s'aider uniquement de son audition pour parcourir un niveau.". En effet, les niveaux musicaux de Rayman Legends, pensés comme des clips musicaux interactifs par Christophe Héral et l'implémentateur son Adrien Pavageau, consistent en une épreuve de jeu de plateforme musical où chaque saut ou action devant être effectuée est annoncé par la construction du morceau, ainsi que soulignés par des feedbacks sonores synchronisés au tempo et à la structure musicale. Il en résulte alors une symbiose image son grisante, instantanément porteuse de l'état de flow, qui reste malgré tout jouable par un public malvoyant.

Pour finir, si certains jeux permettent de manière plus ou moins imprévue d'être expérimentés et d'apporter un état de flow uniquement par le biais du son, d'autres, comme le titre indépendant "A Blind Legend" sorti sur ordinateurs et appareils mobiles en Juin 2015, sont directement pensés pour pouvoir être vécu par des personnes malvoyantes. Malin, le jeu s'articule autour de contrôles simples, à savoir un glissement vers le haut, le bas, la droite ou la gauche de l'écran, permettant au joueur d'avoir un niveau d'interactivité tactile avec l'action. Pour le reste, pas de visuels: tout se passe au niveau du son en binaural, permettant de s'orienter et se déplacer dans l'espace via les feedbacks sonores. Si une expérience telle que celle-ci est évidemment idéale pour une personne aveugle, elle est d'autant plus intéressante pour un joueur à la vision intacte, qui doit alors apprendre une nouvelle manière de jouer, en se connectant de manière plus profonde avec son audition. Alors, les images deviennent mentales, puisant dans l'imaginaire du joueur, rendant ainsi l'expérience unique pour chaque profil. Cette reconnection aux sens primaires de l'être humain permet au joueur de s'immerger bien plus efficacement que si les informations visuelles lui étaient offertes directement, venant alors maintenir l'intérêt et l'engagement du joueur au profit du maintien de son état de flow.

Conclusion

Au cours de ce mémoire sur la relation entre le son et l'apparition ainsi que le maintien de l'état de flow dans l'expérience vidéoludique, nous avons pu comprendre et définir l'état de flow dans le contexte d'un jeu vidéo, explorer les différentes manières de créer et agencer la conception sonore afin de faire apparaître cet état, et passer en revue une multitude de concepts, méthodes et systèmes d'implémentation permettant de le maintenir. Au travers de l'expérience de professionnels de l'industrie du jeu vidéo, ainsi que de déductions basées sur des études scientifiques, nous avons pu comprendre que, si il n'existe pas une réponse applicable à tous les différents types d'expériences vidéoludique en ce qui concerne l'apparition et le maintien de l'état de flow, il existe en revanche plusieurs pistes ou axes d'amélioration qui se rejoignent dans la plupart des conceptions sonores, tout du moins celles réalisées de manière pertinente et réfléchie. On remarque donc que la compréhension, l'altération et la gestion des différentes caractéristiques propres à un son, ainsi que l'élaboration d'une conception sonore décomposée en un ensemble de systèmes flexibles et interactifs, permettent de faciliter l'apparition et le maintien du flow, et ce, même en venant écarter la question du rapport audiovisuel. Si l'élaboration et l'utilisation de tous ces concepts n'est pas chose aisée, et comporte parfois des problématiques budgétaires, elle est en revanche cruciale: c'est elle qui permet de maximiser le potentiel de plaisir et d'engagement du joueur vis-à-vis de son expérience, afin de donner la possibilité aux concepteurs de jeux de véhiculer correctement le ou les messages souhaités (et ce, qu'ils soit émotionnels, personnels ou même politiques).

Je profite de ce dernier paragraphe pour remercier, sans les nommer (la liste était bien trop longue), toutes les personnes m'ayant aidée, d'une manière ou d'une autre, à la conception et la rédaction de ce travail, et sans qui le résultat n'aurait pas été le même.

Enfin, je remercie d'avance le(s) lecteur(s) ou la(les) lectrice(s), qui ont su s'intéresser à ce sujet et tenter de le comprendre à mes côtés.

Amicalement,

Gabriel Sauvage

Bibliographie

“Flow: The Psychology of Optimal Experience” - Mihaly Csikszentmihalyi

“Flow in Games” - Jenova Chen

“The Tuning of the Game” - Bjorn Jacobsen

“Towards an Understanding of Flow in Video Games” - Ben Cowley

“Game Feel” - Steve Swink

Ludographie

- “Doom Eternal” - 2020, Id Software*
- “Manifold Garden” - 2019, William Chyr Studio*
- “Inside” - 2016, Playdead Games*
- “Limbo” - 2010, Playdead Games*
- “140” - 2013, Abstraction Games*
- “Rayman Legends” - 2013, Ubisoft*
- “Get Even” - 2017, The Farm 51*
- “Remember Me” - 2013, Dontnod Entertainment*
- “Silent Hill 2” - 2001, Konami Computer Entertainment Tokyo*
- “No Man’s Sky” - 2016, Hello Games*
- “Grand Theft Auto 5” - 2013, Rockstar Games*
- “Overwatch” - 2016, Blizzard Entertainment*
- “Z” - 1996, The Bitmap Brothers*
- “The Legend of Zelda: Breath of the Wild” - 2017, Nintendo*
- “Thumper” - 2016, Drool*
- “The Secret of Monkey Island” - 1990, Lucasfilm Games*
- “A Blind Legend” - 2015, Dowino*
- “DARQ” - 2019, Unfold Games*
- “Hellblade” - 2017, Ninja Theory*
- “Bioshock: Infinite” - 2013, Irrational Games*

Le rapport entre le son et l'état de Flow dans l'expérience vidéoludique

Résumé

Ce mémoire se concentre sur le rapport entre le son et l'état de Flow au sein d'une expérience vidéoludique. J'y décris, en plusieurs parties et sous parties, le concept de l'état "flow". Puis, j'explore la manière dont les concepteurs sonores contribuent à recréer cet état de flow dans l'expérience vidéoludique, au travers de sa direction sonore. Tout d'abord, je réalise une introduction au flow et entreprend de définir cet état, en le distinguant de l'état d'immersion. Suite à cela, j'explique en quoi la maîtrise des caractéristiques sonores constituent le pilier de l'apparition de l'état de flow, et comment l'intégration de la conception sonore dans le processus de game design, ainsi que son implémentation via l'élaboration de systèmes flexibles et interactifs, permet de véhiculer cet état au joueur.

Mots clés: Flow, Jeu Vidéo, Expérience Vidéoludique, Sound Design, Composition, Conception Sonore, Interactivité, Interactivité Sonore, Game Feel, Wwise