

Université de Montréal

Du son et de la musique à la concrétion sonore : structuration audio du conflit en
jeu vidéo

Par
Ambrune Martin

Département d'histoire de l'art et d'études cinématographiques,
Faculté des Arts et des Sciences

Mémoire présenté en vue de l'obtention du grade de Maîtrise ès Arts (M.A.)
en Maîtrise en cinéma, option études du jeu vidéo

Août 2020

© Ambrune Martin, 2020

Université de Montréal

Unité académique : département d'histoire de l'art et d'études cinématographiques,

Faculté des Arts et des Sciences

Ce mémoire intitulé

Du son et de la musique à la concrétion sonore : structuration audio du conflit en jeu vidéo

Présenté par

Ambrune Martin

A été évalué par un jury composé des personnes suivantes

Bernard Perron

Président-rapporteur

Dominic Arsenault

Directeur de recherche

Irina Kirchberg

Membre du jury

Résumé

Ce mémoire emprunte à la ludomusicologie tout en s'inscrivant dans le cadre des études du jeu vidéo et des études cinématographiques. Il est une exploration du sonore à l'intérieur d'une situation de conflit présente dans un jeu vidéo. Nous y appréhendons la construction sonore de la situation de conflit, caractérisée par un combat entre un personnage jouable et un antagoniste. Nous posons les problématiques suivantes : de quelle manière le sonore se construit-il dans une situation de conflit vidéoludique? Qu'est-ce que cette construction apporte au joueur ou à la joueuse?

En prenant appui sur le cinéma puis le cinéma d'animation, nous envisageons le son dans sa construction audio-visuelle, rattaché à une image mouvante. Il apparaît rapidement que le son en rapport à l'image possède des fonctions qui permettent à l'auditeur-spectateur d'appréhender, voire d'amplifier à la fois l'histoire et les émotions.

À travers le cognitivisme, nous transposons ensuite ces fonctions vers le jeu vidéo, en prenant en compte le paradigme de la jouabilité. En nous intéressant au joueur, nous partons de ses attitudes d'écoute pour proposer la construction d'une concrétion sonore, un ensemble de sons et de musiques inhérents à toute situation audio-visuelle. Il ne s'agit plus de séparer sons, musique et bruits, mais bien de les hiérarchiser en fonction des besoins du joueur, qui sont déterminés par la situation de jeu. Une étude de cas nous permet d'explicitier la concrétion sonore.

Nous précisons enfin les enjeux relatifs à la concrétion sonore dans le cadre de la situation de conflit. Il apparaît au fil de nos recherches que le son, les bruits et la musique possèdent des fonctions qui dépendent moins de leur structure que de l'assemblage sonore complet. De nombreux exemples présenteront des strates sonores aux multiples fonctions. La musique comme le son peuvent constituer aussi bien une aide qu'un frein à la jouabilité. Des techniques de production (volume, spatialisation, complexité du sonore) permettent au joueur de hiérarchiser la concrétion sonore pour profiter d'une meilleure expérience audio-visuelle.

Mots-clés : jeu vidéo ; combat ; conception sonore ; écoute ; audio-visuel ; fonctions sonores ; antagoniste

Abstract

This dissertation is the result of research in ludomusicology, game studies and cinema studies. It is an exploration of sound in the context of a conflict situation present in a video game. We apprehend the sound construction of the conflict situation, characterized by a fight between a playable character and an antagonist. We propose the following questions: how is sound constructed in a video game conflict situation? What does this construction bring to the player?

By studying the cinema and then animation, we consider sound in its audio-visual structure, linked to a moving image. It quickly appears that the sound in relation to the image has functions that allow the listener-spectator to apprehend, even amplify both the story and the emotions.

Through cognitivism, we then transpose these functions to video games, taking into account the paradigm of playability. By focusing on the player, we start from his listening attitudes to suggest the construction of a sound concretion, a set of sounds and music inherent to any audio-visual situation. It is no longer a question of separating sounds, music and noises, but rather of prioritizing them according to the player's needs, which are determined by the gaming situation. A case study allows us to explain the sound concretion.

Finally, we specify the stakes related to sound concretion in the context of the conflict situation. It appears through our research that sound, noises and music have functions that depend less on their structure than on the entire sound assembly. Many examples will present sound layers with multiple functions. Music and sound can be both an aid and a hindrance to playability. Production techniques (volume, spatialization, sound complexity) allow the player to prioritize the sound concretion to enjoy a better audio-visual experience.

Keywords: video game; fighting; sound design; listening; audio-visual; sound functions; antagonist

Table des Matières

Résumé.....	5
Abstract.....	7
Table des Matières.....	9
Liste des figures.....	11
Liste des sigles et abréviations.....	13
Remerciements.....	17
Introduction.....	1
Chapitre 1 – Le son audiovisuel.....	11
1.1 Caractéristiques du son au cinéma.....	13
1.1.1 Audio-visuel et valeur ajoutée.....	14
1.1.2 Sources sonores et acousmatique.....	16
1.1.3 Stratégies d’écoute au cinéma.....	17
1.2 Fonctions du son au cinéma.....	19
1.2.1 Le Congruence-Association Model.....	19
1.2.2 Les fonctions audiovisuelles.....	22
1.3 Cinéma d’animation, image construite et son superposé.....	27
1.3.1 Les fondements du son synchronisé.....	27
1.3.2 Jeu vidéo et cinéma d’animation.....	31
Chapitre 2 – Jeu vidéo et concrétion sonore.....	35
2.1 Spécificités du jeu vidéo.....	36
2.1.1 Jouabilité et cycle magique.....	36
2.1.2 La hiérarchie des attitudes d’écoute.....	39
2.1.3 Le classement de Collins.....	44
2.2 Appréhender le sonore.....	49
2.2.1 La concrétion sonore.....	49
2.2.2 Mise en situation de la hiérarchisation sonore : une étude de cas avec <i>Assassin’s Creed : Brotherhood</i>	53
Chapitre 3 – La concrétion sonore dans la situation de conflit.....	60
3.1 Implémentation de la situation de conflit.....	60
3.1.1 Les transitions sonores et la continuité de l’expérience.....	61

3.1.2 La position du personnage jouable.....	69
3.1.3 L'information donnée : le joueur attentif.....	74
3.2 Combat et concrétion sonore.....	78
3.2.1 Relier l'antagoniste à son environnement.....	79
3.2.2 Alertes sonores et réflexes	82
3.2.3 Indices musicaux et rythmes d'attaque.....	88
Conclusion	93
Références bibliographiques.....	101
Ludographie.....	103

Liste des figures

Figure 1. Congruence-Association Model – Iteration 3. © Cohen <i>et al.</i> , <i>The Psychology of Music in Multimedia</i> , 2013, 33 (issu de l'information audiovisuelle disponible en ligne).	22
Figure 2. Quelques images du film d'Heider et Simmel, 1944. ©Evelyne Thommen 2011.	28
Figure 3. <i>Schéma du cycle heuristique de jouabilité</i> , 2009. ©Arsenault et Perron	38
Figure 4. Mème populaire dont l'image est issue du film © <i>Le Seigneur des Anneaux : la Communauté de l'Anneau</i> (2001)	76

Liste des sigles et abréviations

2D : Deux Dimensions.

3D : Trois Dimensions.

AoE : *Area of Effects*, attaques dont le champ d'effet concerne une aire et non une seule cible.

CAM : *Congruence-Association Model*, modèle cognitif développé par Annabel J. Cohen.

DPS : Dégâts Par Seconde. Est aussi employé pour qualifier un personnage dont le rôle est d'infliger des dégâts.

FPS : *First-Person Shooter*, jeu de tir en vue à la première personne.

GTFO : *Get The Fuck Out*, addon populaire du jeu World of Warcraft.

MMORPG : *Massively Multiplayer Online Role-Playing Game*, jeu de rôle en ligne massivement multijoueur.

PNJ : Personnage Non Jouable.

PvE : *Player versus Environment*, Joueur contre Environnement. Mode de jeu où l'antagoniste est représenté par un ou des personnage(s) non jouable(s).

PvP : *Player versus Player*, Joueur contre Joueur. Mode de jeu où l'antagoniste est représenté par un autre joueur.

RPG : *Role-Playing Game*, jeu de rôle (s'emploie dans le cadre du jeu vidéo).

SMB : *Super Mario Bros.* (Nintendo, 1985).

WoW : *World of Warcraft* (Blizzard Entertainment, 2004).

Sommet venteux et enneigé

Soulagement euphorique de celle qui termine

Ascension

Remerciements

Ce chemin aux mille courbures n'aurait pu commencer sans le soutien de mon directeur de recherche, Dominic Arsenault. L'enthousiasme qui se lisait sur ton visage m'a permis de me lancer, d'abord sur une piste bien éloignée du sujet de ce mémoire. À l'issue d'une conversation qui a remué des souvenirs et des rêves enterrés depuis longtemps, tu m'as proposé de rejoindre à nouveau le monde de la musique, que je croyais avoir quitté. Pour ce rebond inespéré, ton aide, ta bienveillance et tes sourires amusés où tu attendais patiemment, en silence, le bout de mon idée et la fin de mes phrases brouillonnes, merci.

Je remercie ensuite mes parents, Laurence et Christophe, pour m'avoir permis de tracer mon chemin jusqu'ici. Si j'ai parfois navigué à vue, si je me suis perdue en mer, j'ai toujours pu compter sur votre force et votre confiance. Vous m'avez offert tous les outils dont je pouvais rêver, vous m'avez appris à manœuvrer sur une mer d'huile ou par tempête. Vous avez accepté de me laisser voguer en solitaire, à la seule condition d'emporter un téléphone et de partager avec vous mes découvertes. Je mesure chaque jour l'étendue de ma chance. Ce mémoire vous est dédié.

Théophile, merci pour tout. Aucune phrase n'est assez précise ni assez belle pour expliquer ce que tu apportes dans ma vie. Pour tes innombrables encouragements, les promenades pour me vider la tête ou au contraire réfléchir plus intensément, pour les soirées jeux, les conseils de lecture et les mille couleurs dont s'est parée ma route, merci infiniment.

J'adresse enfin mes remerciements à ceux qui ont traversé le chemin de ce mémoire et m'ont toujours apporté de nouvelles pistes, qu'ils en aient conscience ou non. À Mutti, Mamyse, Mimi, Vati, Papy's et Grand-Père pour la sérénité dont vous faisiez (ou dont vous faites toujours!) preuve à chaque nouveau rebondissement de parcours. À Alisée et Malo, qui tracent leur propre sentier avec brio sans jamais s'éloigner de notre route commune. À tous les enseignants qui m'ont offert ce que l'école ne pouvait pas m'apporter. À Mr Revol, Mme Roux et Mme Heberard qui m'ont aidée plus encore qu'ils ne peuvent l'imaginer. À Philippe pour ses précieux conseils. À Kevin et Mathieu sans qui Montréal aurait été bien différente.

À Adrien « ZeratoR » Nougaret qui m'a fait découvrir à quel point les jeux vidéo étaient riches et variés, grâce à qui je me suis lancée à la découverte d'un monde que je connaissais si peu. À Xavier « mistermv » Dang qui, plus tard, m'a permis de relier le jeu vidéo à la musique à travers une culture incroyable et une imagination débordante. Chacun à votre manière, vous m'avez appris que la musique prenait la forme qu'on voulait lui donner, une forme née du plaisir et non de la contrainte.

À tous les auteurs, concepteurs, compositeurs, inventeurs de mondes qui ont fait la personne que je suis aujourd'hui, en m'offrant des routes illuminées par les rêves de ceux qui créent.

Introduction

La recherche en jeu vidéo est un phénomène récent. Considéré dans un premier temps, à l'instar de tous les nouveaux médias, comme un apport culturel moindre en raison de son aspect commercial et de son public de destination, le jeu vidéo a toutefois intéressé les chercheurs et chercheuses qui questionnaient plusieurs sujets de recherche annexes : le concept du jeu, bien entendu, mais aussi ceux de la création informatique, de la simulation ou encore de la narration. Avec l'expansion des études du jeu vidéo, ce dernier a finalement fait son entrée dans le monde académique, comme avant lui le cinéma, le roman ou la musique non savante. L'aspect populaire d'un phénomène culturel n'est désormais plus un frein pour son étude, et nous sommes heureux de nous inscrire dans un champ de recherche en pleine évolution, dont la légitimité n'est plus vraiment questionnée dans l'environnement universitaire nord-américain¹. Le jeu vidéo en tant qu'objet médiatique, culturel, artistique et populaire peut se réjouir d'être le sujet d'une attention grandissante.

Avec l'étude du jeu vidéo sont nées de nouvelles disciplines de recherche, dont les spécificités sont aussi multiples qu'il y a de chercheurs et de chercheuses pour s'y intéresser. Ce mémoire, qui emprunte à plusieurs disciplines, s'intéresse à des problématiques typiques de la ludomusicologie. Étymologiquement construite autour du latin *ludus, i, m* qui signifie « le jeu, ce qui amuse », et du latin *musica, ae, f* qui se rapporte à « musique », la ludomusicologie est tout d'abord l'étude de la musique de jeu vidéo. Dans un sens élargi, ce champ de recherche s'intéresse à la construction sonore présente dans un jeu vidéo, depuis sa conception par une équipe de créateurs sonores ou par un développeur habile, jusqu'à sa

¹ En Europe toutefois, cette légitimité est encore en train de se construire, grâce à des projets de recherche de plus en plus nombreux.

perception par le joueur ou la joueuse une fois que le jeu, proposé comme objet culturel de divertissement, est mis à disposition d'un public. Si les approches ludomusicologiques sont, comme nous le verrons tout au long de cette recherche, extrêmement variées, toutes semblent partager un point commun : réfléchir à l'importante implication du sonore dans un média audiovisuel où nos perceptions tendent à placer naturellement le visuel au premier plan.

Ce mémoire ne s'inscrit toutefois que partiellement en ludomusicologie, se situant plutôt au carrefour des champs que sont les études du jeu vidéo, les études cinématographiques, la narratologie et la ludomusicologie. Nous souhaitons nous concentrer tout autant sur la musique que sur les bruitages ou les sons d'interface, en mettant en avant la perception globale du joueur ou de la joueuse lors de l'expérience vidéoludique. Notre recherche est partie d'une intuition, qui prendra par la suite la forme d'une hypothèse : lorsque nous jouons, nous sommes attentifs, par différents degrés, à tout ce qui constitue l'expérience de jeu. Dans un jeu vidéo, nous suivons du regard les informations visuelles importantes, nous prenons conscience de la position de notre personnage jouable dans l'environnement virtuel qui nous est proposé tout en tenant compte des alertes données par l'interface, qui améliorent ou détériorent la jouabilité dans laquelle nous nous inscrivons en décidant volontairement de jouer. Mais surtout, nous entendons. Nous employons un dispositif d'écoute, perméable ou non à notre environnement réel, transmettant des informations de jouabilité qui sont les mêmes ou qui diffèrent de celles présentes sur l'écran. Nous entendons les pas de notre personnage jouable, le halètement de sa course, lorsque soudain un cri déchire la musique tranquille qui défilait dans l'arrière-plan sonore. Ce cri, nous le savons déjà, annonce la présence d'un antagoniste et l'imminence d'une situation de conflit à laquelle nous devrons, dans la grande majorité des cas, nous confronter. Lors de notre

expérience de jeu, le sonore n'est pas décomposé en fragments détachés les uns des autres. Plusieurs sources sonores s'entremêlent, se croisent, fonctionnent en lien pour former un ensemble que nous entendons, où nous sommes en mesure de repérer certaines informations de jeu. Dans ce mémoire, nous allons nous pencher sur tout ce qui constitue l'expérience sonore vidéoludique.

Mais l'ambition de ces premières intuitions dépasse largement le cadre d'un mémoire de maîtrise. L'exhaustivité est un idéal auquel la jeune chercheuse doit renoncer, pour proposer une recherche claire, efficace et qui ne souffre pas de démesure. Pour cette raison, nous avons choisi de concentrer l'étude du sonore vidéoludique autour d'une seule situation de jeu : la situation de conflit.

Méthodologie et limites de la recherche

Dans un jeu vidéo, le joueur ou la joueuse expérimente de nombreuses situations, qui dépendent à la fois du genre du jeu et des choix de conception qui ont été faits. Nous pouvons passer par des situations d'exploration, d'infiltration, de stratégie, de construction, de récolte. Souvent, plusieurs situations se croisent à l'intérieur d'un seul jeu. Afin de circonscrire cette recherche sans mettre de côté l'étude du sonore dans son intégralité, nous nous concentrons donc sur la situation de conflit vidéoludique, que nous définissons ainsi : il s'agit de la rencontre, dans un environnement de jeu, entre le personnage jouable et un antagoniste, qui peut prendre la forme d'un personnage jouable par un autre joueur ou d'un personnage non jouable, issu de l'imagination des concepteurs du jeu². La situation de conflit implique ici de

² Cette définition est issue d'un travail préalable effectué sur les situations d'opposition en jeu vidéo. À l'aide de théoriciens du jeu tels que Salen et Zimmerman, Caillois, Juul, je me suis intéressée au conflit et aux antagonistes vidéoludiques. J'ai pu déterminer que le jeu vidéo, à l'instar du jeu, du sport ou d'une narration classique, se constitue autour d'une forme d'opposition : « All games involve a conflict, whether that conflict occurs directly between players or whether players work together against the challenging activity presented by the game system » (Salen et Zimmermann 2004, 255). De fait, le joueur intègre implicitement la situation de conflit comme situation de jeu. Dans le cadre de cette recherche, je fais le choix d'évoquer la situation de

manière minimale la présence d'un personnage jouable représenté à l'écran et d'un antagoniste perçu par le joueur comme un être menaçant. Le conflit peut être présent sous d'autres formes (conflit intérieur qui prend la forme d'un dilemme, attaque en groupe d'une base contrôlée par un autre joueur, etc.), mais pour des volontés de simplification de cette recherche nous centrerons celle-ci autour de ces deux considérations minimales. Les jeux vidéo ne représentant pas clairement un personnage jouable à l'écran, à l'instar de ceux où le joueur construit et dirige de nombreuses unités, ne seront malheureusement pas considérés.

Nous faisons le choix ici de ne pas centrer l'étude autour d'un genre de jeu spécifique car notre hypothèse dépasse les caractéristiques imputables à chaque genre, la construction sonore de la situation de conflit employant des tropes communs reconnaissables facilement par le joueur ou la joueuse selon son expérience avec d'autres jeux. Ainsi, nous explorerons l'idée selon laquelle la situation de conflit se déroule, dans sa perception sonore, d'une manière similaire quels que soient le genre et l'esthétique du jeu.

Nous souhaitons dans cette recherche poser les problématiques suivantes : de quelle manière le sonore se construit-il dans une situation de conflit vidéoludique? Qu'est-ce que cette construction apporte au joueur ou à la joueuse? Cette réflexion, partie des perceptions précédemment évoquées, s'est enrichie de nouvelles hypothèses au fil des lectures et des expérimentations racontées dans ce mémoire.

Il serait prétentieux de considérer ce mémoire comme exempt de biais. Si j'ai veillé à apporter tout au long de mon processus de recherche les préconisations instaurées par l'environnement académique, il paraît toutefois nécessaire, avant d'entrer dans le vif du sujet,

conflit dans sa définition la plus simplifiée, c'est-à-dire une opposition ferme entre joueur et antagoniste, qui ne se résoudra que par confrontation lors d'un combat.

de reconnaître les écueils auxquels je me suis heurtée. Les considérer n'enlève rien à la valeur du travail, me permettant au contraire, en admettant les failles, de chercher à les dépasser dans la suite de mes démarches intellectuelles.

Tout d'abord, constatons que cette recherche est rédigée dans le cadre d'une université nord-américaine. Si je prône l'universalité de la recherche académique, force est de constater que les sujets, les concepts et les théories s'ancrent dans leur propre milieu culturel. Ma bibliographie provient d'Europe ou d'Amérique du Nord, les auteurs cités sont en immense majorité caucasiens. Ces ouvrages ont été choisis en fonction de leur pertinence, mais aussi de leur disponibilité et de leur langue. Je crois fortement qu'un apport issu des mondes académiques asiatique ou africain aurait permis à ce mémoire d'entrevoir de nouvelles perspectives, et je souhaite à l'avenir que les milieux académiques fonctionnent avec une meilleure transmission des savoirs, sans domination ethnique.

Puisque je discute d'objets culturels, il est impératif de reconnaître le biais qui en découle, à la fois de goût et de connaissance. Les jeux vidéo sont innombrables, la production dépasse chaque année de nouveaux records. Il m'est impossible, dans ce cadre, d'approcher une quelconque exhaustivité dans mes connaissances et expériences vidéoludiques. S'ajoute à cette considération des choix personnels, qui ont été fait à la fois par pertinence mais aussi par goût. Employer un exemple plutôt qu'un autre, ça n'est pas reconnaître la supériorité de l'objet cité. C'est admettre que j'aurais pu en citer d'autres, mais que je ne l'ai pas fait, par méconnaissance ou par choix. Ma pratique du jeu vidéo a dicté cette recherche, et cela se ressent à travers les exemples. Ainsi, je ne parle pas des jeux vidéo d'horreur ou des jeux nippons parce que je les connais mal. Je m'engage à continuer d'explorer le monde

vidéoludique, pour revenir plus riche et plus ouverte aux expériences méconnues tout en évitant autant que possible la démonstration tautologique.

Je n'ai pas pu, pour des raisons de format et de nécessité, m'intéresser aux spécificités des expériences de jeu. En pratique, chaque joueur et chaque joueuse est unique, et son expérience personnelle diffèrera selon ses goûts, ses connaissances, ses compétences et le dispositif de jeu qu'il ou elle emploie. La théorie se base sur un joueur ou une joueuse idéal.e³, qui finalement n'existe pas vraiment. Je regrette de ne pas avoir exploré des sujets aussi importants que la diversité des dispositifs d'écoute (l'écoute étant conditionnée par la présence ou l'absence de bruits parasites, de stéréo, de perméabilité ou d'imperméabilité du dispositif par rapport aux sons hors du jeu, etc.), les paramètres sonores modifiables par l'utilisateur – généralement une séparation des volumes liés aux effets, aux musiques, aux voix – et les contraintes rencontrées par un joueur ou une joueuse porteur-se de handicap. Si la majorité des jeux vidéo sont construits aujourd'hui pour un public valide, de plus en plus de projets proposent des alternatives pour des joueurs ou des joueuses dont une perception est altérée. Dans le cadre du sonore, cela peut passer par une réduction ou une augmentation du nombre d'informations transmises par un signal audio plutôt que visuel. J'ai préféré laisser ces sujets de côté en raison des contraintes formelles de ce mémoire.

Enfin, il est impératif de préciser que cette recherche ne cherche pas à mettre en avant un genre, une ethnie ou une représentation quelle qu'elle soit. Le jeu vidéo est avant tout un objet culturel destiné à un public élargi. Les preuves de sexisme dans l'industrie et dans la pratique vidéoludique s'accumulent de plus en plus, et je ne cautionne ni n'accepte aucune prévalence du genre masculin sur les autres genres, à la fois dans ma vie et dans mes

³ Pour emprunter un concept maintes fois théorisé en littérature, à travers le « lecteur modèle » présenté par Umberto Eco dans son *Lector in fabula* (Eco, [1979] 1985).

recherches. Il a été démontré à de nombreuses reprises⁴ que les femmes jouent autant que les hommes à des jeux vidéo, et le manque de données sur les personnes s'identifiant à d'autres genres est problématique. L'utilisation sémantique dans ce mémoire d'un masculin générique est donc conditionnée uniquement par l'allègement syntaxique du texte, dans une langue qui autorise peu la neutralité du genre à l'écrit. Cet usage ne saurait refléter une opinion personnelle.

Je souhaite, pour terminer l'état des lieux des limites de ce mémoire, attirer l'attention du lecteur sur un sujet controversé. Ma rédaction est ancrée dans l'année 2020, qui a connu son lot de dénonciations dans le milieu de la production du jeu vidéo. Des mouvements surviennent, ponctuellement mais avec régularité, pour dénoncer de graves formes d'abus de pouvoir dans les chaînes de production vidéoludiques. Des producteurs, concepteurs, compositeurs, directeurs éditoriaux ou parfois même l'ensemble de l'exécutif d'un studio sont accusés de harcèlement moral et sexuel, de manquements au code du travail, voire d'abus d'autorité pouvant conduire à des agressions sexuelles répétées. Il est difficile voire impossible de s'assurer que les exemples employés dans ce mémoire n'ont impliqué dans leur processus de développement aucun problème éthique, juridique ou moral. J'invite le lecteur à rester attentif aux problèmes qui gangrènent le secteur vidéoludique, et souhaite réitérer ici ma volonté de ne pas placer sur un piédestal les exemples employés tout au long de ce mémoire. Si je les utilise parce que je les crois pertinents dans le cadre de ma recherche, je ne cautionne pas pour autant les actes individuels déplorables et pénalement répréhensibles des personnes concernées par de graves allégations.

⁴ https://en.wikipedia.org/wiki/Women_and_video_games#Data_collection , consulté le 7 juillet 2020

Structure du mémoire

Ce travail de recherche a été conçu à travers les apprentissages de son auteure, le développement de son intérêt pour le jeu vidéo tout autant que pour la recherche universitaire. Partant d'une simple curiosité, il se compose autour de l'exploration de nombreux champs de recherche, du cognitivisme aux études cinématographiques en passant bien sûr par l'étude du jeu vidéo. Sa forme naît des questions supplémentaires que les premières hypothèses ont apportées.

Cherchant dans un premier temps (au chapitre 1) à mieux comprendre le sonore et le lien qu'il construit avec le visuel, nous allons nous approcher assez naturellement d'un autre média audiovisuel, le cinéma. Nous appréhendons la co-construction complexe de l'audio et du visuel, deux perceptions qui se mettent à fonctionner ensemble pour proposer une expérience cohérente apportant à la fois des informations narratives et des ressentis. À travers l'exploration de la prise de vue réelle, nous avons commencé à cerner l'importance du sonore. Mais nous faisons ensuite évoluer nos observations en prenant en compte le cinéma d'animation, se rapprochant du jeu vidéo à travers l'image construite de toutes pièces, qui évoque ce qu'elle évoque uniquement parce que le spectateur entre dans une démarche de compréhension volontaire. Dans le jeu vidéo, où un environnement virtuel est créé, inventé par les concepteurs, les enseignements apportés par le cinéma d'animation sur le lien entre le son et l'image constituent des ressources fondamentales et passionnantes.

Il nous faut envisager ensuite les spécificités du jeu vidéo. Ce dernier, qui se détache déjà du jeu à travers plusieurs caractéristiques, se différencie absolument de ses cousins audiovisuels à travers sa perception. Là où le spectateur écoute et regarde, le joueur expérimente. Ses actions découlent de suggestions, de possibilités offertes par le jeu et elles se répercutent sur ce dernier. Dans ce cadre mouvant, la construction du sonore n'est plus

linéaire comme au cinéma, elle s'adapte aux actions du joueur. Dans le chapitre 2, nous étudierons donc plus profondément la situation de conflit en tant que situation de jeu, un espace créé par un concepteur ou une conceptrice et dans lequel un joueur fait évoluer un personnage jouable selon ses propres choix. Cette prise de conscience du jeu en tant qu'expérience a permis de diriger cette recherche vers une considération de la jouabilité et des conséquences spécifiques qui lui sont imputables. La construction sonore fait intervenir de nombreux signaux, de nombreuses sources qui acquièrent des fonctions permettant une meilleure jouabilité. Dans ce cadre, nous développons alors l'idée d'une concrétion sonore, d'une accumulation de strates audio constituant un ensemble solide et perçu comme un tout. Pour présenter ce concept, nous nous livrons à une étude de cas avec le jeu *Assassin's Creed 2 : Brotherhood*.

En étudiant de plus près cette concrétion sonore, il apparaît que le joueur a toutefois besoin d'informations pour vivre une meilleure expérience de jeu. Nos strates sonores, si elles ne sont pas sécables, opèrent toutefois selon des priorités, une hiérarchisation effectuée par le joueur avec l'aide du concepteur qui a pensé l'expérience. En explorant de nombreuses situations de conflit dans le chapitre 3, nous avons considéré ces strates non pas à travers la nature du son entendu (s'agit-il d'une musique? D'un bruit? D'une alerte? La question, finalement, n'est pas vraiment là) mais bien selon sa fonction. La situation de conflit repose sur un enchevêtrement d'alertes et d'informations dirigées vers le joueur, qui peut les appréhender en confrontant ses propres apprentissages à la situation dans laquelle le place son expérience de jeu. De *World of Warcraft* à *Crypt of the Necrodancer*, en passant par *Dark Souls*, *The Legend of Zelda : Breath of the Wild*, *Minecraft* ou *Hollow Knight*, c'est à travers l'expérimentation, en nous appuyant sur les expériences menées par des auteurs et des

auteurs, que nous avons finalement perçu la direction de ce mémoire en mettant à l'épreuve notre concept de concrétion sonore.

Le cadre de la recherche est désormais établi, nos hypothèses, nos problématiques et nos attentes sont fixées. Il nous faut désormais entrer au cœur de notre exploration et découvrir les spécificités de la construction sonore vidéoludique à l'intérieur d'une situation de conflit.

Chapitre 1 – Le son audiovisuel

Nous nous intéressons dans cette recherche au son vidéoludique et à ses apports.

Avant d’y parvenir, nous devons toutefois définir ce qu’est le sonore, et mettre en perspective le champ ludomusicologique avec ses origines : l’étude du son audiovisuel. Dans ce chapitre, nous allons nous pencher sur théories du son relié à l’image, et nous arrêter sur deux aspects constitutifs du son : ses caractéristiques et ses fonctions. Nous avancerons ensuite l’hypothèse que le son vidéoludique emprunte également, si ce n’est plus, à l’esthétique du cinéma d’animation.

Commençons par définir notre objet à travers le lexique qui lui est consacré. Le son est déterminé à la fois par une série de caractéristiques physiques (la période, l’amplitude et la fréquence d’une onde vibratoire) et une sensation physiologique : selon le CNRTL, le son est la « Sensation auditive produite sur l’organe de l’ouïe par la vibration périodique ou quasi-périodique d’une onde matérielle propagée dans un milieu élastique, en particulier dans l’air »⁵. Tout ce que nous entendons est constitué d’un ou de plusieurs sons : une voix humaine qui transmet une information sémantique, un signal d’appareil électronique, les pas de quelqu’un qui marche ou encore le rechargement d’une arme.

Ces deux derniers sons, sur un plan physique, sont en réalité des bruits. Si le son est une onde vibratoire, le bruit est au contraire un « ensemble de sons, d’intensité variable, dépourvus d’harmonie, résultant de vibrations irrégulières »⁶. En physique, le bruit est une onde irrégulière qui s’oppose au son. Sur le plan physiologique, c’est une perception qui ne peut être décomposée en plusieurs signaux auditifs distincts, qui gêne l’intelligibilité. Toutefois, la distinction entre « noise » et « sound » n’est quasiment jamais faite dans les

⁵ <https://www.cnrtl.fr/definition/sons>, consulté le 16 janvier 2020.

⁶ <https://www.cnrtl.fr/definition/bruit>, consulté le 20 janvier 2020.

sources théoriques anglophones de cette recherche. « Sound » est donc employé aussi bien pour un son mécanique (une seule onde transmise et reçue) que pour un bruit dans sa définition physique (plusieurs ondes transmises et reçues simultanément) ou dans son acceptation culturelle (un signal sonore qui ne peut être décomposé). Dans cette recherche rédigée en français, nous emploierons « bruit » et « son » selon les perceptions physiologiques de l'auditeur. Nous préciserons le sens de « sound » lorsque ce sera nécessaire.

La musique serait quant à elle la « combinaison harmonieuse ou expressive de sons »⁷, ou encore l'« Art de combiner des sons d'après des règles (variables selon les lieux et les époques), d'organiser une durée avec des éléments sonores ; production de cet art (sons ou œuvres) »⁸. La musique est une séquence d'une certaine durée de plusieurs sons, perçue à travers un prisme harmonique référentiel. Cognitivement parlant, la musique est l'apport auditif le plus simple et le plus intuitif à repérer. Elle est une organisation volontaire de sons, le résultat d'un compositeur qui a souhaité placer ces sons de la manière qui est perçue. Lorsque Pierre Boulez, Jean-Pierre Changeux et Philippe Manoury tentent de décrire ce qu'est la musique (*Les Neurones enchantés*, 2014), ils se heurtent aux définitions qui ont émaillé les différents siècles, et qui accordent plus d'importance à l'appréciation esthétique – voire physique – de la musique qu'à de potentielles règles de construction formalistes, objectives et pérennes. La musique, à travers les siècles, c'est ce qui est beau, ce qui est agréable à l'oreille (Boulez *et al.* 2014, 9-15). Si cette définition possède un travers subjectif, historique et culturel évident, on peut en tirer toutefois l'importance de la perception de l'auditeur, qui détermine ainsi le statut du signal sonore qu'il écoute. Nous verrons dans la suite de cette recherche que la musique est omniprésente en jeu vidéo, et qu'elle permet souvent au joueur

⁷ <https://www.cnrtl.fr/definition/musique>, consulté le 20 janvier 2020.

⁸ <https://dictionnaire.lerobert.com/definition/musique>, consulté le 20 janvier 2020.

de repérer à la fois le lieu et la situation dans lesquels son avatar se trouve. Il nous faut constater aussi que les frontières entre les trois termes définis – bruit, son et musique – sont de plus en plus perméables, l’irruption de l’électro-acoustique ayant bouleversé les codes de la création sonore et musicale. Le débat sur la modernité et les propriétés sonores a toujours lieu, et il confirme à la fois le brouillage de la séparation et termes et la nécessité pour l’auditeur de déterminer lui-même le sens qu’il confère à ce qu’il entend.

Son et musique constituent la moitié de tout médium audiovisuel. Si c’est bien souvent l’image qui est placée au premier plan dans les études cinématographiques, certains théoriciens rappellent l’importance du sonore pour co-construire un environnement cohérent. Nous allons désormais nous intéresser au son dans sa dimension audiovisuelle.

1.1 Caractéristiques du son au cinéma

L’une des principales caractéristiques du son vidéoludique est bien entendu son rapport à l’image. En effet, le son seul tout autant que l’image seule ne forment pas la base vidéoludique sur laquelle nous appuyons cette recherche. C’est bel et bien le mélange de deux sens, l’ouïe et la vue, qui constitue l’expérience de jeu du joueur. Il convient alors de caractériser le son à l’image et les différentes théories construites autour de cette mixité médiatique. Dans ce cadre, le jeu vidéo se rapproche explicitement d’un autre média audiovisuel : le cinéma. Parce que ce sont deux médias audiovisuels, plusieurs théoriciens ont longtemps appuyé leurs propos sur l’étude du cinéma et son influence sur le développement vidéoludique. Notre étude sur le son devrait s’enrichir fortement des études menées notamment par Michel Chion, qui a cherché à décrire le son lorsqu’il est corrélé à l’image.

1.1.1 Audio-visuel et valeur ajoutée

Michel Chion, lorsqu'il évoque les principaux liens qui constituent un média « audio-visuel »⁹, s'appuie sur la naissance de fonctions du son. Pour lui, le son constitue une « valeur ajoutée » ([1990] 2013, 16) qu'il définit ainsi : « La valeur expressive et informative dont un son enrichit une image donnée, jusqu'à donner à croire [...] que cette information ou cette expression se dégage "naturellement" de ce qu'on voit, et est déjà contenue dans l'image seule. » Selon Michel Chion, cette impression est fautive, car on pourrait croire « que le son est inutile, et qu'il redouble un sens qu'en réalité il amène et crée. » (16)

Le chercheur caractérise ensuite les différentes valeurs ajoutées du son sur l'image : tout d'abord à travers le langage (« voco-centrisme et verbo-centrisme » (17), qui vont structurer l'expérience visuelle vécue grâce à la sémantique de la « voix comme support de l'expression verbale » (17). Pour Michel Chion, « si le son au cinéma est voco- et verbo-centriste, c'est d'abord parce que l'être humain, dans sa conduite et ses réactions quotidiennes, l'est aussi. » (17) L'auteur caractérise ensuite la valeur ajoutée de la musique grâce à son « effet empathique et anempathique » (19) Il s'agit de la structuration de l'émotion vécue par l'auditeur-spectateur. Cette émotion peut résider dans le rapport direct entre la musique jouée et l'image vue (effet empathique) : « la musique exprime directement sa participation à l'émotion de la scène, en revêtant le rythme, le ton, le phrasé adaptés, cela évidemment en fonction de codes culturels de la tristesse, de la gaieté, de l'émotion et du mouvement. » (19) Notons que l'on retrouve la valeur culturelle accordée à la musique que nous avons précisée dans les définitions des termes employés. Dans le cadre de la musique

⁹ Nous empruntons la forme orthographique « audio-visuel », qui constitue le point majeur avancé par Michel Chion dans son ouvrage *L'audio-vision* où il développe l'importance d'une co-construction de l'image et du sonore pour former un ensemble cohérent.

anempathique, ce sont des musiques « dont la frivolité et la naïveté étudiées renforcent dans les films l'émotion individuelle des personnages et du spectateur, dans la mesure même où elles affectent de les ignorer. » (19) L'auteur précise également que des musiques peuvent être « ni empathiques ni anempathiques, [elles] ont soit un sens abstrait, soit une simple fonction de présence, une valeur de poteau indicateur – en tout cas pas de résonance émotionnelle précise. » (20), puis il ajoute que des bruits peuvent aussi être anempathiques : « lorsque par exemple, dans une scène très violente ou après la mort d'un personnage, un processus quelconque (bruit de machine, ronronnement d'un ventilateur, jet d'une douche, etc.) continue de se dérouler *comme si de rien n'était* » (20). On distingue dans ces analyses que le son relié à l'image, dans le cinéma, peut guider le sens de la scène développée.

Toutes ces perceptions développées par Michel Chion reposent sur le phénomène de synchrèse : « La synchrèse (mot que nous forgeons en combinant « synchronisme » et « synthèse ») est la soudure irrésistible et spontanée qui se produit entre un phénomène sonore et un phénomène visuel ponctuel lorsque ceux-ci tombent en même temps, cela indépendamment de toute logique rationnelle. » ([1990] 2013, 64) Celle-ci découle principalement d'un point de synchronisation, défini par Chion de cette manière : « c'est, dans une chaîne audio-visuelle, un moment saillant de rencontre synchrone entre un moment sonore et un moment visuel » (60). Il prend pour exemple le coup :

Dans le réel, un coup de poing, même s'il fait du mal, ne fait pas forcément du bruit. Alors que dans une audio-image de cinéma ou de télévision, le bruit d'impact est quasiment obligatoire – autrement on n'y croirait pas [...]. Ceux-ci sont donc presque systématiquement accompagnés de bruit. Cette rencontre ponctuelle, instantanée, abrupte d'un son et d'un impact visible devient alors la représentation la plus directe et la plus immédiate du point de synchronisation audio-visuel, en tant que repère, que ponctuation [...]. [Le coup] est le point audio-visuel vers lequel tout converge, à partir duquel ça se diffuse. Et il est aussi, dans l'audio-image, l'expression privilégiée de l'instantanéité. (Chion [1990] 2013, 63)

La valeur ajoutée du son passerait ainsi par une synchronisation ponctuelle du son attendu et de l'image visualisée. Cette synchrèse est au cœur de l'expérience audio-visuelle et permet de créer un lien logique entre ce que nous entendons et ce que nous voyons. Mais il arrive aussi que nous ne puissions voir ce que nous entendons. À travers l'exploration des sources sonores et la définition de l'acousmatique, nous allons désormais nous intéresser au son lorsqu'il n'est pas directement relié à l'image.

1.1.2 Sources sonores et acousmatique

Les sons que nous identifions au cinéma proviennent de sources, qu'elles soient visibles dans le cadre ou non. Chaque source sonore génère un ou plusieurs sons, et « Il ne faut pas oublier qu'un son n'a souvent pas qu'une source, mais deux au moins, voire trois ou plus encore. Prenons le crissement du stylo feutre avec lequel est écrit le brouillon de ce texte ; les deux sources du son sont le stylo et le papier – mais aussi bien le geste d'écrire, et encore nous qui écrivons et ainsi de suite. » (Chion [1990] 2013, 35) Toutefois, le repérage des sources sonores au cinéma n'est pas forcément évident puisque s'immisce la question du hors-champ dans l'exploration du cadre.

Au cinéma, si la source sonore est visible, Michel Chion parle d'une « écoute *visualisée* (c'est-à-dire accompagnée de la vision de la cause/source) » ([1990] 2013, 72). Quant au son acousmatique, il est défini ainsi : « signifie "que l'on entend sans voir la cause originare du son", ou "qui fait entendre des sons sans la vision de leurs causes" » (71). Bien entendu, une source sonore ne reste pas visible ou cachée, puisqu'elle peut se déplacer. Toutefois, au cinéma, c'est le créateur du film qui décide si la source devient visible ou non : le contrôle de la caméra assure le contrôle de la présence de la source sonore à l'écran. L'auditeur-spectateur subit la décision du créateur de montrer ou non la source sonore. Si elle n'est pas

présente à l'écran, il va d'ailleurs en chercher plus activement la source en invoquant son encyclopédie personnelle, mais cette démarche est seulement cognitive : si la source sonore n'apparaît pas à l'écran, l'auditeur-spectateur ne peut pas choisir de la rendre visible.

À cet égard, Michel Chion détermine six trajets du son, du visualisé à l'acousmatique et inversement. Il prend aussi en compte les stratégies de *in*, de *off* et de son hors-champ pour déterminer des trajets. Nous estimons cependant que ces échanges ne fonctionnent pas du tout de la même manière en jeu vidéo, où le joueur possède un contrôle sur la caméra et peut donc très souvent choisir de rendre ou non la source sonore apparente.

1.1.3 Stratégies d'écoute au cinéma

Michel Chion identifie dans *L'audio-vision* « trois attitudes d'écoute différentes, visant des objets différents : l'écoute causale, l'écoute sémantique et l'écoute réduite » ([1990] 2013, 33). L'écoute causale, tout d'abord, constitue notre capacité à « se servir du son pour se renseigner, autant que possible, sur sa cause » (33). Les sources sonores, comme nous l'avons vu ci-dessus, peuvent être visibles ou invisibles dans le cas du hors-champ. Toutefois, Michel Chion précise que cette attitude d'écoute n'est pas forcément fiable : nous ne jugeons pas d'un seul son mais d'un son entendu dans un contexte, issu de plusieurs sources sonores (c'est donc plutôt un bruit). Cette écoute peut donc devenir trompeuse (précisons que la tromperie peut être souhaitée par les créateurs de l'expérience).

Michel Chion évoque ensuite l'écoute sémantique. Il s'agit de l'écoute « qui se réfère à un code ou un langage pour interpréter un message » (35). Nous passons brièvement sur cette écoute car elle est plutôt logique : dans *Hellblade : Senua's Sacrifice* (2017, Ninja Theory) par exemple, le joueur doit être attentif à des conseils vocaux donnés par une narratrice pour

comprendre comment vaincre l'antagoniste auquel son avatar est confronté. Il est cependant pertinent de noter que « l'écoute causale et l'écoute sémantique peuvent s'exercer parallèlement et indépendamment sur une même chaîne sonore. Nous entendons à la fois ce que quelqu'un nous dit, et comment il le dit » (Chion [1990] 2013, 36). Les écoutes s'entrecroisent donc, elles ne sont pas le résultat d'une attitude consciente mais bien un mélange d'attitudes ayant pour but de créer du sens, ou de décoder ce dernier. Dans un jeu vidéo, nous serons parfois capables d'identifier ou de supposer le caractère antagoniste d'un personnage qui n'est pas encore agressif, en fonction de son timbre de voix et d'une précédente rencontre auditive avec ce dernier. Nous nous arrêterons plus en détail sur ce principe dans les chapitres 2 et 3.

Enfin, l'écoute réduite est un concept emprunté à Pierre Schaeffer et défini par Chion comme tel : c'est une écoute « qui se porte sur les qualités et les formes propres du son, indépendamment de sa cause et de son sens ; et qui prend le son – verbal, instrumental, anecdotique ou quelconque – comme objet d'observation, au lieu de le traverser en visant à travers lui autre chose » ([1990] 2013, 36). Comme cette attitude d'écoute repose sur une réécoute et une attitude volontairement savante, et surtout qu'elle « implique la fixation des sons » (37), nous ne pensons pas qu'elle est pertinente dans le cadre de la situation de conflit vidéoludique où le joueur agit plutôt dans l'instant, sans s'arrêter pour se concentrer, réécouter, revivre un signal sonore particulier. Michel Chion lui-même, dans le cadre du cinéma, la met de côté en argumentant que son utilisation permet « d'affiner l'oreille du réalisateur, du chercheur ou du technicien, qui ainsi connaîtront le matériau dont ils se servent et le maîtriseront mieux » (37). Nous pouvons seulement affirmer ainsi que les concepteurs de l'expérience de jeu sont conscients de ce qu'ils font, et que l'environnement ainsi créé est

volontairement construit de manière à mettre plus ou moins en avant certains sons pour aider ou tromper le joueur.

1.2 Fonctions du son au cinéma

Après avoir détaillé les caractéristiques techniques du son audiovisuel, nous allons nous intéresser aux fonctions qui en découlent. Au cinéma, plusieurs types de sons cohabitent pour former un ensemble sonore complexe et cohérent. Des chercheurs vont argumenter sur les possibles fonctions des sons, reliant ainsi volonté du créateur et perception du spectateur. Parmi ces classifications, nous avons choisi d'explorer plus en détail celle d'Annabel J. Cohen, qui s'intéresse au son audiovisuel à travers une approche cognitive de la perception de l'auditeur-spectateur.

1.2.1 Le Congruence-Association Model

Annabel J. Cohen, lorsqu'elle s'intéresse à l'interprétation d'un film à travers sa musique, va identifier deux processus cognitifs qui semblent se répéter dans la compréhension d'une image cinématographique et de son accompagnement sonore :

One process exploited the shared (congruent) temporal structure of the music and film, guiding attention to a particular part of the film. The second process directed the meanings (or associations) of the music to the place on the screen where visual attention has been directed by the first process. (Cohen *et al.* 2013, 21).

Elle développe alors, sur plus de vingt-cinq ans, plusieurs itérations du *Congruence-Association Model*. Celui-ci, issu directement des études cognitivistes, fait état des processus de compréhension qui se déroulent chez l'auditeur-spectateur durant le visionnage d'un film.

Le *Congruence-Association Model* (abrégé en CAM) part du principe selon lequel nous cherchons à donner un sens au monde fictif que nous propose l'expérience audiovisuelle (Cohen *et al.* 2013, 21). Pour ce faire, le cerveau humain va prioriser les informations qu'il

reçoit à travers plusieurs modalités de perceptions (qui sont donc congruentes). Les congruences sont des informations structurelles entre plusieurs canaux, tandis que les associations sont les informations de sens transmises par l'expérience audiovisuelle et recherchées activement par notre cerveau pour construire un sens logique.

So there are three parts to the notion of congruence : (1) a brain response to coordinated information across modalities, (2) assignment of privileged status to this information (it has a higher priority for attention than non-congruent activity), and (3) conscious attention that may be disproportionately directed to the congruent information across modalities. (Cohen *et al.* 2013, 25).

Les associations, quant à elles, sont formées d'après les processus cognitifs de *top-down* et de *bottom-up*. L'auditeur-spectateur construit une attente de sens en fonction des informations issues de sa mémoire (*top-down*). Ces informations, si elles passent forcément par des biais culturels et sociologiques, permettent de créer un sens rapidement face à l'expérience audiovisuelle. Dans le processus *bottom-up*, ce sont les informations issues de l'expérience audiovisuelle qui vont construire les données mémorielles de l'auditeur-spectateur. Celui-ci actualise alors ses connaissances, et donc ses attentes lors de la prochaine itération culturelle à laquelle il sera confronté.

Dans le modèle CAM-3, que nous présentons plus bas (fig. 1), Annabel J. Cohen identifie cinq canaux médiatiques (les modalités qu'elle évoquait précédemment) selon le modèle des langages cinématographiques développé par le sémiologue en audiovisuel Christian Metz (1974) : le texte, le discours, le visuel, la musique et les effets sonores. Ces cinq canaux, qui possèdent chacun une structure et un sens, vont permettre à l'auditeur-spectateur de reconstituer l'expérience narrative à travers trois processus : un processus *bottom-up*, depuis les canaux (ligne A) vers les attentes générées par la mémoire à long terme (ligne D) ; un processus *top-down*, depuis la ligne D vers le travail de constitution de sens

(*working narrative*¹⁰, ligne C) ; et enfin un processus de congruence qui permet de récupérer plus rapidement la structure médiatique en place lorsqu'elle est représentée par plusieurs canaux : « These congruencies can contribute to the direction of attention. This is a parsimonious system, because the same process is carried out on all domains of information¹¹. In other words, the model assumes a structural analysis independent of that of meaning. »

(Cohen *et al.* 2013, 31) De la même manière :

an associationist analysis of meaning applies to all channels of information, here referred to as *meaning* at level B. Music brings certain associations to mind, but so do sound effects, words as text or speech, and visual scenes. The brain does not necessarily care where the semantic information comes from. (32).

¹⁰ Ce terme, qui vient remplacer celui de *visual narrative* dans les précédentes itérations du CAM, s'appuie sur les travaux de Joan E. Foley autour des *working maps*: « Her exposure to Baddeley's ideas during the early 1980s, led to her proposing a notion of the *working map* that an individual mentally generates to guide his or her wayfinding in an environment. Such working maps are continually updated by information received by the senses or by information about large-scale space stored in LTM (e.g., floor plans). Some persons may rely on visual or auditory imagery while others might make use of more abstract constructions based on their individual thinking styles, abilities, and experiences (Foley & Cohen, 1984). The notion of the working narrative captured this dynamic construction of reality, or pseudo-reality, of the film as shown in Figure 2.4 at level C of CAM-3. » (Cohen *et al.* 2013, 31)

¹¹ Annabel J. Cohen précise à la page 31 que les congruences s'appliquent à tous les canaux d'informations, elles ne sont pas toutes représentées pour faciliter la lecture du schéma.

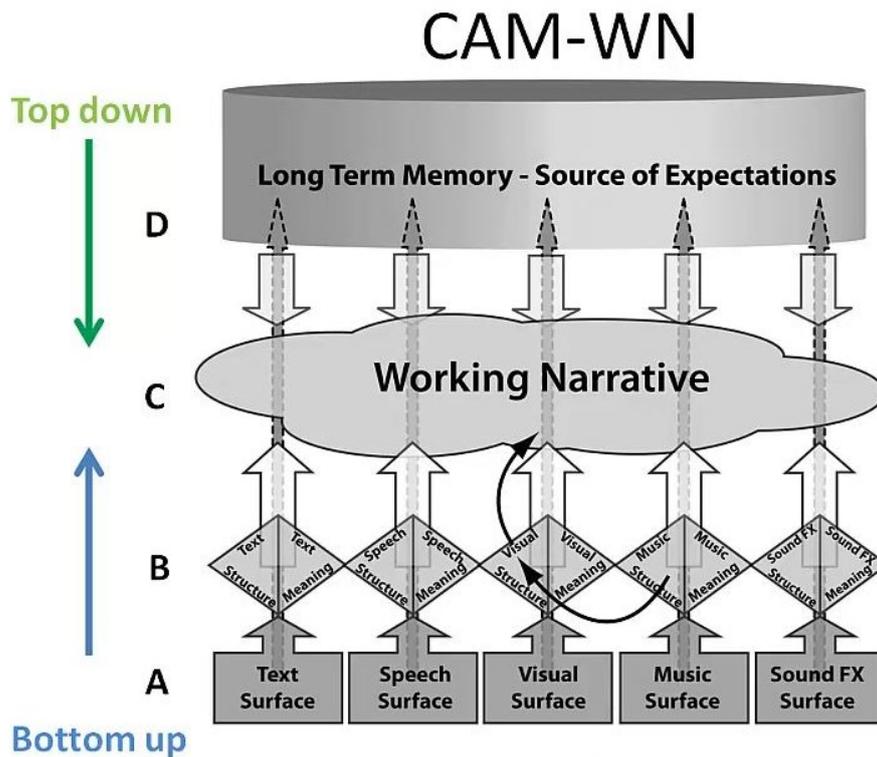


Figure 1. Congruence-Association Model – Iteration 3. © Cohen *et al.*, *The Psychology of Music in Multimedia*, 2013, 33 (issu de l’information audiovisuelle¹² disponible en ligne).

Le modèle présenté par Annabel J. Cohen permet de mettre en avant la multiplicité des canaux d’information dans une expérience audio-visuelle. Pris indépendamment les uns des autres, ces canaux véhiculent un sens et une structure qui peuvent différer selon le canal. Le cerveau humain, qui va travailler continuellement pour obtenir une cohérence narrative de l’expérience audiovisuelle, privilégie les structures communes entre plusieurs canaux pour construire sa compréhension.

1.2.2 Les fonctions audiovisuelles

Annabel J. Cohen va aussi s’intéresser aux fonctions de la musique dans une expérience audiovisuelle. Elle justifie cet intérêt par les théories cognitives qui présentent la

¹² Le même modèle est représenté directement dans le livre, à l’exception des flèches *top-down* et *bottom-up*. Nous trouvons cette version plus lisible. Le titre « CAM-WN » réfère à « Congruence-Association Model – Working Narrative ». <https://global.oup.com/booksites/content/9780199608157/>, consulté le 23 juin 2020.

compréhension de la musique comme un phénomène à part dans le traitement des divers canaux d'information par le cerveau (1998, 13). Nous choisissons également de nous arrêter sur ces fonctions car certaines d'entre elles sont omniprésentes dans l'analyse des sons vidéoludiques. L'article de Cohen, qui possède l'atout de distinguer ces fonctions de manière très claire, va nous servir de base pour la suite de cette recherche.

La première fonction développée par Annabel J. Cohen est une fonction de masquage, « masking » (1998, 14). En effet, dans un environnement multimédia, la complexité sonore de la musique va attirer l'attention de l'auditoire, et ainsi l'empêcher de se disperser vers des sources sonores moins importantes : « As music is coherent it will occupy limited auditory attention that would otherwise have been dissipated on extraneous sounds. » (14) Cette fonction est à double tranchant car la musique ne doit cependant pas masquer d'autres signaux sonores porteurs de sens, notamment les dialogues cinématographiques ou toute autre source sonore sémantique.

L'auteure évoque ensuite la « provision of continuity » (14). Il s'agit de la capacité de la musique à créer une continuité temporelle sonore, à laquelle des images peuvent se rattacher pour former une histoire cohérente : « this organization helps to connect disparate events in other domains. » (14)

La troisième fonction de la musique appliquée au multimédia est la « direction of attention » (Cohen 1998, 14). Cohen exploite ici l'idée que le cerveau, en manque constant de signification, va créer des liens de sens et de structure grâce à la musique qu'il entend. Il s'agit en fait de la situation de congruence que Cohen développe dans son modèle cognitif : la synchronèse entre l'audio et le visuel va donner un sens de lecture rapide à l'auditeur-

spectateur. Le cerveau est plus susceptible de diriger son attention vers une congruence audio-visuelle que vers un signal isolé.

De cette troisième fonction découle directement la quatrième, la « mood induction » (15) que de nombreux auteurs que nous citons dans ce mémoire évoquent. La musique, à travers son apport culturel, va transmettre une émotion à celui qui l'écoute. Cohen se base sur les travaux de Pignatello, Camp & Rasar (1986) et sur ceux d'Albersnagel (1988) pour définir ce que nous traduirons par un changement de ressenti émotionnel : « Music can alter the mood or feelings of the listener-viewer » (Cohen 1998, 15). Non seulement ce changement de ressenti est démontré, mais les auteurs précédemment cités par Cohen ont également mis en avant que ce changement peut être prévu en fonction des caractéristiques musicales présentes. Ainsi, un certain type de musique sera presque toujours, culturellement, relié à la même émotion. En ludomusicologie, le changement de ressenti est très souvent amené par les auteurs qui développent les fonctions du son. Le reste de notre analyse tournera constamment autour de ce changement d'état induit par la musique, voire – mais nous le développerons par la suite – par le son lui-même.

« Mood induction changes how one is feeling, while communication of meaning simply conveys information » (Cohen 1998, 15). Cette cinquième fonction élaborée par Cohen est très proche de la quatrième. Cependant, la « communication de sens » (15, notre traduction) va se baser principalement sur l'image pour véhiculer un sens et non plus un ressenti émotionnel. La musique va acquérir un sens en fonction de l'image donnée : « Music can take on meanings through association with events (Cohen 1993) » (1998, 15). Cette idée est basée sur des perceptions culturelles suffisamment partagées pour acquérir une valeur relativement universelle. Il faut toutefois l'affirmer avec prudence, car les théoriciens qui explicitent cette

communication de sens sont avant tout issus de pays européens ou nord-américains partageant des références historiques et culturelles communes. Réutilisons l'exemple proposé par Cohen : « sadness is conveyed by slow pace, falling contour, low pitch, and the minor mode and happiness is conveyed by fast tempo, rising tempo, high pitch and the major mode (Levi, 1982 ; Trehub, Cohen & Guerriero, 1985). » (Cohen 1998, 15)

Cohen développe ensuite la sixième fonction de la « Music as a cue for memory » (16) en donnant l'exemple d'un spectateur-auditeur qui peut imaginer la présence du personnage associé à un thème musical, même si ledit personnage n'est pas présent. Il s'agit communément du concept de *leitmotiv*¹³. D'après Cohen, ces musiques comme indices mémoriels sont si efficaces qu'ils sont particulièrement utilisés en publicité (16) : en effet, une seule écoute permet déjà l'association à un visuel, et donc un ancrage dans la mémoire de l'auditeur-spectateur.

La dernière fonction majeure sur laquelle Annabel J. Cohen se concentre est le développement de l'« arousal and focal attention » (16). Nous devons rentrer de manière plus explicite dans le cognitivisme pour démontrer cet effet : « when there is music, more of the brain is active. Increased activity of the brain may increase concentration on the primary attentional focus and filter out distractions. » (16) On revient de manière plus développée aux première et troisième fonctions de la musique, la capacité à masquer d'autres signaux sonores et la direction de l'attention du spectateur-auditeur. Cohen démontre moins qu'elle évoque cette capacité de la musique à rendre l'auditeur plus attentif : « Heightening arousal level may

¹³ Dans la musique et la musique dite "classique", il est très habituel qu'un thème représente un personnage ou une émotion. Ce paradigme, popularisé au XXe siècle par Richard Wagner qui en fait un usage systématique dans ses compositions, s'appelle le *leitmotiv*. Dans ses opéras, le compositeur n'hésite pas à faire retentir le thème d'un personnage absent de la scène pour montrer son influence sur ce qui se déroule hors de son contrôle direct.

thus increase the intended impact of the multimedia experience, making the experience seem real. It can be suggested that heightened arousal temporarily suspends contact with “real” reality and facilitates belief in the virtual reality » (17). Sans musique, l’environnement créé nous semblerait moins réaliste car nous ne serions pas assez occupés pour nous concentrer sur la proposition fictionnelle, et nous garderions un lien plus important avec les stimuli qui nous entourent hors de l’expérience fictive. On comprend mieux, dès lors, pourquoi la musique possède un si grand impact sur notre ressenti émotionnel ou sur notre capacité à croire en l’environnement virtuel proposé.

Cohen conclut son résumé par une huitième fonction qu’elle développe beaucoup moins, qui est l’aspect esthétique de la musique. Si cette fonction doit, selon elle, « ne jamais être oubliée » (17, notre traduction), son traitement dépend bien moins de la situation multimédiale que du statut d’art que véhicule la musique : « music is an art form and its presence enhances every situation » (17). Il s’agit de la partie culturelle et subjective attribuée à la définition de la musique que nous évoquions au début de cette recherche : la musique est susceptible d’apporter un plaisir, une émotion à l’auditeur.

Les travaux d’Annabel J. Cohen sont une base pour étudier les fonctions de la musique dans les œuvres multimédias. Si l’environnement sonore vidéoludique est rempli de caractéristiques qui lui sont propres, principalement en raison de la modification du déroulement de l’œuvre dans le temps en fonction des informations reçues de la part du joueur actif, on peut toutefois constater à la lecture de Cohen que la musique et le sonore ont développé une affinité avec le visuel. Le résultat de cette affinité est une interdépendance entre ce que nous voyons et ce que nous entendons. L’un sans l’autre, ni visuel ni audio ne peuvent créer un environnement si riche de sens et de perceptions.

1.3 Cinéma d'animation, image construite et son superposé

Tout comme la plupart des auteurs en ludomusicologie que nous citerons dans cette recherche, notre premier réflexe pour comprendre les spécificités du son lorsqu'il est couplé à une image en mouvement a été de se tourner vers les études cinématographiques. Celles-ci induisent souvent l'utilisation d'un seul genre de cinéma, celui en prise de vue réelle. Or, nous trouvons dans le cinéma d'animation des liens plus concrets encore avec le jeu vidéo : l'image rendue à l'écran est intégralement construite, les effets sonores sont obligatoirement enregistrés à part puis superposés sur l'image d'animation. Nous allons ici tenter de comprendre de quelle manière les liens tissés entre son et image dans le cinéma d'animation constituent les fondements du rapport son-image en jeu vidéo.

1.3.1 Les fondements du son synchronisé

Il n'est pas inintéressant de signaler que les réflexions d'Annabel J. Cohen autour du lien entre sonore et visuel sont parties de la visualisation d'un film d'animation de 2 minutes mettant en scène des figures géométriques simples. Dans ce film de 1944 (qui ne porte pas de titre) réalisé par Heider et Simmel, des triangles et un disque interagissent entre eux et avec un espace ouvert représenté par cinq traits.

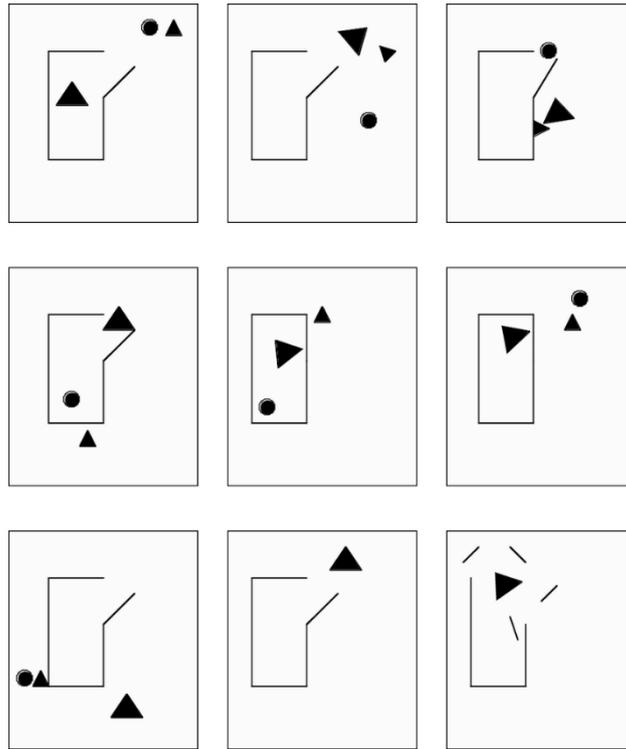


Figure 2. Quelques images du film d'Heider et Simmel, 1944. ©Evelyne Thommen 2011.

En montrant ce film à ses étudiants et en l'accompagnant de plusieurs musiques, Annabel J. Cohen fait le constat suivant :

Their motion patterns [ceux des figures géométriques dans le film] provoked attitudes that were normally associated with complex human interactions. [...] The stable finding allowed us to ask, "could different background music alter the meaning of or attitudes toward these geometric figures?" (2013, 20).

Ce questionnement de départ est explicité par Zach Whalen, qui s'intéresse plus particulièrement aux liens entre le cinéma d'animation et le jeu vidéo. D'après lui, la musique du cinéma d'animation a permis le développement de tropes encore utilisés en jeu vidéo aujourd'hui (2004). Il évoque principalement deux aspects : le sentiment de vie attribué au personnage animé et la crédibilité d'une diégèse spatiale :

Studies of cognition and animation suggest that objects are perceived as alive and exhibiting anthropomorphic behaviour when their motions are accompanied by a synchronized soundtrack (Cohen, 2000). This phenomenon relates to Chion's (1994) poetic description of the sublime incorporation of sound into the space of

the animation, and the two observations (Chion's theory and Cohen's research) are linked because, for Chion, one of the key functions of sound is to aid in an audience's perception of a spatial diegesis. The cartoon character acting in the space of the animation relies on musical accentuation to make the illusion compelling, and the result is that the musical cues and non-musical sound effects instill objects with even more life than the simple appearance of figures in motion. (2004)

Pour mieux expliquer l'importance de la musique et du son dans le cinéma d'animation, ainsi que le rapport entre ce dernier et le jeu vidéo, Zach Whalen commence par expliquer le principe du *mickey mousing*, qui renforce l'impression de vie donnée par le personnage à l'écran. La musique du cinéma d'animation fonctionne dans une attitude kinesthésique, ce qui signifie qu'elle cherche à reproduire le mouvement de la scène avant de transmettre l'émotion (2004) :

Mickey mousing, or "mickeymousing," occurs in both animated and live-action cinema when the music provides a synchronized, aural imitation of what is happening on the screen (Neumeyer and Buhler, 2001). The goal in simple mickey mousing certainly seems to be more physically or kinesthetically oriented, but in that it represents a character's relation to its fictive universe, mickey mousing roots the cartoon character in a whimsical world whose space is responsive only to the constraints of the character. [...] A very common example is a simple ascending sequence of light staccato notes drawing attention to a character tiptoeing. Music which synchronizes musical cues to physical, slapstick violence also places the character within the aural characterization of the story's space. (2004)

On constate la présence de *mickey mousing* dans de très nombreux films d'animation et quelquefois en prise de vue réelle. Ce procédé, selon Michel Chion, permet de pallier le temps de traitement de l'image par l'œil en apportant l'information du mouvement à travers le canal audio. Notre cerveau traiterait les données audio plus rapidement que les données visuelles. On peut en tout cas rapporter la congruence des informations sonores et visuelles au modèle d'Annabel J. Cohen, et déterminer ainsi que la redondance de l'information du mouvement permet un traitement et une compréhension rapides de la scène présentée. Avec un son qui fait entendre les mouvements des personnages, ceux-ci paraissent plus réalistes –

Michel Chion suggère d'enlever le son d'un film d'animation de Tex Avery pour s'apercevoir que les mouvements dessinés semblent saccadés ou trop rapides ([1990] 2013, 114).

Selon Zach Whalen, le *mickey mousing* permet à la fois d'ancrer le personnage et le décor entier du film et de relier les deux pour proposer une fiction cohérente (2004). Pour expliciter son propos, il fait une analyse sonore qui s'inscrit directement dans la lignée de l'expérience d'Annabel J. Cohen. Le film d'animation *Skeleton Dance*, réalisé en 1929 par Walt Disney et qui s'inscrit dans la série des *Silly Symphonies*, est emblématique dans l'utilisation du *mickey mousing* et le rapport entre le décor, les personnages et la musique. L'ambiance sonore du court-métrage est composée par Carl W. Stalling, qui y intègre aussi un extrait de la *Suite lyrique* d'Edvard Grieg (op. 54, « Marche des trolls »). Dans le court-métrage d'animation, on peut voir une bande de squelettes sortir de leur tombe pour danser sur la musique de Grieg. Selon Whalen, « *Skeleton Dance* (1929) demonstrates the complicated blend of diegetic and non-diegetic music and sound that merges to create an immersive story. » (2004) L'auteur explique tout d'abord que l'atmosphère « spooky » est créée par le mélange de la musique et des sons du court-métrage, qui reproduisent les mouvements des personnages ainsi que leurs réactions : deux chats paniqués lors de l'éveil d'un squelette sont représentés musicalement par une figure sonore descendante, jouée par un violon qui produit un son très strident (2004). Les choix d'instrumentation, les bruitages et la musique sont continuellement synchronisés sur les personnages du court-métrage, et ils transmettent non pas un son réaliste mais un son figuratif, qui accompagne à la fois les personnages et le visuel représenté.

Mais Whalen met ensuite en avant une caractéristique fondamentale du sonore dans le cinéma d'animation : la frontière très perméable entre le diégétique et le non-diégétique.

Les sons de *Skeleton Dance*, qu'on peut considérer comme destinés à l'auditeur-spectateur et absents de la diégèse dans un premier temps, migrent vers la diégèse lorsque les squelettes dansent en rythme sur la musique de Grieg :

These harmonic and instrumental choices demonstrate uses of music to convey a specific mood in accompanying the visuals. [...] The music that acts as a non-diegetic underscore of the skeleton's action soon blends into a choreographed dance number where the skeletons clearly respond to and play the obviously diegetic music that we hear as well. Thus the "source" of the music has become diegetic, whereas the initial mickey mousing gestures are non-diegetic underscores. (2004)

Si le sonore au cinéma est traditionnellement étudié en fonction de son rapport à la diégèse, la séparation diégétique est toutefois plus floue dans le cinéma d'animation. D'après Zach Whalen, c'est d'ailleurs cette caractéristique qui rapproche cinéma d'animation et jeu vidéo :

The fact that this shift [entre diégétique et non-diégétique] is accomplished seamlessly corresponds to the contiguity of the videogame in that atmospheric or "tone setting" sound of the videogame's worlds quickly and smoothly give way to didactic or motivational implementations that emphasize the videogame's sequence. (2004)

Nous allons pour terminer ce chapitre explorer le sonore dans son rapport à l'espace visuel présenté, et ainsi faire le lien entre image cinématographique d'animation et jeu vidéo.

1.3.2 Jeu vidéo et cinéma d'animation

Whalen fait le parallèle entre cinéma d'animation et jeu vidéo à travers l'exemple de *Super Mario Bros.* (1985, Nintendo). Selon lui, le son vidéoludique possède deux fonctions : « Specifically instances of musical sound in videogames generally follow one of two trajectories: to expand the concept of a game's fictional world or to draw the player forward through the sequence of gameplay. » (2004) S'il a conscience que les parallèles entre jeu vidéo et cinéma

sont complexes, Whalen met toutefois en avant la réception de ces deux médias, qui se fait grâce à deux sens principaux, la vue et l'ouïe : « The key, fundamental overlap between videogames and films is the fact that films and videogames basically rely on both aural and visual cues to convey a sense of a consistent diegesis or gameworld. » (2004) Et dans le parallèle entre cinéma d'animation et jeu vidéo, l'image n'est pas captée mais constituée de toute pièces. Le son, qui n'en est pas issu de manière naturelle, est enregistré séparément puis rattaché à l'image. Dans ce cadre, le lien entre son et image n'est pas dicté par le réalisme, mais plutôt par la vraisemblance, la création d'une cohérence entre l'univers présenté visuellement et les sons reproduits. En analysant *Super Mario Bros.*, Whalen expose des mécanismes qu'il a déjà évoqués dans sa présentation de *Skeleton Dance*.

On retrouve des occurrences de *mickey mousing* dans *Super Mario Bros.*. Le son des sauts de Mario dans *SMB*, par exemple, est un *glissando* chromatique ascendant. Dans une définition moins musicologique, il s'agit d'un son court où l'oreille reconnaît un mouvement depuis un début grave vers une fin plus aigüe. La caractéristique principale du *glissando* est l'absence de paliers dans cette montée sonore, elle est continue. On peut aussi concevoir comme du *mickey mousing* l'avertissement sonore qui résonne lorsque Mario récupère une pièce. Dans ce son ascendant, on peut reconnaître deux notes distinctes (qui forment une quinte). Ces deux notes rebondissent sur l'effet visuel perçu par le joueur : si la pièce était visible dans le niveau, elle disparaît seulement, mais si elle apparaît depuis une brique que Mario a frappée en sautant (dans un mouvement du bas vers le haut), alors la pièce s'envole avant de disparaître. Dans les deux cas, le score de la découverte de la pièce s'affiche brièvement à l'endroit de l'impact, avant de disparaître vers le haut en imitant la pièce. Le son, qui possède quasiment la même durée que l'animation (il est légèrement plus long), témoigne de ce mouvement ascendant. En haut de l'écran du joueur apparaissent les

informations sur le nombre de pièces et le score de la partie. Ce trope est un moyen, selon Whalen, de renforcer la crédibilité de l'environnement fictif proposé. « The cartoon character acting in the space of the animation relies on musical accentuation to make the illusion compelling, and the result is that the musical cues and non-musical sound effects instill objects with even more life than the simple appearance of figures in motion. » (2004)

Ainsi, le saut de Mario serait mieux intégré à son environnement numérique grâce au son qui l'accompagne. Cette analyse nous renvoie à la manière de composer de Koji Kondo, qui crée les six thèmes musicaux accompagnant *Super Mario Bros*. Dans une entrevue avec *Wired*¹⁴, on peut apercevoir une partie de son processus de composition :

“I wanted to create something that had never been heard before”, he says, “something not like game music at all.” Some of the tunes came easy, like the lilting, soothing music that trills when Mario is swimming underwater, or the staccato bass line that accompanies his travels underground. But that signature theme, the first one in the game, took the longest to compose. Kondo would write one version, and then he and the team would put it into the game. If it didn't accentuate the action perfectly, didn't time up right with Mario's running and jumping, didn't harmonize with the different sound effects, he'd scrap it and start over. (Wired 2007)

La musique renforcerait donc, à l'instar de l'exemple de *Skeleton Dance*, les différents sons du jeu. Zach Whalen va également développer l'utilisation de la musique pour accentuer une émotion, un sentiment transmis par l'univers représenté à l'écran. Il utilise en exemple la modification importante du thème lorsque le personnage jouable, Mario, se retrouve face à Bowser, le boss du jeu (un antagoniste particulièrement puissant). Lorsque Mario entre dans les niveaux qui se concluent sur un face à face avec Bowser, on note un changement d'ambiance radical, notamment de rythme et d'harmonie chromatique :

At a broad level, the musical character's *[sic]* of each world's levels advances *[sic]* concludes with the ominous fourth level which acts as a motivational cue similar to

¹⁴ <https://www.wired.com/2007/03/behind-the-mario-maestros-music/>, consulté en ligne le 21 janvier 2020

the tempo increase at the end of each individual level. These "castle" levels build up to an ultimate battle with a boss character, a subordinate manifestation of Bowser. This music is similar to the Underworld theme in its lack of tonal center and reliance on chromatics. Here, the confined space afforded to the player [...] is mirrored in the dense cluster of notes that carry the theme. (Whalen 2004)

Les changements d'ambiance de *SMB* sont construits par rapport aux niveaux, aux situations ; en d'autres termes, par rapport à la situation de jeu proposée : « Composer Koji Kondo again uses particular melodic themes to identify specific areas of the gameworld in something like Wagner's leitmotifs acting in reverse. » (2004) Dans *SMB*, les thèmes sont représentatifs des environnements dans lesquels Mario évolue.

Le lien entre son cinématographique et son vidéoludique semble s'éloigner de l'appartenance à la diégèse pour se concentrer sur la capacité à rendre un univers fictionnel cohérent tout en permettant au joueur, à travers des situations de congruence, de comprendre rapidement la situation de jeu dans laquelle il se trouve. Dans ce chapitre, nous avons mis en avant l'apport du sonore sur le visuel, le traitement cognitif des informations sonores lorsqu'elles sont synchronisées avec l'image, déterminé les fonctions attribuables au son et notamment à la musique lorsqu'elle est reliée à l'image cinématographique, et présenté les rapprochements entre le cinéma d'animation et le jeu vidéo. Nous allons désormais nous concentrer sur les spécificités du son vidéoludique pour comprendre de quelle manière il se construit d'après les modèles évoqués dans ce chapitre.

Chapitre 2 – Jeu vidéo et concrétion sonore

Le chapitre précédent nous a permis de circonscrire le sonore lorsqu'il est relié à l'image, mais plutôt à l'image cinématographique. À travers l'étude de *Super Mario Bros.*, Zach Whalen s'intéresse à la fois à la caractérisation sonore des personnages et au traitement de l'espace par la musique du cinéma d'animation. Toutefois, il ajoute dans son analyse un point que nous avons encore peu abordé, et qui va constituer le pilier central de cette recherche : les sons et la musique du jeu vidéo ont un rapport avec sa jouabilité. Whalen l'évoque ainsi : « music and sound effects function in encouraging successful game play by providing positive reinforcement as consequences for actions in the game » (2004) en prenant pour exemple le son de mort qui rappelle la musique d'échec à un jeu télévisé, ou encore les sons décrits comme « positifs » de récupération d'une pièce. Selon Whalen, les sons encouragent à poursuivre l'expérience de jeu en transmettant des signaux de succès. Nous allons dans ce chapitre chercher à comprendre de quelle manière le son peut véhiculer des informations au joueur, et argumenter sur la nécessité d'aborder son et musique dans un ensemble cohérent puisqu'ils semblent se rejoindre pour former l'expérience de jeu.

Dans un premier temps, il nous faudra repartir du cognitivisme pour déterminer quels sont les mécanismes en fonctionnement lorsqu'un joueur joue. Nous devons également, afin de circonscrire cette recherche, définir notre cadre autour de la jouabilité. Par la suite, nous prendrons en compte les attitudes d'écoute du joueur pour déterminer des fonctions du son vidéoludique. À travers quelques exemples, nous souhaitons présenter le sonore comme un ensemble formé de plusieurs strates, qui si elles ont chacune leur spécificité ne peuvent pas être envisagées de manière totalement indépendante. Musique et sons semblent reliés pour former une expérience sonore qui participe à l'expérience vidéoludique tout entière. En prenant en compte cette analyse, nous nous pencherons sur plusieurs modèles d'études du

sonore et nous arrêterons sur celui proposé par Karen Collins, qui étudie les sons en fonction de leur lien avec le joueur.

2.1 Spécificités du jeu vidéo

L'approche cognitive d'Annabel J. Cohen, si elle est consacrée à la musique audiovisuelle, va servir de point d'attache pour le reste des théories développées dans cette recherche. Nous avons établi que le sonore, lorsqu'il est en synchronisme – selon Michel Chion – avec l'image, permet à l'auditeur-spectateur de mieux appréhender la scène cinématographique qui se déroule. D'après Cohen, ces congruences entre deux systèmes d'information sont les vecteurs de la direction de l'attention, et d'une compréhension plus rapide de la scène. Dans le jeu vidéo, nous ajoutons toutefois certains paramètres : le joueur agit sur le jeu à travers une interface de contrôle qui prend une forme kinesthésique (une manette, un clavier ou une souris). Le jeu réagit aux actions du joueur, son déroulement n'est pas linéaire. Dans ce cadre, la prise en compte des informations données par le média est modifiée. Nous allons voir quelles sont les modalités spécifiques du sonore dans le jeu vidéo.

2.1.1 Jouabilité et cycle magique

Roux-Girard, pour expliquer la jouabilité, nous ramène à la définition qui en est donnée dans *Half-Real*, de Jesper Juul, selon les propos de Richard Rouse (2001) : « A game's gameplay is the degree and nature of the interactivity that the game includes, i.e., how the player is able to interact with the game-world and how that game-world reacts to the choices the player makes. » (Rouse cité dans Juul 2005, 122) Ce *gameplay*, que nous traduisons par *jouabilité*, serait donc la clef qui relie le joueur aux concepteurs de jeu, à travers un environnement créé pour être exploré plus ou moins librement par le joueur. Cette jouabilité découle du processus de jeu. Guillaume Roux-Girard, en s'appuyant sur les travaux d'Arsenault et Perron (2009),

explicite bien que la jouabilité ne prend en compte que ce que le joueur, dans son expérience ludique, est en mesure de percevoir :

[L]a jouabilité ne doit pas être perçue comme « le » système de jeu, mais plutôt comme « l'expérience ludique » qui émerge de la relation se tissant entre le joueur et le système de jeu. Ainsi, bien que l'exposition de la structure interne de l'algorithme de jeu ou encore l'étude de l'implémentation du son dans un jeu donné puissent mettre de l'avant plusieurs faits intéressants visant à l'amélioration des techniques de conception, il est très important de comprendre, qu'à travers les yeux d'un simple joueur, l'expérience de la jouabilité n'est pas représentée par une série de codes défilant à l'écran, ni par une représentation exacte de l'implémentation sonore à l'intérieur de ce code. (Roux-Girard 2009, 17)

En nous concentrant sur cette jouabilité, nous pouvons ainsi appréhender la construction sonore telle que reçue dans son entièreté par un joueur attentif. Celui-ci, à travers une écoute de l'intégralité des signaux sonores à sa disposition, peut démêler les informations implémentées à travers différentes strates sonores pour identifier une intention et un travail du concepteur.

Si la jouabilité devient le prisme à travers lequel nous allons étudier l'espace sonore et ses fonctions en jeu, il nous faut alors entrer, de manière simplifiée et minimaliste, dans la discipline du cognitivisme pour mieux comprendre les mécanismes d'intégration et d'apprentissage du joueur. Pour ce faire, il faut nous intéresser à nouveau au résumé effectué par Guillaume Roux-Girard des travaux d'Arsenault et Perron sur le « cycle magique ».

Arsenault et Perron définissent les jeux vidéo comme une « réaction en chaîne » dans laquelle « [l]e joueur n'agit pas autant qu'il réagit à ce que le jeu lui présente, et similairement, le jeu réagit à ces commandes » (2009, p. 119-120, notre traduction). En d'autres mots, le joueur répond à des événements qui ont été programmés par un concepteur (dont une partie du travail consistait à prédire les réactions du joueur aux dits événements) et réciproquement, le jeu agit en réponse aux manipulations du joueur avec de nouveaux événements programmés correspondant au nouvel état de jeu. (2009, 18)

Cette réciprocité est à la base du travail effectué non seulement sur la jouabilité, mais aussi sur le sonore tel qu'il est actuellement créé. De plus en plus, le joueur est amené à comprendre l'environnement dans lequel il évolue à travers une écoute attentive de celui-ci. Et plus l'écoute est attentive, plus elle permet d'identifier rapidement les signaux sonores qui fonctionnent en rapport avec une situation ou une source particulière. Cette capacité provient directement des spirales interconnectées du modèle du cycle magique d'Arsenault et Perron (fig. 03). Ces trois spirales, la spirale heuristique de la jouabilité, la spirale heuristique de la narration et la spirale herméneutique mettent en avant les capacités d'apprentissage du joueur à l'intérieur de l'expérience ludique dont il profite.

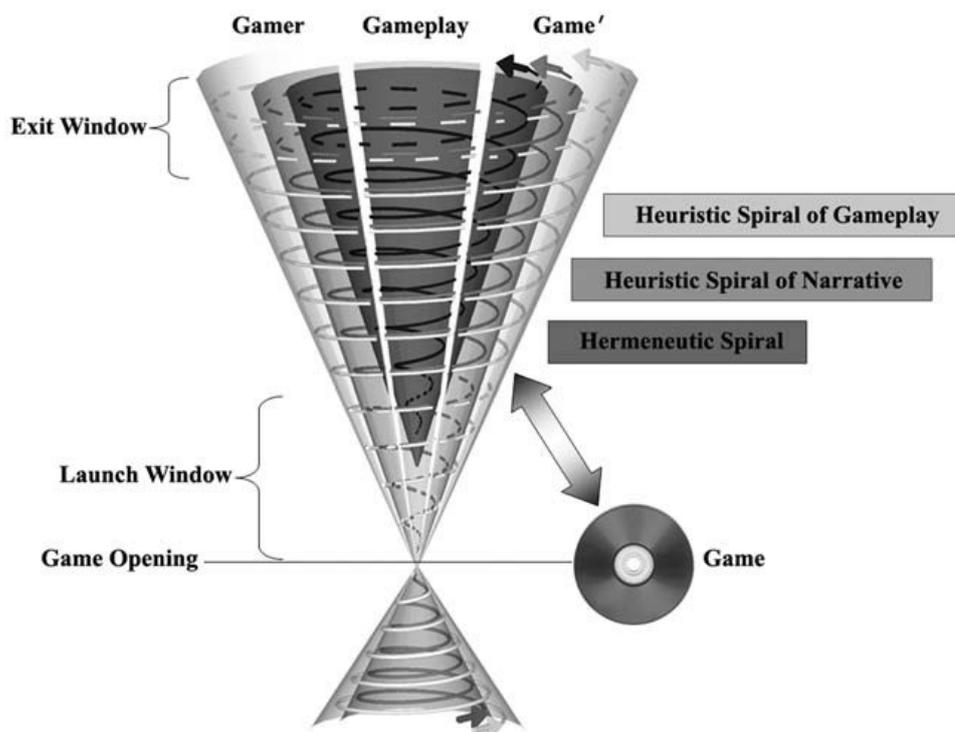


Figure 3. Schéma du cycle heuristique de jouabilité, 2009. ©Arsenault et Perron

Roux-Girard décrit et traduit les procédés cognitifs découlant de cette spirale ainsi :

Il faut préciser que « la relation qui les unit en est une d'inclusion : la jouabilité mène au déploiement de la narration et, ensemble, la jouabilité et la narration rendent possible une interprétation » (Arsenault et Perron 2009, p.118, notre traduction). Ce modèle offre l'avantage de considérer l'encyclopédie du joueur et son horizon d'attentes qui sont respectivement représentés par les lignes pointillées appelées « fenêtre d'entrée » (*launch window*) et par la spirale inversée. À partir du point d'entrée du joueur dans le jeu, la boucle de la jouabilité décrite préalablement sera « répétée un nombre incalculable de fois pour créer le cycle magique » (*ibid.* p. 121, notre traduction) et construire l'image mentale du jeu telle que développée par le joueur (représentée par le *Game'* du modèle). (Roux-Girard 2009, p.20)

Ajoutons à ce schéma un détail sur lequel nous reviendrons de nombreuses fois dans cette recherche. Si le cycle magique se constitue autour de la connaissance que le joueur construit du jeu, il nous semble toutefois que le joueur construit aussi son encyclopédie personnelle¹⁵ en fonction des jeux auxquels il a joué, et de son expérience vidéoludique globale. Roux-Girard le précise également : « le joueur structure les sons dans le jeu selon ses connaissances antérieures et son apprentissage » (2009, 21). L'apprentissage des situations qu'il rencontre dans un jeu peut provenir, comme nous allons le présenter avec l'étude des situations de conflit, de toutes les autres situations similaires rencontrées dans d'autres jeux. La connaissance se construit à l'intérieur du jeu ainsi qu'à l'extérieur, dans un environnement vidéoludique et médiatique plus large.

2.1.2 La hiérarchie des attitudes d'écoute

Intéressons-nous maintenant à la manière dont le joueur écoute l'environnement de jeu. En effet, nous pourrions alors constater que si le joueur effectue un tri des informations sonores, il ne semble pas le faire en fonction de la nature du son entendu : musique, sons et bruits sont traités simultanément, selon les informations de jouabilité qu'ils sont susceptibles

¹⁵ Nous employons « encyclopédie » dans le sens explicité par Umberto Eco dans *Lector in Fabula*. Pour Eco, le lecteur accumule au fil de ses lectures et de ses expériences un bagage de connaissances qu'il peut réutiliser ensuite pour mieux comprendre d'autres lectures. L'encyclopédie personnelle permet une réutilisation des connaissances et une interprétation plus rapide des nouvelles expériences.

de transmettre. Si nous avons évoqué précédemment les modèles d'écoute proposés par Michel Chion, les processus d'écoute dans le jeu semblent différents, puisque l'environnement sonore est inventé par des concepteurs qui imaginent quels bruits, quels sons fonctionneraient bien avec les personnages et l'univers créés par les développeurs : comme dans le cinéma d'animation, aucun son n'est issu directement de l'image construite.

Guillaume Roux-Girard, qui s'intéresse aux fonctions sonores dans le jeu vidéo d'horreur, met d'ailleurs en garde contre l'approche phénoménologique de Michel Chion, qui exclurait des modes d'écoute « en tant que signes d'expression, signes d'affects » (2009, 31). De plus, et comme Roux-Girard l'explique tout au long de sa recherche, le joueur est actif dans le jeu : ses actions vont avoir une influence sur la « partition sonore » (31) qu'il entend. Dans cette perspective, il y a une causalité directement induite entre le joueur et ce qu'il entend. L'écoute causale, nous explique Roux-Girard, est donc teintée par les connaissances du joueur, qui peut identifier un antagoniste s'il a déjà perçu le son qui le représente précédemment dans son expérience de jeu (31).

L'écoute causale va nous pousser à utiliser les sons entendus pour repérer les informations de l'espace de jeu. Michiel Kamp le souligne :

[W]e are normally more concerned with the source of the sounds we hear, and how it relates to our situation, than with the sounds themselves. Upon hearing the sound of a car when trying to cross a street, we do not stop to admire the slow crescendo, but identify it as approaching us, gauging its direction and its speed. (2013, 235).

Mais si Kamp argumente sur un phénomène d'écoute qu'on pourrait relier à l'écoute réduite de Schaeffer, nous allons étendre cette réflexion et formuler une nouvelle hypothèse : l'écoute causale prend le dessus en jeu vidéo, et le joueur va plutôt chercher à hiérarchiser

les informations données par l'ensemble des sons entendus, plutôt qu'à les trier en fonction de leur type.

Prenons pour exemple le jeu *Minecraft* (Markus Persson, puis Mojang, 2009). *Minecraft* est un jeu de survie qui s'est démarqué par une jouabilité dite « *sandbox* », qui laisse une grande part de liberté au joueur. L'avatar de celui-ci évolue dans un environnement gigantesque, cubique, qu'il peut modeler à sa convenance, équipé de quelques outils. En collectant des ressources pouvant être ensuite transformées en objets, ou en réorganisant et remodelant l'environnement, le joueur a la possibilité de créer un espace à son image, où les limites de création sont confinées à celle de son imagination, chaque objet et cube pouvant trouver du sens en fonction de l'usage souhaité par le joueur. Mais la conception d'un environnement unique n'est pas de tout repos : dans *Minecraft*, le joueur est son seul allié. Les créatures rencontrées sont très rarement amicales et attaquent plutôt à vue. La mort est punitive : le personnage jouable réapparaît à son *spawn*, un point préenregistré qui peut être bien éloigné du lieu de mort, tandis que le contenu de l'inventaire attend le personnage à l'endroit où le décès est survenu.

Comment prévenir le joueur du danger qui peut guetter son personnage? Tout d'abord à travers la représentation de l'environnement. Le joueur apprendra vite, souvent à ses dépens, que c'est l'obscurité qui fait apparaître les antagonistes (qu'il s'agisse de celle créée par le cycle jour/nuit ou l'absence de lumière à l'intérieur d'une grotte ou sous terre). Plus que l'espace de jeu lui-même, immense et regorgeant d'hors-champ et d'angles mort, c'est le sonore qui va permettre au joueur d'adapter son comportement. La musique, par exemple, n'est pas omniprésente : elle se déclenche au fil des explorations et des activités du joueur. La musique de *Minecraft* est caractérisée par des thèmes mélodiques lents, harmonieux,

souvent joués au piano et accompagnés par des cordes. Les thèmes sont assez longs, ils ont pour but d'accompagner le joueur dans son expérience de jeu sans entraîner de lassitude, car chaque progression est un pas de fourmi et les sessions de jeu peuvent être longues ou s'étaler sur plusieurs jours, semaines ou mois. L'ambiance ainsi créée fonctionne parfaitement avec l'exploration tranquille, la récolte et la construction caractérisées par des sons distincts omniprésents : bruits de récolte, d'outil qui récupère un bloc, d'expérience gagnée, etc. Mais lorsqu'un grain de sable, représenté par un *creeper*, une araignée ou un zombie, vient se glisser dans l'engrenage paisible de la production, l'espace sonore va jouer son rôle d'avertisseur. Tout d'abord à travers le silence : si la musique douce et mélodique était présente (ce qui est rarement le cas, puisqu'elle n'est pas continue), elle s'arrête assez brutalement, après un *decrecendo* court. L'environnement sonore à proximité de la menace est rempli par des bruits et des sons que le joueur, après quelques journées passées à guider son personnage dans le monde de *Minecraft* et grâce à l'écoute causale, va rapidement identifier comme appartenant à tel ou tel antagoniste. Le râle guttural est celui d'un zombie, le sifflement rapide est caractéristique d'une araignée, ou encore le bruit de la flèche tirée par un squelette. Les sons emplissent rapidement l'espace sonore, déjà saturé visuellement par les cubes qui composent l'environnement et qui réduisent le point de vue. Tous les bruits sont amplifiés, le niveau sonore est plus élevé¹⁶.

Les chercheurs Tuomas Eerola et Kai Tuuri vont proposer une nouvelle taxonomie des modes d'écoute en 2012. Celle-ci s'appuie sur la *cognition incarnée* (« *embodied cognition* », notre traduction) (Varela, Thompson et Rosch, 1991) et l'approche écologique des

¹⁶ On peut entendre une des transitions sonores de *Minecraft* dans la vidéo suivante : la musique est coupée par l'arrivée d'une flèche, puis, lorsque le joueur se rapproche de l'antagoniste, elle s'arrête. Surviennent alors des sons caractéristiques du conflit, dont le volume est plus élevé que celui des autres sons présents (eau, pioche, etc.) <https://youtu.be/HeU84dRmpZQ?t=120> (consulté le 23 août 2020).

perceptions. Cette taxonomie vise la détermination des modes d'écoute selon une hiérarchie cognitive. Les chercheurs étudient l'écoute dans le cadre des perceptions de l'auditeur par rapport au monde qui l'entoure, en s'appuyant notamment sur le lien kinesthésique de l'auditeur avec ledit environnement. Si nous ne partageons pas ici l'objectif de rendre compte de l'activité d'écoute de manière extrêmement précise, le fonctionnement de la taxonomie de Tuuri et Eerola nous paraît cohérent pour identifier les stratégies sonores mises en place dans un environnement médiatique : la situation de conflit vidéoludique.

En premier lieu, nous constatons que Tuuri et Eerola cherchent à comprendre le fonctionnement de l'écoute dans une situation où les sources sonores ne sont pas séparées : son, musique et bruit constituent ce qui peut être écouté. Le mode d'écoute ne dépend pas de la nature du son, mais plutôt des besoins de l'auditeur : « Listening is an active process, through which we gather information to meet our needs for interacting with the environment. » (2012, 137) En insistant sur l'aspect actif de l'écoute, les deux auteurs expliquent aussi que l'écoute se fait en continu, et que les modes d'écoute de leur taxonomie ne sont pas séparés. L'auditeur va changer de mode d'écoute selon les informations dont il a besoin, ce changement n'intervient pas selon la nature du son (138).

Pour les deux chercheurs, le travail cognitif s'arrime à l'environnement et aux possibilités d'action, qui forment un tout :

rather than perceiving structural features of sound first, we are naturally sensitive to action-relevant values of the environment. Our ecological knowledge of action-sound couplings provides a perceptual basis for various action-relevant meanings, such as sound sources, gestural signatures in actions (manifesting affect), and conditioned/learned associations. Perception and action are thus intertwined together since (1) the meanings of an environment are structured through embodied subject-environment interactions and (2) meaning-structures, such as action-sound couplings, are organized in terms of directly perceivable action-relevant values, i.e. *affordances* (Gibson, 1979; Varela et al., 1991). In other words,

our cognition, including sound-meaning structures, is integrally coupled with an ongoing world and embodied experiences of interaction. (2012, 138)

La taxonomie d'Eerola et Tuuri permet donc d'envisager l'environnement en fonction des actions motivées par la création de sens. L'auditeur, pour appréhender son environnement, se concentre sur les sons dont il a besoin rapidement pour traiter la situation de conflit.

Dans une situation de conflit vidéoludique, le joueur est concentré sur la résolution du défi. La victoire contre un antagoniste demande à la fois des réflexes kinesthésiques pour effectuer des actions (attaque, parade, etc.), et une appréhension rapide et claire de l'environnement de combat. Il est indispensable que le joueur parvienne à faire le tri dans toutes les informations sonores qu'il reçoit. À cet égard, son attitude d'écoute peut lui permettre, selon Eerola et Tuuri, de prioriser les sources sonores qui pourraient influencer ses actions. En jeu, il est souvent crucial de parvenir à identifier la bonne information au bon moment. Nous argumentons ici, en prenant appui sur le modèle de Karen Collins explicité ci-dessous, que le joueur effectue un tri dans les informations sonores en fonction de leur importance dans la situation de conflit qu'il rencontre, pour se concentrer sur celles qui peuvent l'alerter sur la nature d'un antagoniste ou sur le déclenchement d'une attaque.

2.1.3 Le classement de Collins

En ludomusicologie, chaque chercheur va proposer des stratégies différentes pour analyser la totalité des sons d'un jeu : identification des fonctions sonores, séparation des sons selon leur nature, importance vis-à-vis de la diégèse. Le modèle qui nous semble à la fois le plus simple et le plus efficace est celui de Karen Collins. L'auteure va en effet classer les sons

dynamiques en fonction de leur « degré d'interactivité avec le joueur » (2008, 125). Tout en soulignant la modularité de la spécificité de chaque son, qui peut passer d'un statut à un autre en fonction du contexte – et notamment de l'espace de jeu traversé par le personnage jouable – Collins présente les sons et leurs fonctions selon leur rapport au joueur et à la diégèse¹⁷. Nous allons passer à travers les différentes catégories utilisées par Collins, qui nous aideront à traduire la hiérarchisation des sons et des musiques entendus par le joueur.

La musique qui s'intègre le moins bien à l'espace de jeu est la musique de cinématique, considérée par Collins comme « linéaire, non-dynamique et non-diégétique » (2008, 126, notre traduction). Une cinématique va introduire, avec plus ou moins de narration, un élément important du jeu. Pendant celle-ci, le joueur n'a pas de contrôle sur son personnage, il est simple spectateur. Il paraît alors logique que les sons et la musique d'une cinématique ne soient pas réactifs. Ils s'apparentent à l'image cinématographique qui ne présente aucun caractère de jouabilité.

Les sons non diégétiques peuvent être adaptatifs et interactifs. Dans le premier cas, le son change en fonction d'un élément scripté sur lequel le joueur n'a aucun impact : la modification d'un thème musical dans un cycle jour/nuit, par exemple. Dans le second cas, c'est le déplacement ou l'action du joueur qui va créer une modification de ce qui est entendu : s'approcher d'un antagoniste va induire l'irruption de nouveaux signaux sonores.

Collins va ensuite classifier les sons diégétiques avec les mêmes étiquettes. On note ainsi que c'est bien le rapport à la jouabilité que Collins met en avant. Pour illustrer la présence

¹⁷ Le rapport à la diégèse est une question complexe dans le jeu vidéo : les événements qui s'y déroulent permettent à la fois de faire avancer l'histoire, mais aussi de donner des clés de jouabilité au joueur qui agit sur le déroulement de celle-ci. Kristine Jorgensen parle de « transdiégétique » (2009) pour définir l'état d'un son qui fait partie de l'environnement fictionnel tout en informant directement le joueur.

d'un son diégétique non-dynamique, l'auteure propose la recherche d'une station de radio par un personnage secondaire dans *Grim Fandango* (LucasArts, 1998) : « Manny (the player's character) has no contact with the radio: its sound is diegetic, but nondynamic » (2008, 126). Dans le cadre d'une situation de conflit, il s'agirait des sons caractérisant les différentes attaques d'un antagoniste, si elles sont générées de manière aléatoire et qu'elles ne dépendent pas de la position du personnage dans l'espace. Un son diégétique et adaptatif serait quant à lui constitutif de l'environnement sonore du jeu : il s'agit des modifications n'appartenant pas à la musique constituant le cycle jour/nuit d'un jeu, par exemple le bruit d'insectes différents selon l'heure de la journée dans *The Legend of Zelda : Breath of the Wild* (2017, Nintendo). Enfin, les sons interactifs diégétiques sont très présents dans l'espace sonore, et contribuent à la crédibilité de celui-ci. Ce sont les sons créés par les actions du joueur sur son personnage. Dans *The Legend of Zelda : Breath of the Wild*, il s'agit par exemple du bruit discret des pas lorsque Link marche, ou le halètement créé par sa course. Les deux possibilités sonores sont le résultat de l'action du joueur, qui décide de faire marcher ou courir son personnage.

Collins évoque également le cas des sons kinesthésiques, lorsque le joueur « participe à travers son corps aux sons à l'écran » en imitant un instrument à travers le dispositif de contrôle kinétique dont il dispose (2008, 127, notre traduction). Ceux-ci sont employés principalement dans des jeux ou des mini-jeux de rythme ou de danse, et ils ne sont pas très utilisés dans les situations de conflit. Nous ne nous y attarderons donc pas.

Nous serons amenés dans cette recherche à évoquer les sons dits *transdiégétiques*. Cette caractéristique développée par l'auteure Kristine Jørgensen rend compte de l'espace

entre le jeu et le joueur, spécifique au médium vidéoludique. Jørgensen définit le transdiégétique ainsi :

Computer games often questions the division between what should be understood as belonging to the virtual game world, and what should be seen as messages from the system or the interface. This can be explained by the presence of a player, who is allowed to step into the virtual game world while still remaining on the outside. This positioning enables the game to transcend information from the outside and into the virtual world, thereby creating a communicate [*sic*] channel with the functional purpose of easing the player's choices of actions in the game world. [...] *Transdiegetic space* will temporarily be understood as the conceptual space that comes into being when communication in a game questions the boundaries of the virtual world. (2009, 97).

Le transdiégétique est un état, ici appliqué à l'espace, qui justifie l'emploi d'un signal représentant quelque chose pour le joueur, mais qui soit également attribuable à un objet dans la diégèse. Habituellement, la diégèse compartimente l'action dans le monde imaginaire (inradiégétique) et l'information dirigée vers le spectateur qui ne fait pas directement partie dudit monde imaginaire (extradiégétique). Cette catégorisation binaire ne fonctionne pas dans beaucoup de jeux vidéo, où des informations extradiégétiques qui parviennent au joueur l'amènent à modifier le comportement de son personnage et ainsi les actions intradiégétiques. Par exemple, en entendant le déclenchement d'une musique de combat, le joueur peut comprendre qu'il s'approche d'un ennemi et ainsi dégainer une arme, préparer une attaque ou tenter de se camoufler.

L'opposition entre intra- et extradiégétique ne permet donc pas d'expliquer une écoute causale relative à la musique extradiégétique. Or, cette explication ne satisfait pas non plus Kristine Jørgensen en raison du joueur et de son rapport au personnage jouable¹⁸ : si ce

¹⁸ Jørgensen utilise le mot « avatar » pour parler du personnage jouable. En langue française, l'utilisation de ce mot implique une réflexion sur le rapport entre le jouer et le personnage à travers l'agentivité portée par ce dernier. Il ne s'agit pas du sujet de cette recherche, c'est pourquoi nous préférons employer le terme moins conceptuel de « personnage jouable ».

dernier fait partie de la diégèse, le joueur quant à lui n'y est pas intégré. Mais il pourra prendre des décisions vis-à-vis de son personnage jouable en fonction d'informations extradiégétiques dont il dispose. Ainsi, le caractère transdiégétique paraît pertinent pour expliquer cette dualité dans laquelle se trouve le joueur, entre son action réelle et la réaction de son personnage jouable dans la diégèse proposée¹⁹.

The player may use information available in extradiegetic sound when evaluating the avatar's possible actions, and he may thus influence the course of action in the game. [...] This leads to the interesting situation that although the game character does not hear extradiegetic sound due to its game-internal position, in effect it may react to extradiegetic sound because of the direct control link between player and avatar. In this respect, the game character can evaluate and act upon information that it should not be able to hear. (2009, 99).

À travers des degrés de jouabilité et des modifications scriptées ou non, les sons et la musique d'un jeu rendent compte d'une situation à laquelle le joueur fait face. À ce titre, la distinction précise entre musique, son et bruit devient moins importante pour le joueur qui doit hiérarchiser les informations qu'il reçoit dans le but de survivre : si la musique change brutalement, le joueur va l'utiliser pour comprendre la modification de son état de jeu, avant de se reconcentrer sur les sons diégétiques pour déterminer leur source. Son attitude d'écoute est modifiée par ses besoins selon l'action qui a lieu dans le jeu. Dans ce cadre, c'est le rapport des sons à la jouabilité qui va hiérarchiser l'information. Si un son entendu fait partie de la bande sonore et qu'il symbolise la présence d'un antagoniste, sa spatialisation va devenir l'information importante. Dans un contexte de combat, c'est plus directement le bruit

¹⁹ Notons cependant que le concept de transdiégétique est flou dans certaines situations et peut être questionné. Le jeu est créé par un concepteur pour un joueur. Les sons y sont donc implémentés pour être entendus par le joueur. Nous verrons ensuite que les fonctions du son ne limitent pas les possibilités de transmettre de l'information. Nous exposons ici le transdiégétique car il permet de comprendre une différence fondamentale entre le dispositif cinématographique et le dispositif vidéoludique, mais nous sommes conscients de ses limites.

caractérisant l'attaque ou le schéma d'attaque d'un antagoniste qui va nécessiter l'attention du joueur. Ainsi, les référentiels sonores bougent au fil de l'expérience vécue et des situations rencontrées. Ils constituent donc un ensemble qui contient des informations nécessitant un tri en fonction de la situation.

2.2 Appréhender le sonore

Dans notre exploration du sonore relié à l'espace de conflit vidéoludique, il était nécessaire de définir le son, ainsi que le rapport que le sonore entretient avec l'image. L'approche cognitiviste nous a semblé probante pour établir que l'auditeur-spectateur ou l'auditeur-joueur, à travers des processus d'écoute multiples fonctionnant en lien les uns avec les autres, reçoit des informations audio qu'il hiérarchise afin de se concentrer sur les points importants pour la situation vécue. Les sons seront privilégiés selon leur rapport à l'action entre l'auditeur et l'environnement.

2.2.1 La concrétion sonore

Nous définissons la situation de conflit vidéoludique ainsi : dans le cadre d'un jeu vidéo, le joueur est amené à rencontrer un ou plusieurs antagonistes. Ces antagonistes, expressément hostiles, nécessitent l'implication du personnage jouable dans un combat. La situation de conflit est une situation parmi d'autres dans le jeu vidéo. Elle intervient dans beaucoup de jeux où le joueur contrôle un personnage jouable. Elle ne constitue pas l'ensemble du jeu mais des moments particuliers, qui diffèrent d'autres situations jouables : une situation d'exploration, de construction ou d'infiltration par exemple. L'antagoniste se décline quant à lui selon plusieurs possibilités : il est un PNJ (un personnage non jouable) ou un personnage jouable contrôlé par un autre joueur dans le cadre d'un jeu multijoueur. En tant que PNJ, il peut également prendre deux formes principales : les antagonistes-poubelle

(« *trash mobs* », notre traduction) sont des antagonistes peu puissants qui ralentissent la progression du joueur sans vraiment la stopper. Ce sont des modèles d'ennemis génériques, qui peuvent intervenir à plusieurs reprises dans le même jeu. Un *boss*, en revanche, est un antagoniste puissant qui arrête complètement la progression du jeu. Le joueur n'a d'autre choix que de vaincre le *boss*, souvent après un nombre d'essais/erreurs conséquent puisque le *boss* emploie plus d'attaques qu'un antagoniste habituel, est souvent plus rapide ou possède bien plus de points de vie que le personnage jouable lui-même.

À la lumière des points évoqués précédemment dans cette recherche, nous croyons que les facteurs pour que le joueur identifie une situation de conflit sont multiples, et que certains indices proviennent du sonore. Le jeu vidéo étant un média fondamentalement audio-visuel, qui fonctionne sur un principe de jouabilité entre le joueur et le jeu, il est cohérent de penser que l'identification d'une situation particulière doit se faire le plus rapidement possible, afin que le joueur puisse y répondre adéquatement. Nous évoquerons ici la construction de la situation de conflit à travers la formation de ce que nous appellerons une concrétion sonore, concept que nous avançons maintenant.

Parmi les aspects évoqués dans les sous-parties précédentes, nous retenons deux idées principales. La première, c'est que l'auditeur d'un environnement audio-visuel cherche à créer du sens dans l'expérience qui lui est proposée. Il va spontanément relier ce qu'il entend à ce qu'il voit, ce qui permet de créer un sentiment de cohérence pour adhérer à la proposition audio-visuelle. Michel Chion d'abord, puis Annabel J. Cohen s'accordent sur l'attribution de fonctions sonores qui semblent guider l'écoute de l'auditeur-spectateur. Nous pouvons séparer ces fonctions en deux grandes catégories : des fonctions ayant pour but de diriger l'attention de l'auditeur, et d'autres permettant plutôt une cohérence narrative et

émotionnelle au sein de l'expérience proposée. L'auditeur-spectateur est en mesure, grâce à ce qu'il entend, de comprendre l'expérience audio-visuelle et d'en tirer une cohérence narrative. Dans le jeu vidéo, l'auditeur-joueur a en plus un but à accomplir dans son expérience de jeu. La résolution du défi passe par l'appréhension, puis la compréhension de la situation de jeu dans laquelle il se trouve. Le son du jeu vidéo, à l'instar de toute expérience audio-visuelle, est constitué d'une multitude de sources sonores ayant des formes très différentes. Des bruits de pas peuvent se superposer à une musique d'ambiance extradiégétique, les rôles d'un antagoniste peuvent indiquer sa position dans l'espace, tandis que des sons d'interface peuvent apporter au joueur des informations sur le ramassage d'un objet ou des points de vie faibles.

De cette multitude de sources sonores découle la deuxième idée développée dans cette recherche: l'auditeur ne se préoccupe pas vraiment de la nature de la source audio, mais plutôt de son rapport à la situation audio-visuelle. Son écoute est dirigée en fonction de ses besoins. Eerola et Tuuri mettent clairement en avant cette idée en déterminant que l'écoute est reliée à l'action de l'auditeur dans son environnement. Nous pensons que cette idée est entièrement transposable au jeu vidéo, où l'action du personnage-jouable dans l'environnement de jeu va dépendre du joueur et de ses perceptions. Le joueur qui crée du sens et de la structure dans son expérience de jeu va se concentrer sur les indices sonores et visuels lui permettant de le faire. À cet égard, nous croyons qu'il est plus pertinent de parler d'une hiérarchisation des sources sonores en fonction du lien à l'action, que de séparer les différentes sources selon leur nature. Les caractéristiques mises en avant tout au long de cette recherche montrent clairement que l'écoute est un processus mouvant, et que l'auditeur oriente son écoute selon ses besoins. La nature du son entendu n'intervient pas autant dans le processus d'écoute que son lien avec l'environnement. Dans le jeu vidéo, et d'autant plus

dans une situation de conflit, le joueur est amené à réagir rapidement face à la situation dans laquelle son personnage jouable se trouve. En écoutant et en hiérarchisant les différentes sources sonores, le joueur établit des priorités, de manière plus ou moins consciente, pour traiter la situation le plus efficacement possible. Bien entendu, le traitement sonore fait en amont par les concepteurs du jeu facilite ou entrave de manière volontaire la compréhension, comme nous le verrons plus tard. Il nous semble important de retenir ici que le sonore informe le joueur sur la situation de jeu, et que le joueur peut ensuite grâce à ses connaissances préalables s'intéresser principalement aux signaux sonores nécessitant son attention rapidement.

Parce que le sonore est traité dans son ensemble par l'auditeur, nous croyons qu'il est intéressant de considérer le son de la situation de conflit comme une *concrétion*. En géologie, la concrétion est un « Amas de particules solides se trouvant dans les roches ou les sols et résultant de la formation successive et de l'agglomération de particules nouvelles sous l'action d'agents physiques ou chimiques. »²⁰ L'analogie avec le fonctionnement du système sonore nous paraît pertinente dans la mesure où les sources sonores, de natures diverses, se réunissent pour former un tout cohérent. Si une roche est constituée de couches sédimentaires superposées, la concrétion sonore fonctionne de la même manière, avec des sources sonores qui s'agglomèrent autour de la situation de jeu²¹. Chaque strate possède des particularités différentes, des fonctions différentes, mais lorsque le joueur écoute, c'est l'intégralité de la concrétion qu'il perçoit. Il peut ensuite, grâce à son écoute dirigée vers un

²⁰ <https://www.cnrtl.fr/definition/concr%C3%A9tion//0>, consulté le 15 juin 2020

²¹ Cette notion est introduite, créée, conceptualisée au fil de la réflexion autour de cette recherche. Les prolégomènes les plus importants sont posés ici, toutefois nous sommes conscients que cette conceptualisation demande peaufinage, précision et réflexion pour évoluer en un concept plus complet. Ainsi, le choix de la métaphore géologique, si elle nous semble attrayante en raison des liens qu'elle partage avec la matière sonore (on parle de musique concrète, de morphologie du son, etc.), gagnerait à être mise à l'épreuve dans de futurs travaux de recherche.

but, accorder plus d'importance à une strate sonore ou à une autre en fonction de la situation de jeu à laquelle il est confronté. Nous n'accordons pas la même importance à toutes les sources sonores, mais nous les entendons toutes puisqu'il nous est impossible de les séparer totalement. Nous pouvons nous concentrer sur des sons qui paraissent prioritaires, mais nous entendrons toujours, en même temps, les autres strates sonores qui constituent la concrétion. Dans ce cadre, le travail d'écoute ne se fait pas par soustraction mais par hiérarchisation : il ne s'agit pas de supprimer certaines sources sonores, mais bien de leur accorder des priorités d'attention.

Dans une situation de conflit, le joueur peut entendre à la fois la musique d'ambiance, le placement de l'antagoniste par rapport à son personnage-jouable, le bruit de différentes attaques, des signaux sonores relatifs à l'interface de jeu ou encore les bruits issus du personnage-jouable lui-même. Tous ces sons proviennent de différentes sources sonores, qui constituent un ensemble, une concrétion dans laquelle plusieurs strates d'informations cohabitent. Dans la suite de ce mémoire, nous nous attarderons sur les différentes strates existantes et voir de quelle manière celles-ci se superposent, afin d'appréhender le sonore de la situation de conflit dans son ensemble, sans le dénaturer en lui attribuant des séparations qui, si elles sont structurelles – les différentes strates étant assemblées, puis superposées lors de la conception du jeu – ne sont pas directement perceptibles par l'auditeur-joueur.

2.2.2 Mise en situation de la hiérarchisation sonore : une étude de cas avec *Assassin's Creed : Brotherhood*

Plaçons-nous maintenant en situation de jeu. Comme explicité avec le modèle proposé par Collins, il nous semble que la hiérarchie des informations sonores se fait en fonction d'un lien à la jouabilité, qui constituerait selon nous l'équivalent du lien à l'action proposé par

Eerola et Tuuri. Le joueur qui fait évoluer son personnage va être attentif à son environnement selon des degrés de jouabilité, et la priorisation des sons se fera en conséquence. Prenons l'exemple d'un jeu où le joueur entend les pas du personnage qu'il dirige. En tout logique, le joueur comprend très rapidement que le bruit des pas est lié à l'action de marche de son personnage. Il peut donc se désintéresser de ce son et concentrer son écoute sur les autres couches sonores de la concrétion. Dans celles-ci, nous comptons par exemple : une musique environnementale, les différents sons qui composent l'environnement (qui sont donc diégétiques) et plusieurs niveaux d'alertes sonores qui vont, selon nous, conditionner l'attention et le ressenti du joueur. À travers l'analyse de plusieurs de ces strates, nous allons nous pencher sur les différents sons et tenter de mieux percevoir leurs fonctions par rapport à la situation de jeu.

À travers l'examen du jeu *Hitman : Contracts* (2004, Eidos Interactive), Kristine Jørgensen – qui s'intéresse directement à la manière dont l'audio peut influencer les actions du joueur – va traiter le cas d'un jeu solo guidé par une narration scénaristique explicite, où le personnage jouable évolue dans un environnement 3D dont une des mécaniques principales de jeu est l'infiltration. En s'inspirant de l'analyse de Jørgensen, nous nous pencherons sur le jeu *Assassin's Creed : Brotherhood* (2011, Ubisoft Montréal) car les deux jeux possèdent des points communs de design qui nous paraissent cohérents. Le scénario de la trilogie *Assassin's Creed II* se concentre sur l'histoire d'Ezio Auditore, un jeune nanti qui évolue à Florence durant la Renaissance italienne, qui se retrouve dans un conflit armé entre les Assassins (dont il fait partie) et les Templiers. *Assassin's Creed : Brotherhood* est le deuxième opus d'une série de trois jeux qui ont les mêmes protagonistes. Au début du jeu, Ezio rend visite à son oncle Mario qui habite la petite ville de Monteriggioni. Nous allons voir qu'au fil du scénario, l'ambiance

sonore est totalement modifiée et contribue à générer un sentiment d'apaisement ou d'urgence pour le joueur.

Ezio commence par déambuler tranquillement dans la ville de son oncle, Monteriggioni. Sur le plan visuel, la lumière est celle d'un coucher de soleil, et le personnage jouable arrive de l'extérieur de la ville, juché sur son cheval²². Sur le plan sonore, peu de bruits sont audibles : on écoute surtout le dialogue entre Ezio et son oncle qui l'accompagne, les acclamations des villageois visiblement heureux de retrouver Ezio, et en fond, une musique très calme, très simple²³, composée par Jesper Kyd. Le seul autre bruit présent est celui des sabots des chevaux ou des pas d'Ezio. Le jeu est ensuite entrecoupé de séquences narratives sur lesquelles nous ne nous attarderons pas : elles contribuent à donner l'ambiance du jeu et elles apportent surtout des informations narratives à travers des dialogues. Elles forment d'ailleurs la transition entre les deux états auxquels nous nous intéressons ici. Le joueur est ensuite invité, à travers l'interface, à aider plusieurs habitants de la ville. C'est l'occasion de noter la présence de sons d'interfaces, qui sont donc extradiégétiques et qui informent directement le joueur. La découverte de Monteriggioni s'effectue tranquillement.

Le joueur, qui vient de commencer le jeu, est accueilli par un environnement chaleureux, rendu tel à la fois par le visuel et par une concrétion sonore fluide, qui n'est pas saturée d'informations et qui invite au calme. Dans une des missions à effectuer, Ezio doit cependant rattraper un cheval qui s'est enfui²⁴. La musique change du tout au tout et devient un thème rapide, entraînant, mais qui reste dans une tonalité majeure. Elle ressemble à une

²² Pour illustrer ce segment, il est possible de consulter une partie enregistrée par un joueur et diffusée sur Youtube : <https://youtu.be/ZvWzk4MBXpY?t=732> (consulté le 20 août 2020).

²³ Le joueur du précédent opus, qui aura pu convoquer son encyclopédie personnelle, pourra reconnaître cette musique comme le « Monteriggioni theme » puisqu'elle est déjà présente dans *Assassin's Creed II*.

²⁴ Une séquence que l'on retrouve à 16'23'' : <https://youtu.be/ZvWzk4MBXpY?t=983> (consulté le 20 août 2020).

danse, fluide et enlevée. Dans le style, on peut d'ailleurs la rapprocher des danses traditionnelles de la Renaissance : une guitare classique et un violon avec un thème très simple. Si l'impression de vitesse peut être ressentie par le joueur à travers cette musique, on ne constate aucun caractère anxiogène. Les rares bruits étant identifiés très rapidement par le joueur comme appartenant à l'environnement tranquille avec lequel il est familier, il n'a aucune raison d'être sur ses gardes.

Lors d'une autre quête, Ezio doit manipuler les canons de la ville²⁵. Immédiatement, l'espace sonore change : une musique rythmée, guerrière, avec une omniprésence de percussions et un thème grave joué par des cordes retentit. Ce thème est ponctué par le bruit des canons qui ont un volume élevé : ils couvrent totalement la musique. Non seulement l'ajustement du volume des différentes strates sonores va ajouter à la crédibilité de la scène, mais il est également un indicateur de la hiérarchisation adressée au joueur : il est plus important pour lui d'entendre que l'action de tirer au canon est effectuée (grâce au bruit des canons) que d'entendre le thème musical.

L'ambiance tranquille de Monteriggioni est toutefois complètement modifiée au petit matin suivant. Après une cinématique où Ezio est réveillé en sursaut par des canons et voit la ville assiégée par l'ennemi, le joueur reprend le contrôle du personnage²⁶. Cette fois-ci, la lumière est nocturne. Les nuages du ciel sont très bas et des explosions retentissent sur le chemin du personnage. C'est l'occasion de constater que la modification de l'espace sonore se fait rarement seule, l'espace visuel est lui aussi important pour apporter un ressenti. Nous nous appuyons toutefois sur l'analyse de Michel Chion pour affirmer que le sonore apporte de la cohérence à l'image. Tout d'abord, le volume de la scène de jeu est globalement

²⁵ <https://youtu.be/ZvWzk4MBXpY?t=1239> (consulté le 20 août 2020).

²⁶ <https://youtu.be/ZvWzk4MBXpY?t=1708> (consulté le 20 août 2020).

augmenté. La musique qui retentit est proche de celle de la quête de manipulation des canons : très rythmique, percussive avec un thème grave. On constate ensuite une saturation de la concrétion sonore : la musique est plus forte, mais les bruits sont désormais omniprésents. L'atmosphère est remplie par les cris des villageois qui résonnent même hors-champ. Le cheval est lancé au galop et le bruit de ses sabots qui claquent sur les pavés résonne sur la même fréquence que l'accompagnement rythmique des cordes graves (contrebasses et violoncelles). De plus, on entend très régulièrement des explosions qui se mêlent à la musique extradiégétique : impossible de séparer vraiment les deux. La hiérarchisation des informations sonores est bien plus complexe, le joueur étant noyé dans une quantité très importante d'informations. Toutefois, le sens général de l'espace est clair : il faut fuir.

Les rues qui paraissaient si larges sont désormais étroites et jonchées de gravats. Le son amplifie le sentiment de désordre, et crée celui de l'urgence. Mais c'est bien l'ensemble de la concrétion sonore qui est modifiée pour générer un ressenti émotionnel, pas uniquement la musique. Et cette fois-ci, lorsqu'Ezio retourne aux canons de la ville et doit à nouveau tirer à l'extérieur, le son rajoute du stress²⁷. Techniquement, la situation de jeu est strictement la même : le joueur tire sur des cibles immobiles à l'extérieur de la ville. Cependant, les sons des canons ennemis sont désormais omniprésents et donnent l'impression que les forces militaires des assaillants sont bien plus rapides que celles que le joueur contrôle. Dans une concrétion sonore saturée, les coups de canon déclenchés par le joueur résonnent à peine aux oreilles. L'information donnée au joueur n'est plus aussi claire, et c'est voulu : la confusion créée par l'amplification sonore contribue à la fois au sentiment

²⁷ <https://youtu.be/ZvWzk4MBXpY?t=1823> (consulté le 20 août 2020).

d'urgence et à l'impuissance ressentie par le joueur, qui contrôle certes un héros, mais un héros bien seul face à une armée.

Le joueur est ensuite plongé dans une scène de bataille au corps à corps²⁸. C'est l'occasion de constater encore une fois d'importants changements dans la concrétion sonore. À partir du moment où le joueur effectue une action de combat, le volume sonore diminue brusquement. Tout l'arrière-plan sonore, les bruits d'explosion et la musique, restent présents mais sont moins audibles. En revanche, on entend très distinctement, au début de la situation de conflit, le bruit des épées sorties de leur fourreau. La hiérarchisation des sons est réeffectuée afin d'apporter une information directe au joueur. La concentration reprend le dessus sur la panique et la désorganisation. Et lorsqu'Ezio est cerné par des assaillants²⁹, les bruits qui résonnent sont de trois sortes. On entend tout d'abord le bruit d'une épée contre de la ferraille, qui informe toutefois le joueur qu'il a bel et bien touché une cible. Il faut dire que dans le tourbillon visuel du combat, et en raison du placement automatique de la caméra du joueur (à la troisième personne), l'angle de vue ne permet pas toujours de savoir si l'ennemi est touché – il y a tout de même une répétition de cette information au niveau du visuel avec une barre de vie des ennemis qui diminue. Ensuite, le deuxième son qui est perçu rapidement par le joueur est le cri poussé par un assaillant, soit lorsque ce dernier va attaquer, soit lorsqu'il meurt. Cet indice sonore va permettre au joueur de comprendre en une fraction de seconde qu'il doit s'intéresser à, ou qu'il peut se désintéresser d'un antagoniste. Enfin, le reste de la concrétion sonore reste présent avec une musique très rythmée qui apporte à la fois induction de ressenti émotionnel et communication du sens de la situation de jeu.

²⁸ <https://youtu.be/ZvWzk4MBXpY?t=1978> (consulté le 20 août 2020).

²⁹ <https://youtu.be/ZvWzk4MBXpY?t=2003> (consulté le 20 août 2020).

Ce tutoriel dans Monteriggioni, qui permet de circonscrire sur une courte durée l'ensemble des situations principales présentes dans *Assassin's Creed: Brotherhood* en présentant au joueur un rappel sur la jouabilité des précédents opus, nous permet de voir le fonctionnement du son dans le jeu vidéo et d'illustrer en quoi la concrétion sonore est une manière efficace de rendre compte de l'activité d'audition du joueur, tournée vers l'action et engagée dans une activité de hiérarchisation des sons. On y a vu aussi l'importance de l'induction d'un ressenti, une fonction présentée par Cohen dans le cadre cinématographique et traitée par Collins, Kamp, Jørgensen ou Roux-Girard dans un contexte vidéoludique. L'implémentation d'une expérience crédible passe par un apport émotionnel auquel le son contribue. Les strates sonores véhiculent à la fois des fonctions permettant une meilleure jouabilité et des informations narratives pour une expérience audio-visuelle engageante et crédible. Cet exemple nous permet de constater aussi que la construction sonore est à la fois plurielle et complexe. Grâce aux indices établis par les concepteurs, le joueur peut traiter les informations dont il a besoin pour une bonne jouabilité et concentrer son attention sur telle ou telle strate sonore, sans pour autant omettre le reste des signaux qui sont importants. En ce sens, le prochain chapitre nous permettra de réfléchir aux diverses stratégies déployées par les concepteurs, dans différents types de jeux, pour structurer ou guider l'expérience audio-visuelle du joueur.

Chapitre 3 – La concrétion sonore dans la situation de conflit

Concentrons-nous désormais sur l'étude du sonore dans la situation de conflit, afin de comprendre comment la stratification de la concrétion sonore peut fonctionner dans ce cadre. La situation de conflit vidéoludique demande de nombreuses compétences de la part du joueur, qui doit résoudre rapidement un défi faisant appel à ses réflexes, ses capacités d'analyse, son endurance et sa dextérité. Pour permettre la résolution de ce défi, les concepteurs placent dans la situation de conflit de nombreux indices, tout aussi visuels que sonores. Dans ce dernier chapitre, nous allons voir quels sont les indices sonores à la portée du joueur pour qu'il puisse analyser correctement la situation à laquelle son personnage jouable est confronté. À travers des exemples issus de jeux très différents, nous mettrons également en avant l'homogénéité de la concrétion sonore dans des situations de conflit diversifiées. Il nous faudra s'attarder sur les indices qui permettent de comprendre rapidement dans *Assassin's Creed* que nous passons d'une situation d'exploration à une situation de conflit. Nous verrons avec *Overwatch* que des stratégies de spatialisation sonores peuvent être employées pour préciser ladite situation de conflit. À l'aide de *Minecraft*, de *The Legend of Zelda : Breath of the Wild* ou encore de *The Elder Scrolls V : Skyrim*, nous nous pencherons sur les moyens mis en place par les concepteurs pour rendre un joueur plus attentif, pour transformer son ressenti émotionnel. Enfin, nous nous jetterons au cœur du combat pour identifier des aides, des contraintes ou des pièges qui aident ou qui brouillent volontairement la compréhension du joueur, à travers des exemples tels que *Crypt of the Necrodancer*, *Dark Souls* ou encore *World of Warcraft*.

3.1 Implémentation de la situation de conflit

La situation de conflit intervient entre d'autres situations de jeu à l'intérieur d'une expérience vidéoludique. De fait, on peut parfois distinguer la situation de conflit à travers le

sonore qui y est rattaché plus que grâce au visuel. Il est rare que le visuel change du tout au tout pour prévenir d'un danger, tandis que le sonore, lui, peut être totalement et subitement transformé. Nous avons commencé à le voir dans notre exemple d'*Assassin's Creed : Brotherhood* : les changements d'ambiance dirigent les possibilités de jeu et peuvent prévenir d'une éventuelle agression envers le personnage jouable. Nous nous concentrerons ici sur la transition qui a lieu d'une situation de jeu vers une situation de conflit. À travers plusieurs exemples, nous verrons que les sources sonores peuvent diriger l'attention du joueur vers un conflit potentiel, quelles que soient leur nature. Nous nous intéresserons ensuite à un aspect du son qui est primordial dans de nombreuses situations de conflit, et que nous avons laissé de côté pour le moment : dans les jeux en trois dimensions, la spatialisation du son peut permettre au joueur de localiser avec efficacité et rapidité la direction de l'agression qu'il subit ou va subir. Enfin, nous nous intéresserons plus particulièrement aux strates sonores incitant le joueur à diriger son attention rapidement. Ces sources, qui fonctionnent comme des alertes, sont des indices cruciaux pour la jouabilité de la situation de conflit.

3.1.1 Les transitions sonores et la continuité de l'expérience

Dans le cadre de la situation de conflit, nous allons particulièrement nous intéresser à la transition entre un ressenti et un autre. En effet, c'est à partir de cette transition que le comportement du joueur est modifié et qu'il se prépare à combattre un antagoniste. Ajoutons toutefois directement une nuance : dans de nombreux jeux, l'espace sonore n'est pas particulièrement modifié par l'apparition d'un antagoniste classique, un antagoniste-poubelle (*trash mob*) qui apparaît plusieurs fois dans le jeu. La création d'une transition sonore n'est pas non plus déclenchée uniquement par l'apparition d'un antagoniste, elle peut être créée en raison d'un changement d'environnement de jeu (entrée dans un donjon, dans un nouveau

niveau, dans un nouveau biome). À ce titre, de nombreux ludomusicologues s'intéressent au « town theme »³⁰, un thème musical précis relié à une ville traversé par le joueur, qui active les fonctions de rappel mémoriel et d'induction de ressenti présentés par Cohen. Les jeux de la franchise *Assassin's Creed* (Ubisoft) présentent d'intéressants changements d'ambiance sonore, directement liés à un changement de paradigme de jouabilité.

Assassin's Creed se veut une alternance entre phases de combat et phases d'infiltration. Sans changer spécialement d'environnement, le joueur est amené à être discret ou plus conquérant en fonction de la situation dans laquelle se trouve son personnage. Si le jeu ne se veut pas pacifique, le joueur est toutefois très encouragé au silence et à la furtivité, à la fois par le scénario (à travers des objectifs secondaires demandant expressément de « Ne pas se faire repérer » dans la plupart des quêtes) et par l'environnement dans lequel le personnage jouable évolue, qui propose des chemins alternatifs à celui sur lequel se déplacent les antagonistes (toits, escalade, hautes herbes, charrettes de foin, etc.). Les antagonistes d'*Assassin's Creed* attaquent très majoritairement en groupe, ce qui peut rapidement submerger le joueur. Pour toutes ces raisons, une approche discrète est privilégiée pour la plupart des missions proposées par le jeu. Mais pour ajouter au défi du joueur, les antagonistes sont à la recherche du personnage principal – ou en tout cas, de tout comportement étrange, comme grimper sur les façades des bâtiments, dans leur périmètre de surveillance. Et lorsque les antagonistes repèrent le personnage jouable, ils l'attaquent à vue. À partir de ce moment-là, le joueur a deux choix : combattre des groupes d'antagonistes

³⁰ La récente thèse de Gregory James Rossetti, intitulée *Overworlds, towns, and battles : how music develops the worlds of role-playing video games* met en avant la construction de sens inhérente aux thèmes musicaux reliés à des lieux ou des situations de jouabilité.

ou fuir. En effet, les assaillants ne cesseront leur traque que si le personnage reste caché pendant un certain temps.

Il est très intéressant de constater dans les jeux *Assassin's Creed* que l'alternance entre la situation de furtivité et la situation où le personnage est repéré se fait grâce à une alarme extradiégétique qui retentit et vient troubler le calme de la phase de discrétion³¹. Ce son d'alerte a pour seul but d'avertir le joueur. Il est totalement déconnecté de l'espace sonore et possède une hiérarchie élevée par rapport aux autres sons, ce qui permet au joueur de comprendre qu'on s'adresse directement à lui et lui donne la possibilité de rediriger immédiatement son attention vers une agression qu'il sait imminente. Ce son est précédé ou suivi, selon la situation, d'une alerte diégétique : le cri des assaillants qui s'adressent au personnage par un « Hey! You! » (*Assassin's Creed Origins*, 2017, Ubisoft Montréal)³². La transition entre le personnage furtif et celui découvert passe essentiellement à travers l'espace sonore : l'espace de jeu est très peu modifié, on note juste l'apparition d'une interface signalant la direction des antagonistes. L'état du joueur est cependant totalement transformé, puisqu'il passe d'une approche lente et discrète à un état plus vif, celui de la fuite ou du combat. On note qu'à l'instar de l'exemple de Monteriggioni, la concrétion sonore devient plus complexe lorsque la situation de conflit se déclenche : à la musique s'ajoutent les voix des soldats qui cherchent le personnage et le bruit des armes si ce dernier est en combat. La transition sonore est ici fluide, efficace et rapide.

³¹ Ce système d'alerte sonore a été rendu populaire par le célèbre « ! » de *Metal Gear Solid*, un des jeux ayant servi de modèle pour le type de jouabilité de furtivité (*stealth game*) : <https://medium.com/the-space-ape-games-experience/-423c6bcf9336>, consulté en ligne le 17 août 2020.

³² <https://youtu.be/B2Bi7KdGpnU?t=4098> (consulté le 20 août 2020).

La mixité des informations sonores est un point mis en avant par Jørgensen dans son analyse d'*Hitman : Contracts* (2004, Eidos Interactive). Pour l'auteure, la musique ne donne pas seulement une amplification dramatique à la scène, mais elle est aussi un informateur important :

When guards direct negative attention towards the avatar, the music will change into a higher-paced up-beat melody. This music will continue as long as the situation is tense or the combat is going on. What is interesting with this music, is that the melody changes according to how the player is doing in the specific situation. In this respect, the music works to give the player feedback on the combat status. (2009, 90).

Selon la classification employée par Collins, la musique devient alors à la fois interactive et adaptative. Elle permet d'obtenir une information précise tout en n'étant pas directement reliée à la diégèse du jeu. Son intégration dans le reste de la concrétion sonore va permettre une redondance des informations, et le joueur pourra savoir très rapidement si son personnage est détecté ou non par les gardes.

À l'instar d'*Hollow Knight* (2017, TeamCherry), la transition sonore peut s'effectuer en fonction de l'environnement et non précisément de l'antagoniste. Elle passe parfois même par un grand oublié de la concrétion sonore : le silence. Ce jeu de plateformes 2D fonctionne comme un *metroidvania*, c'est-à-dire que le joueur se fraie un passage dans un environnement labyrinthique où des chemins se débloquent au fil de sa progression et des améliorations de son personnage. La particularité d'*Hollow Knight* réside dans sa difficulté. Tout au long du jeu, le joueur voit son personnage confronté à des *boss* particulièrement acharnés. Le personnage jouable évolue dans un monde souterrain où jardins enchanteurs succèdent à nids d'araignées et repaires de méduses. Mais ces différentes zones ne sont pas séparées par des niveaux : le joueur va simplement observer un changement majeur dans la direction artistique de l'environnement visité, qu'il s'agisse de modifications visuelles ou sonores.

L'accompagnement musical d'*Hollow Knight* est extrêmement riche. Le compositeur Christopher Larkin emprunte beaucoup à la composition orchestrale et propose des morceaux très divers, où l'on peut constater les influences de la musique classique, baroque et romantique. La bande musicale est agréable à écouter, elle rythme à la fois les combats et les espaces : selon la difficulté de l'environnement traversé, on passera d'une musique douce, calme, jouée à la harpe ou au piano³³, à des morceaux plus rythmiques³⁴, avec la présence de percussions et de thèmes moins mélodiques, auxquels s'ajoutent les bruits qui émanent des antagonistes rencontrés. Toutefois, certains espaces sont exempts de tout accompagnement musical. Ils interviennent à plusieurs reprises dans le jeu et suscitent la méfiance chez le joueur habitué à se promener sur un fond musical riche. Lorsque l'espace sonore est silencieux, l'espace de jeu est lui aussi plus inquiétant : il est moins coloré, plus restreint, plus agressif également (pics, bestioles invincibles présentes au premier plan, etc.). La concrétion sonore est ensuite construite de manière minimaliste, par des bruits qui viennent amplifier le contraste entre la présence ou l'absence de musique³⁵. On peut entendre le vent qui souffle, les grignotements de ce qu'on devine être des insectes peu sympathiques, ou encore le « plic ploc » d'une goutte d'eau qui s'écrase sur les parois d'une caverne. Dans le contraste entre la complexité d'une concrétion sonore riche, et la solitude de quelques strates vidées de tout accompagnement agréable, la transition appelle à la vigilance. Lorsque l'on entend moins de choses, tout détail sonore devient suspect – et indique potentiellement une menace ou une agression dans un environnement qui n'était déjà pas très amical.

³³ https://youtu.be/e_ijd5g9q_w?t=8452 (consulté le 20 août 2020).

³⁴ https://youtu.be/e_ijd5g9q_w?t=11312 (consulté le 20 août 2020).

³⁵ https://youtu.be/e_ijd5g9q_w?t=5623 (consulté le 20 août 2020).

En comparaison, nous allons voir que certaines transitions sonores sont moins bien intégrées à l'expérience de jeu. Parfois même, elles peuvent sortir le joueur de son appréciation de l'expérience en étant trop brutales ou disproportionnées. Ceci peut constituer une menace pour la continuité de l'expérience, qui est créée par une cohérence globale entre les différentes phases de jeu. Michel Chion évoque l'importance du sonore pour la cohérence cinématographique entre les plans, et nous émettons l'hypothèse que la cohérence du jeu est assurée en partie par la cohérence de la concrétion sonore lorsque les paradigmes de jouabilité sont modifiés. Prenons pour exemple les transitions sonores vers une situation de conflit dans le jeu *The Elder Scrolls V : Skyrim* (2011, Bethesda Softworks).

Skyrim est un RPG complexe où de nombreuses factions, guildes, et entités politiques parsèment le chemin de l'Enfant de Dragon. À cet égard, les antagonistes prennent aussi des formes très diverses : des loups aux mammoths, des vampires aux chevaliers, ou même les gardes des différentes villes visitées si le joueur contrevient aux bonnes manières dans celles-ci. Un moyen sûr de repérer un antagoniste est de vérifier son statut sur la barre horizontale servant de boussole, en haut de l'interface : si quelqu'un est agressif, il est symbolisé par un point rouge (alors que les PNJ neutres ne sont pas symbolisés du tout). Sur le plan sonore, l'affaire n'est pas plus compliquée : lorsqu'un danger menace, une musique d'avertissement retentit, avec un thème court, incisif, joué aux cordes et intégrant une rythmique grave, percussive et rapide³⁶. Le thème n'est pas vraiment mélodieux, surtout en comparaison du reste de la bande sonore du jeu, extrêmement riche et variée. C'est cette musique qui conditionne dans son entièreté la situation de conflit. Les autres indices sonores sont parfois

³⁶ Il y a plusieurs musiques de combat dans *Skyrim*, en voici un exemple : <https://youtu.be/nr62-GnrrOs?t=13661> (consulté le 20 août 2020).

le cri de l'antagoniste rencontré ou le bruit de personnages tirant leurs armes dans le cas peu fréquent où le personnage jouable est en compagnie d'alliés.

Cette transition sonore souffre d'un problème de cohérence et d'intensité. En effet, le thème musical de conflit arrive sans avertissement, sans qu'une transition musicale soit faite entre la bande sonore précédente et celle-ci. Son volume est sensiblement plus élevé, et il prend plus de place dans la concrétion sonore que les thèmes d'accompagnement habituels. Notons qu'il ne survient pas toujours au bon moment : il arrive fréquemment que le thème se déclenche bien après le début du combat³⁷. Contrairement à d'autres jeux où la trame sonore, adaptative, permet de superposer plusieurs strates sonores pour créer une transition, celle de *Skyrim* n'est pas idéale pour que le joueur mesure l'ampleur, ou même la présence, de la situation de conflit. Ces facteurs, comme dans le cas d'*Assassin's Creed*, permettent pourtant au joueur de développer un nouvel état émotionnel et il va en toute logique rechercher la cause de l'agression sonore qu'il subit.

C'est ici que les choses se compliquent dans *Skyrim* : sans redondance visuelle, ni indice sonore intradiégétique, il devient difficile de repérer l'origine de l'attaque. Si notre seule indication reste le point rouge sur la boussole, cela ne résout pas totalement le manque d'informations. Tout d'abord, cette boussole ne permet pas de déterminer avec précision la distance entre le personnage jouable et l'agresseur, d'autant plus si l'antagoniste arrive hors du champ de vision du personnage jouable. De plus, le seul axe couvert par la boussole est l'axe des abscisses, alors que l'agression peut arriver de partout, y compris du ciel – auquel cas l'antagoniste est un dragon, un puissant ennemi contre lequel le joueur doit se concentrer.

³⁷ On peut voir dans cet exemple le déclenchement de la musique de combat à la fin de la situation de conflit. Cette transition tardive peut laisser supposer de l'arrivée d'un nouvel antagoniste alors que ce n'est pas le cas : <https://youtu.be/Dt6dY7O82-g?t=1711> (consulté le 20 août 2020).

Et c'est là le dernier problème que nous constatons dans les transitions sonores de *Skyrim* : il n'est pas possible de déterminer, à l'écoute, l'ampleur du danger qui guette notre joueur. Si les mêmes musiques retentissent avec la même intensité pour l'attaque d'un loup ou celle d'un dragon, comment faire pour hiérarchiser les informations nécessaires à l'appréhension rapide d'un combat? Les différentes musiques de conflit possèdent toutes les mêmes caractéristiques : un rythme rapide et percussif, une intensité sonore et un volume conséquent, un thème joué aux cordes et des percussions omniprésentes. Il n'y a pas non plus d'adéquation entre le ressenti émotionnel du joueur et la situation de conflit : dans un cas, la musique est bien trop agressive pour rendre compte d'une attaque anodine, dans l'autre, son caractère épique est amoindri par son omniprésence dans chaque situation de conflit, sans hiérarchie de ces dernières. Il est important de noter que ce dispositif peut être volontaire de la part des concepteurs pour créer un effet de surprise : le processus est fréquemment employé dans les jeux d'horreur. Mais le problème d'irruption de la musique à la fin de la situation de conflit nous fait plutôt pencher vers l'hypothèse d'une incohérence involontaire.

Utilisons comme contre-exemple un autre RPG dans lequel le joueur est amené à traverser, à pied ou à cheval, des environnements susceptibles d'abriter des antagonistes. Dans *The Witcher 3* (2015, CD Projekt), la situation de conflit s'annonce par la concrétion sonore suivante : le bruit associé à un antagoniste (dans cet exemple³⁸, des loups) survient par-dessus la musique d'exploration, avant que ledit antagoniste n'apparaisse par un point rouge sur la mini-carte de l'interface. Quelques secondes après, lorsque le joueur voit les loups sur l'écran et s'approche suffisamment d'eux pour qu'une animation d'attaque se déclenche,

³⁸ https://youtu.be/j_93xvcbhSA?t=245 (consulté le 23 août 2020).

la musique d'exploration s'arrête pour une musique de combat, plus dynamique. Le volume général de la concrétion sonore augmente aussi, le changement d'ambiance est très clair.

Si le joueur ressent de la lassitude, voire de la frustration, en raison d'indications peu précises, la continuité de l'expérience s'en trouve affaiblie. Ce problème proviendrait-il de l'appui sur une seule composante de l'espace sonore pour créer une transition? Pour pallier ce manque de précisions, certains concepteurs vont utiliser une caractéristique sonore différente et que nous avons peu évoquée jusqu'ici : la spatialisation.

3.1.2 La position du personnage jouable

Dans les environnements en 3D, la position du personnage jouable par rapport à la menace qui l'attend est particulièrement importante : si le joueur possède le contrôle sur la caméra, que ce soit à travers la position du personnage ou en pouvant modifier la direction du regard de celui-ci, la présence d'antagonistes est souvent signifiée par les sources sonores, qui vont combler le manque d'informations du hors-champ. Dans ce contexte, et de manière amplifiée dans des jeux où la réactivité du joueur est cruciale (les FPS, les jeux d'horreur), nous allons étudier l'impact du sonore sur les perceptions du joueur. Ce dernier va pouvoir guider son personnage jouable en fonction de la localisation des sons entendus. Dans ce cadre, la concrétion sonore devient encore plus importante, et le fait de jouer sans le son peut être très problématique, voire nuire totalement à la jouabilité de l'expérience.

Le jeu FPS multijoueur en ligne *Overwatch* (2015, Blizzard Entertainment) se démarque par une concrétion sonore particulièrement riche. Dans ce jeu, le joueur évolue en compagnie de cinq joueurs alliés, et l'équipe ainsi constituée doit réussir des objectifs qui dépendent de la carte (donc de l'espace de jeu) sur laquelle se déroule la partie. En face, une autre équipe

est elle aussi composée de six joueurs et doit réussir à défendre lesdits objectifs. Cet environnement de jeu entraîne forcément une confrontation perpétuelle. La situation de conflit d'*Overwatch* est complexe. Tout d'abord, elle est composée uniquement de personnages qui sont joués par des joueurs. Les déplacements et les choix de ces derniers sont moins prévisibles que ceux d'une intelligence artificielle. Ensuite, les personnages sont répartis selon trois rôles différents : des soigneurs, des tanks (qui absorbent les dégâts pour éviter que soigneurs et DPS ne meurent trop vite) et des DPS (dégâts par seconde). En fonction du rôle choisi par le joueur, celui-ci n'effectue pas les mêmes actions dans le jeu. Le positionnement dans l'environnement et l'attention face à l'agression seront différentes : un tank va se placer au-devant de son équipe et faire face à l'adversité tandis qu'un DPS, avec peu de points de vie, va profiter de l'espace de jeu pour se cacher et faire le plus de dégâts possibles sans en subir lui-même. Les soigneurs quant à eux sont plus centrés sur leur propre équipe, puisqu'ils doivent la maintenir en vie, et vont chercher à éviter le combat afin d'accomplir leur rôle. Comment, dans ces conditions, créer une concrétion sonore contenant toutes les informations nécessaires à la survie des deux équipes?

Prenons d'abord pour acquis que tous les joueurs n'entendent pas la même chose³⁹. La hiérarchisation des informations sonores se fait en fonction de priorités d'équipes et de positionnement : le joueur entend plus fortement un opposant qu'un allié, ce qui permet de déterminer rapidement la source de la menace. Mais cette source est mouvante, et cherche aussi à localiser son propre opposant. C'est ici que la spatialisation du son joue un rôle important dans la lecture du jeu. À l'instar des premiers jeux 3D, *Overwatch* va développer

³⁹ Afin d'aider la compréhension de cette sous-partie, nous proposons une séquence de jeu enregistrée en ligne et diffusée sur Youtube. Il est difficile de marquer temporellement chaque strate sonore décrite, en raison de la complexité de la concrétion sonore, mais nous espérons que le lecteur appréciera cette mise en situation audio-visuelle : https://youtu.be/k_nlPcHSJf8?t=19 (consulté en ligne le 21 août 2020).

la question des sources sonores pour non seulement les placer aux premier et second plans, mais également pour permettre au joueur d'identifier leur provenance sur une modélisation en trois dimensions. En fonction de sa position dans l'espace, et de la position de l'opposant, le joueur va entendre ledit opposant avec des critères de position : un volume plus ou moins fort en fonction de la distance du personnage à la source sonore, et surtout, une position sonore de l'opposant qui reflète son positionnement dans l'espace de jeu. Si l'opposant est à gauche du personnage joué, le joueur entendra dans son système d'écoute des bruits de pas ou la provenance de tirs du côté gauche. Et si le joueur change la direction de la caméra de son personnage, le son de l'opposant s'adapte à cette nouvelle position. Cela permet de déterminer très rapidement la source sonore, et de repérer dans un espace complexe où est situé, sur un plan visuel, l'assaillant que nous souhaitons contrer.

Bien entendu, l'espace visuel est redondant de l'espace sonore et donne au joueur les mêmes informations de positionnement, en affichant par exemple les barres de vie des antagonistes (l'interface va permettre de les séparer du reste du décor) ou la provenance des tirs dans certains cas. Cette répétition favorise la compréhension du joueur, et c'est l'association entre le son et le visuel qui crée de l'efficacité dans la résolution du défi. Le sonore pallie cependant les limitations imposées par le champ de vision, et permet d'informer le joueur d'attaques surgissant hors-champ. Dans un FPS multijoueur, la concrétion sonore est saturée d'informations que le joueur doit trier le plus rapidement possible : de ce tri dépend à la fois sa survie et son expérience de jeu, puisque la mort est punitive (dans *Overwatch*, elle empêche de jouer pendant six secondes dans une partie chronométrée, puis demande au joueur de rejoindre son équipe depuis son point de *respawn*, fixe et parfois très éloigné des combats).

Outre les bruits de tirs, le joueur d'*Overwatch* entend de nombreuses choses pendant sa partie. Tout d'abord, il possède un retour très direct sur ses propres actions : il entend les bruits de pas de son personnage, de même que le bruit de ses compétences de tir et de ses compétences spéciales⁴⁰. Il s'agit ici, comme l'évoque Jørgensen (2009, 93), d'une information de confirmation de commande permettant au joueur de savoir sans avoir besoin de le vérifier visuellement que son action manuelle sur le périphérique de contrôle a bien créé une action dans le jeu. Les compétences spéciales sont quant à elles des actions à l'effet conséquent sur la partie, qui ne sont disponibles qu'une fois toutes les X minutes (elles se construisent en fonction de l'impact du joueur sur sa partie). Chaque personnage possède une compétence spéciale, toujours la même, qui peut modifier la bataille de manière décisive.

Lorsqu'un allié ou un antagoniste déclenche cette capacité spéciale, un son intradiégétique qui lui est propre résonne dans tout l'espace de jeu, quelle que soit la position du joueur. Ce dernier peut déterminer si la compétence spéciale a été activée par l'un de ses alliés ou non en fonction de l'indication sonore : il y en a une pour la compétence d'un allié, et une autre si le personnage est un opposant. Cela permet, dans un jeu où les deux équipes peuvent choisir les mêmes personnages, de donner avec beaucoup de clarté une indication sur la source de l'attaque. Par exemple, le personnage de Dva (un tank) lance une bombe qui explose au bout de quatre secondes. Au lancer de la capacité, les ennemis de Dva vont entendre « Nerf this! », une injonction péremptoire relativement taquine qui non seulement demande de stopper une compétence invincible, mais surtout qui avertit que les joueurs ont à peine quelques secondes pour réagir et se cacher lorsqu'ils entendent ce son, sous peine de

⁴⁰ Précisons que dans *Overwatch*, aucun personnage ne possède les mêmes compétences. Entendre un certain type de tir permet aussi de renseigner avec précision le joueur aguerris qui pourra déterminer non seulement où est son assaillant, mais quel personnage du jeu l'attaque. Il pourra ainsi adapter sa propre attitude en fonction de cette information.

mourir immédiatement. Compte tenu du peu de temps de réaction disponible, la priorisation de cette source sonore est absolue, ce qui justifie à la fois son volume et sa priorité sur les autres sons de combat. De leur côté, les alliés de Dva entendent une commande lancée par le personnage : « Activating self-destruct sequence », une formulation interne, un peu comme une commande vocale, qui informe donc les alliés de la manipulation effectuée par le joueur, mais qui ne les enjoint pas de manière agressive à rediriger leur attention.

Pour terminer cette analyse de l'espace sonore d'*Overwatch*, nous nous arrêtons sur deux autres strates sonores. Tout d'abord, on va noter l'usage par les joueurs alliés d'indications stratégiques vocales. Sans parler du système vocal de discussion – que nous laissons de côté dans le cadre de cette recherche car les informations sonores qu'il donne n'ont pas été créées par les concepteurs – les joueurs peuvent envoyer des indications vocales préprogrammées à leurs alliés en appuyant sur une touche de leur clavier ou manette : « I need a doctor » demande expressément à un soigneur de l'équipe de tourner son attention vers un allié blessé, tandis qu'un « Stop the payload » indique aux alliés de concentrer leur attention sur l'objectif, qui passe souvent au second plan lorsque les combats font rage. Ces indications transdiégétiques sont offertes au joueur, qui peut choisir de les écouter ou continuer à se concentrer sur les sons de combat. Enfin, pour revenir vers un exemple de transition sonore, nous pouvons constater qu'une musique retentit de plus en plus fort à la fin du chronomètre de la partie, signalant au joueur qu'il est temps de se concentrer sur l'objectif de jeu à quelques secondes de sa fin. La musique, jusqu'alors anecdotique face à tous les signaux sonores, prend peu à peu une place très importante dans l'espace, et dans la hiérarchisation des indices audibles.

La combinaison des strates sonores pour former une concrétion cohérente et remplie d'informations est particulièrement perceptible ici. En donnant au joueur des indications très claires lui permettant de classer les sources de danger en temps réel et de réagir à la menace la plus importante, la concrétion sonore intègre une strate de spatialisation pour augmenter son potentiel d'avertissement et offrir la meilleure expérience possible au joueur. Bien entendu, cette spatialisation passe par l'usage d'un dispositif d'écoute adapté, au minimum en stéréo. Mais comme nous l'avons fait remarquer dans notre introduction, étudier l'emploi de dispositifs d'écoute précis dépasse l'ambition de cette recherche, et nous cherchons toujours à trouver le point d'accord entre le concepteur et le joueur, au prisme de la jouabilité idéale. Dans ces circonstances, on peut prendre en compte à la fois l'écoute attentive du joueur et l'utilisation idéale de la concrétion sonore pour offrir un système d'alertes précis, fiable et efficace. La redondance avec l'environnement visuel permet non seulement une cohérence, mais une expérience de jeu plus précise. Comme nous l'avons vu, la spatialisation sonore lorsqu'elle est utilisée couvre à la fois la situation de conflit et les situations de repos : elle alerte en réalité sur la position dudit conflit dans l'espace de jeu tout entier.

3.1.3 L'information donnée : le joueur attentif

Dans la situation de conflit direct avec un antagoniste, la concrétion sonore va à la fois préparer le joueur au conflit, en apportant des modifications qui entraînent un changement d'attitude, tout en conférant de l'ampleur à l'antagoniste présenté. Nous étudierons ici les différentes informations transmises au joueur pour créer un état de concentration, ce qui permet ensuite de développer la concrétion sonore autour de signaux plus précis qui vont dicter le combat lui-même.

Pour Karen Collins, « game sound can be a significant element of gameplay » (2008, 128). Si elle s'appuie également sur les fonctions déterminées du son au cinéma pour présenter sa théorie vidéoludique, l'auteure explore plus en détail la fonction préparatoire de la musique, puis du son. Collins discute en effet de la difficulté de la musique linéaire à activer des fonctions d'alerte pour le joueur (c'est d'ailleurs probablement un des aspects qui explique le dysfonctionnement de la musique d'attaque dans *Skyrim*).

A crucial role of music and sound effects in games is the preparatory function that the audio serves, for instance, to alert the player to an upcoming event, or to forewarn the player of approaching enemies. Anticipating action is a critical part of being successful in many games, particularly adventure and action games. Notably, *acousmatic* sound – sound with no clear origin visually – may inspire us to look to the direction of a sound, to “incite the look to go there and find out” (Chion 1994, pp.71, 85). This function, though present in films, [...] is far more pronounced in games, as sound gives the player cues to head in a particular direction or to run the other way, therefore affecting player decision making. (Collins 2008, p.130)

À cet égard, on peut noter que le joueur, dans sa courbe d'apprentissage du jeu, est de plus en plus attentif à des signaux qu'il peut reconnaître. Lorsque le joueur entre dans une situation de conflit, le son va être un des moyens importants de prendre la mesure de l'ampleur dudit conflit. Cela peut passer par un *leitmotiv* déjà entendu précédemment : « In *Ocarina [The Legend of Zelda : Ocarina of Time]*, for instance, the lesser enemies all have the same or similar music, and beneficial items like gems or pieces of heart likewise all have the same or similar-sounding cues. » (Collins 2008, 131)

Il est intéressant de constater que Collins relie l'indication sonore d'un antagoniste directement au même processus employé pour les récompenses données au joueur. Nous avançons l'idée que l'espace de conflit dans un jeu est facilement reconnaissable, comme nous permettent de le supposer les quantités de mèmes circulant sur internet et qui évoquent avec

humour la similarité de design entre les espaces de combat de différents jeux⁴¹ : « Big open room ... No enemies ... This is a boss room », déclare le plus célèbre d'entre eux sur fond d'un Gandalf suspicieux.



Figure 4. Mème populaire dont l'image est issue du film © *Le Seigneur des Anneaux : la Communauté de l'Anneau* (2001)

Le joueur est attentif aux signaux que le jeu lui transmet, et l'absence d'antagonistes, de même que l'omniprésence de bonus pour le joueur (soins, munitions, etc.) sont autant d'informations qui vont être prises en compte par le joueur pour anticiper ce qui l'attend.

⁴¹ <https://tvtropes.org/pmwiki/pmwiki.php/Main/BossRoom> consulté en ligne le 18 août 2020

Guillaume Roux-Girard parvient à la même conclusion dans sa recherche qui concerne spécifiquement le son dans les jeux vidéo d'horreur :

les attentes que le joueur entretient relativement à la jouabilité et la narration des jeux vidéo d'horreur esquissent une pluralité de schémas qui affectent sa façon de percevoir l'environnement du jeu et influencent directement ses actions en réaction à cet environnement. [...] En d'autres mots, nous pouvons affirmer que les schémas du jeu vidéo d'horreur, par le régime de conventions qu'ils impliquent, agissent tout d'abord comme une contrainte de lecture et d'écoute qui organise le travail cognitif et le processus d'apprentissage du joueur en fonction des attentes que ce dernier entretient avec la jouabilité et la narration du jeu. Ce travail, qui, d'un point de vue général, correspond au procédé descendant (*top-down*) actif dans le modèle en spirale d'Arsenault et Perron, force aussi le joueur à ajuster sa posture d'écoute selon les différents états de jeu qui lui sont présentés. Puisque le schéma principal de jouabilité des jeux vidéo d'horreur cherche à opposer le joueur à des menaces monstrueuses, l'objectif prépondérant de ce dernier devient, par la même occasion, la survie de son personnage-joueur à ces affrontements. Pour ce faire, le joueur doit amasser suffisamment d'information avec le procédé ascendant (*bottom-up*) afin de déterminer l'origine et la cause des sons dans l'espace, pour ensuite comparer ces données à son schéma général. (2009, 35-36)

Dans ce cadre, la concrétion sonore joue elle aussi son rôle à partir de la posture d'écoute du joueur. Un silence pesant (*Hollow Knight*), des rôles anxiogènes sans source visible (*Minecraft*) ou encore une indication sémantique claire, sous forme d'un dialogue d'avertissement (*Assassin's Creed*) vont pousser le joueur, à travers sa situation d'écoute, à trouver une explication dans le jeu par rapport à ce qu'il entend. Le point commun entre ces différentes sources sonores réside plus dans le changement que dans leurs similarités. Ce qui met la puce à l'oreille du joueur, c'est en effet la modification du processus de jeu. Si nous sommes en train de nettoyer un donjon de centaines d'ennemis qui se ruent sous nos pas depuis quelques dizaines de minutes dans *World of Warcraft* (2004, Blizzard Entertainment), c'est l'absence soudaine de ces antagonistes qui va nous interpeller.

Un combat plus important dans un jeu, caractérisé souvent en tant que *boss fight*, nécessite de la part du joueur des ressources d'investissement plus importantes. À cet égard, la préparation de l'attitude du joueur est primordiale pour éviter la frustration d'un combat

mal préparé dont l'issue sera très probablement fatale. Pour cette raison, la concrétion sonore se modifie en même temps que l'environnement de jeu pour rendre le joueur plus attentif. L'un des moyens employés le plus souvent est la soustraction de strates sonores : si la concrétion sonore est riche depuis plusieurs minutes ou heures, le calme relatif précédant un combat de boss risque d'être remarqué par le joueur. D'une autre façon, les concepteurs peuvent employer le suspense pour générer une série d'alertes et d'évènements sonores attirant l'attention du joueur. Bernard Perron, en étudiant les systèmes d'alerte présents dans les jeux vidéo d'horreur, évoque l'exemple de *Silent Hill* (1999, Konami) où le signal d'alerte est fondamentalement sonore : le bruit blanc d'une radio possédée par le personnage jouable se modifie selon la proximité dudit personnage avec des monstres, jusqu'à laisser la place aux bruits provenant directement du monstre lorsque le personnage arrive à leur rencontre (2004, 5). Ce bruit permet de créer ou de conserver l'attention du joueur, tout en augmentant la tension qu'il peut ressentir. L'alerte sonore possède alors une double fonction d'induction de ressenti émotionnel et d'information de jouabilité.

3.2 Combat et concrétion sonore

La concrétion sonore intervient aussi directement dans la situation de conflit entre l'antagoniste et le joueur par le biais du personnage jouable. En effet, si les indices sonores sont cruciaux pour signaler la modification de la situation de jeu, le passage d'un état vers un autre et la nécessité pour le joueur de modifier son attitude de jeu, le son va surtout permettre d'appréhender le combat lui-même⁴². Nous allons voir ici de quelle manière la concrétion sonore permet à la fois de situer l'antagoniste dans l'environnement de jeu, d'intégrer la difficulté du combat, et comment le sonore est inclus à l'intérieur du combat lui-même pour

⁴² Rappelons que la situation de conflit intègre tous les moments où le joueur est alerté sur une situation agressive, et que cette situation de conflit mène à un combat direct selon la définition simplifiée établie en introduction.

transmettre de nouvelles informations de jouabilité. Nous nous intéresserons plus en détail aux antagonistes à apparition unique, les *boss*, qui sont des antagonistes plus puissants et demandant plus de ressources de la part du joueur pour les combattre. Les *boss* possèdent généralement plus de points de vie que le personnage jouable, attaquent de plusieurs façons différentes et les combats sont déséquilibrés en faveur de l'antagoniste, dont les coups sont plus puissants que ceux portés par le personnage jouable.

3.2.1 Relier l'antagoniste à son environnement

Dans une situation de conflit, le joueur rencontre à la fois un antagoniste mais aussi un espace. Il peut y avoir un lien intéressant entre l'environnement traversé et l'antagoniste rencontré, même si la co-dépendance n'est pas toujours présente. En effet, certains antagonistes sont présents un peu partout à l'intérieur du jeu, tandis que d'autres se rapportent à un espace (qui peut être constitué d'un niveau, d'une salle ou d'un lieu traversé) plus précisément. Le joueur qui traverse l'espace par le biais de son personnage jouable fait le choix de (ou est guidé par le concepteur pour) traverser ou de s'opposer aux antagonistes en les combattant. Dans les deux dispositions, la concrétion sonore doit prendre en compte la totalité de l'environnement afin de permettre au joueur d'envisager les bonnes stratégies pour surmonter le défi proposé. L'hypothèse émise ici est la suivante : l'antagoniste doit être relié à son environnement, notamment par le son, afin de donner au joueur une expérience cohérente.

Le rôle du sonore peut être de rendre ce lien plus explicite pour le joueur. Pour Kristine Jørgensen, l'espace sonore fourmille de liens entre les sons, le joueur et l'environnement de jeu. « sounds that seem to be diegetically motivated also have usability functions, and sounds that seem to be motivated by usability are anchored in the virtual environment. » (2009, 82)

Comme nous l'avons évoqué, les catégories sonores ne sont pas fermées et leur perméabilité offre à la situation de conflit un équilibre entre jouabilité, induction de ressenti et narration. L'intégralité de la situation de conflit s'ancre dans le reste du jeu à travers les liens tissés par l'environnement et le sonore.

World of Warcraft propose deux types de combat : le *PvE (Player versus Environment)* et le *PvP (Player versus Player)*. Le premier regroupe tous les combats faits par les joueurs contre des créatures non contrôlables, qu'il s'agisse de dragons démesurés ou de sangliers a priori inoffensifs. La seconde catégorie fait état des combats de joueurs contre joueurs. Pour mieux expliquer le lien entre antagoniste et environnement, nous allons nous concentrer sur le PvE, et plus particulièrement sur le combat en raid et en donjon. Dans ceux-ci, un groupe de joueurs (de 5 à 40 personnes) est téléporté dans une instance, un lieu clos qui n'est pas directement relié au reste du monde proposé par le MMORPG. Cette instance est remplie d'antagonistes, principalement sous deux formes : des antagonistes-poubelle qui remplissent les couloirs labyrinthiques du donjon et qu'il faut combattre groupe par groupe⁴³, et des *boss* qui apparaissent quant à eux dans des salles aux dimensions gigantesques. Le groupe de joueurs alterne entre le labyrinthe et les salles de boss. On constate déjà ici la différence de design entre les combats contre les antagonistes-poubelle et ceux contre les boss, puisqu'ils n'évoluent pas dans les mêmes espaces de jeux. Les couloirs demandent à être traversés tandis qu'une salle vide, on l'a vu dans la sous-partie précédente, incite selon les connaissances préalables à la méfiance du joueur et le prépare à un affrontement plus important.

⁴³ <https://youtu.be/61Fb1wHRGyU?t=342> (consulté le 23 août 2020) pour percevoir la différence de design entre les couloirs et la salle du boss, dans laquelle le personnage jouable apparaît à 6'08''.

L'environnement transmet des indices sur le combat qui va avoir lieu. Mais le son participe grandement au combat. Dans les conflits les plus difficiles de *WoW*, la moindre attaque est susceptible de tuer d'un seul coup le personnage jouable. *WoW* est construit comme un jeu d'évitement, et l'espace de combat aussi vaste permet une liberté de mouvements du joueur faite pour répondre au besoin constant de ne pas être touché par une attaque. La coordination de l'équipe de joueurs se fait à travers un chat vocal, qui n'était pas implémenté dans le jeu mais qui est rapidement devenu indispensable. Sans pouvoir parler de réflexe car les combats n'ont pas un rythme très rapide, les sons dans l'environnement de conflit sont des alertes qui indiquent au joueur qu'il est une cible pour le boss, et qu'il va subir des dégâts à la suite d'une attaque. Ces sons sont extradiégétiques et informent le joueur sur différents aspects tactiques : manque de ressources pour lancer un sort, coup critique, ciblage par l'ennemi, etc. La différence principale qui réside entre ces sons et les sons réflexe que nous verrons plus tard est principalement une gestion du temps différente, le joueur de *WoW* bénéficiant d'un laps de temps plus long pour prendre des décisions pendant le combat.

Dans le cas de jeux dits en monde ouvert, la distinction entre la situation de conflit et le reste de l'environnement est moins évidente, elle passera souvent par l'intervention d'un changement dans la concrétion sonore. Si un ennemi repère le personnage jouable dans *Breath of the Wild*, une musique de combat arrive aussitôt et vient créer une transition avec les musiques environnementales calmes et apaisantes qui constituent la concrétion précédente. C'est le cas aussi dans *The Witcher 3*, où il est fréquent de se faire attaquer par un monstre en se promenant entre deux villes. La concrétion sonore ne contient pas que des strates dont le but est d'alerter : d'autres strates peuvent modifier l'état d'esprit du joueur et amplifier l'importance d'un combat plutôt qu'un autre. Nous le faisons remarquer depuis le

début de ce mémoire, les changements sonores véhiculent des changements d'ambiance et semblent encourager une modification du comportement du joueur. Pour Jørgensen, « the reason why the music so easily can change spatial status is that we create hypotheses about the meaning of the music related to the filmic context on screen (Branigan 1992:97) » (2009, 101).

La recherche de sens, de laquelle provient l'écoute causale, est au centre du comportement du joueur en jeu. Pour rendre cohérent l'environnement dans lequel son personnage jouable évolue, il est nécessaire de maintenir une attitude de crédulité, qui va passer aussi par la construction d'un sens. La concrétion sonore, qui va créer un lien entre les différents espaces traversés par le personnage jouable, renforce à la fois la cohérence de l'expérience et le ressenti que peut y éprouver le joueur. Toutefois, il est difficile d'étudier le sonore sans le relier au visuel : « game audio development follows the premises of ecological psychoacoustics in that it understands sound as a feature that cannot be regarded as isolated from the situations in which it appears (Neuhoff 2004, 2-3) » (Jørgensen 2009, 135). L'espace de jeu est co-construit avec la concrétion sonore, et c'est la combinaison audio-visuelle qui propose une expérience cohérente et riche de sens.

3.2.2 Alertes sonores et réflexes

Le son enclenche de fait des attitudes réflexes, conditionnées à la fois par les compétences du joueur et par son apprentissage du jeu. Lors de combats difficiles contre des antagonistes complexes, ces réflexes sont souvent les seuls moyens d'éviter une mort du personnage jouable. Comment alerter de la façon la plus efficace possible d'un danger imminent, et de quelle manière le joueur peut-il réagir? Étudions dans cette sous-partie, à travers des exemples de combat, de quelle manière la concrétion sonore devient constitutive

de la situation de conflit. Nous verrons que le début d'une animation d'attaque est souvent postérieur à un signal sonore ayant permis au joueur de réagir immédiatement et de se placer en sécurité ou de profiter de la situation pour attaquer lui-même.

Il est crucial pour le joueur d'analyser correctement les informations sonores qui lui sont transmises. Mais tout le monde n'en est pas au même point dans la spirale magique de l'apprentissage, et présumer des connaissances du joueur peut constituer une erreur de design. Pour créer la meilleure expérience de jouabilité possible, c'est le concepteur qui doit mettre toutes les chances de son côté en offrant une expérience de combat qui soit fluide, où la frustration engendrée par l'essai/erreur ne soit pas suffisante pour que le joueur quitte le jeu. Dans cette expérience, le son prend une importance encore plus grande. Lorsqu'il se penche sur les attitudes réflexes et la rapidité d'interprétation dans les jeux d'horreur, Guillaume Roux-Girard explique que « le jeu vidéo (et le film) sollicite d'abord la perception sous un mode préconscient. » (2009, 32) Kristine Jørgensen s'appuie sur les travaux de Sanders pour déterminer dans quelles situations les informations sonores peuvent être les plus pertinentes : « utilizing sound is a method for making the interface more intuitive and more user-friendly in situations where the visual system is busy, overburdened, or limited in other respects. » (2009, p.62) Grâce à Carrie Heeter et Pericles Gomes (1992), Jørgensen va également distinguer plusieurs signaux sonores différents :

The inclusion of sound seems to ease the ability to pick up and interpret information, as well as relieving the user from attentive tasks. Heeter & Gomes also put emphasis on the idea that providing redundant information through both the visual and the auditory channel, increases the likelihood that a certain message is received. On this basis, sound may be implemented in two basic ways as support for usability; either *proactively* or *reactively*. The proactive implementation is emphasized by McCormick & Sanders (1986:138-40), who point to the use of sound for providing information about situations that need immediate attention. In such cases, sound is used as notices, warnings or alerts. (2009, 62)

Ces sons d'alerte sont amplement utilisés dans les situations de conflit, où le nombre d'informations visuelles peut devenir trop important pour mener à bien le combat. D'autre part, les sons « réactifs » identifiés par Heeter et Gomes permettent de prodiguer rapidement une information de jouabilité supplémentaire, puisqu'ils autorisent l'identification des réponses du système de jeu aux inputs du joueur (Jørgensen 2009, 62). Dans une situation d'urgence, « the sound should be able to attract a person's attention as fast as possible. » (Jørgensen 2009, 92)

Le concepteur peut donc se servir du son pour activer des réflexes et donner au joueur des moyens de gérer un combat rapide et agressif. Dans un jeu multijoueur comme *Counter Strike* (2010 à 2012, Valve), il devient vital (selon les critères du jeu) de déterminer avec précision la position d'un antagoniste – un joueur faisant partie de l'équipe adverse – grâce à la spatialisation sonore évoquée précédemment. L'environnement de jeu étant volontairement créé pour permettre de se cacher dans de nombreux recoins, il est souvent délicat de déterminer avec précision la position d'un joueur. Les coups de feu tirés par l'antagoniste seront un moyen efficace de déterminer à la fois l'ampleur de l'attaque, si elle nous concerne ou non, et de récupérer des indices sonores sur la position de l'antagoniste.

Les jeux solo emploient le même principe à des échelles différentes. Sans forcément utiliser la spatialisation, le son va être un excellent moyen de rendre une attaque explicite. Lors d'un combat, le joueur doit gérer de multiples informations : sa propre barre de vie, pour commencer, et le nombre de coups qu'il peut prendre avant de mourir. Dans la plupart des jeux où l'antagoniste est un *boss*, les conflits sont volontairement déséquilibrés, et le personnage jouable est bien moins résistant que l'antagoniste qu'il combat. La situation de conflit demande notamment de l'endurance au joueur, qui doit faire de nombreux dégâts à

l'antagoniste tout en évitant ses coups. Le joueur doit ensuite gérer ses propres mouvements et les mouvements de l'antagoniste dans l'espace réservé au conflit. Il doit aussi parfois s'occuper de ses armes : les recharger (*Doom* (2016, Bethesda)), les changer si leur durabilité est épuisée (*The Legend of Zelda : Breath of the wild* (2017)) ou s'il passe d'une situation de conflit à distance à une situation au corps à corps (*Skyrim* (2011)). Dans ce déluge d'informations, le joueur s'appuie évidemment sur l'interface disponible et sur le visuel. Mais ces informations sont doublées, voire précédées par des signaux sonores d'alerte, intra- ou extradiégétiques, que l'attitude d'écoute du joueur va permettre d'appréhender rapidement.

Nous notons deux types d'alertes principaux : les alertes d'attaques, souvent intradiégétiques, sont des sons et des bruits poussés par l'antagoniste quelques dixièmes de seconde avant qu'il n'attaque (avant que son animation d'attaque ne soit déclenchée, ou pendant les animations précédant le déclenchement de ladite attaque : prise d'élan, concentration du tir, etc.). Ce système est employé à foison dans *Hollow Knight*, où les boss poussent différents cris durant les combats, qui aident parfois le joueur à comprendre quel type d'attaque va être lancé. Toutefois, ces cris ne sont pas forcément précis : le boss Hornet⁴⁴, par exemple, dispose de trois cris différents, qui ne sont pas liés à une attaque chacun. Le cri bref est simplement un moyen d'avertir le joueur : il n'est plus temps d'attaquer, il faut s'éloigner de l'antagoniste sous peine de subir des dégâts. Un signal bref, rapide et agressif lui permet donc de comprendre clairement qu'il doit reculer dans l'espace de combat qui lui est offert, mais pas d'anticiper avec précision quelle attaque il va subir.

⁴⁴ <https://youtu.be/CtI9XBX8aPU?t=5> (consulté en ligne le 23 août 2020).

Dans *The Legend of Zelda : Breath of the Wild*, le personnage jouable Link doit parfois combattre des Gardiens, des antagonistes particulièrement vifs et dont les attaques sont extrêmement puissantes. L'une d'elle, un rayon laser très difficile à éviter met à terre Link et ne lui laisse qu'une fraction minimale de ses points de vie. Ce rayon met quelques secondes à s'activer, pendant lesquelles le gardien s'immobilise pour rechercher sa cible. Ces secondes peuvent être mises à profit par le joueur qui peut courir et s'éloigner du gardien pour éviter le tir qui va suivre. Ce temps est symbolisé par une série de sons de plus en plus rapides, qui se détachent de la musique de combat et du bruit des armes. La même note aigue est répétée plusieurs fois, jusqu'à monter d'une tierce et résonner de manière continue, ce qui signifie que le gardien a verrouillé sa cible et va tirer. Le joueur peut alors, grâce à ce signal sonore, ne plus conserver l'antagoniste dans son champ de vision et se concentrer sur la mise d'un écart suffisant entre ce dernier et le personnage jouable. Il n'a pas non plus besoin de compter lui-même le court laps de temps entre le verrouillage et l'attaque, il peut se fier à son oreille pour parer ou esquiver au dernier moment⁴⁵. On constate ici l'utilisation d'un signal transdiégétique qui informe directement le joueur de la situation critique dans laquelle son personnage jouable est empêtré.

Une autre alerte sonore, en grande majorité extradiégétique, est très présente aussi durant les combats : c'est celle qui alerte le joueur sur le peu de vie de son personnage jouable. Pour rester dans *Breath of the Wild*, lorsque Link subit trop de dégâts, une alerte sonore forte résonne brutalement dans l'espace sonore. Ce son, qui peut couvrir d'autres informations sonores, est loin d'être agréable. Il est plus rapide que la musique de combat et génère donc un décalage rythmique entre les deux, qui résonne désagréablement à l'oreille. Par la suite,

⁴⁵ <https://youtu.be/EiOCerktVFA> (consulté le 23 août 2020).

une autre alerte moins puissante va continuer à résonner, sur un rythme lent qui évoque des battements de cœur, tant que les points de vie de Link ne seront pas remontés. Ils rajoutent à la fois une information cruciale, la mise en danger du personnage jouable, ainsi qu'une source de stress supplémentaire.

Ces alertes de points de vie faibles sont utilisées un peu partout dans les jeux vidéo et prennent souvent la même forme que dans *Breath of the Wild*. L'intégration d'un rythme cardiaque sonore émet un signal clair, largement utilisé au cinéma, qui symbolise une mise en danger importante. Les stratagèmes employés sont toutefois multiples : dans *Hollow Knight*, chaque coup subi par le personnage jouable est amplifié soudainement par un filtre sonore passe-bas, qui étouffe soudainement le son. La musique et les différents sons de combat sont instantanément coupés, comme si une bulle sonore éclatait subitement pour revenir quelques instants plus tard. Cette modification brutale permet d'intégrer immédiatement l'attaque subie.

Terminons cette partie sur un exemple différent, mais explicite. Reprenons l'exemple des raids de *Wow*. Les boss attaquent souvent avec des *AoE (Area of Effect)*, des attaques touchant une aire plutôt qu'une seule personne. Celles-ci sont symbolisées dans l'espace de jeu sur le sol, mais la multiplication des informations de combat ne permet pas toujours de bien les distinguer : le joueur doit gérer son personnage, veiller sur ses compagnons qui ne doivent pas mourir, utiliser de nombreux sorts, et voit sur son écran les effets qui affectent tous ses camarades en plus des siens. Afin d'obtenir des informations plus claires, les joueurs téléchargent des *addons*, autorisés par l'éditeur du jeu (*Blizzard Entertainment*), qui mettent l'emphase sur certains aspects cruciaux du combat en modifiant l'interface du jeu. L'un d'entre eux nous intéresse particulièrement. Il s'agit de *GTFO* (dont l'acronyme signifie *Get*

the Fuck Out, que l'on peut traduire par « Crisse ton camp! ») qui se présente de lui-même ainsi :

GTFO provides an audible alert when you're standing in something you're not supposed to be standing in. In some cases, you'll be warned before you start taking damage. This mod improves your situational awareness and is recommended for dungeon divers and raiders of all skill levels as even the most seasoned veteran sometimes needs a reminder to GTFO.⁴⁶

Certains joueurs vont ainsi ajouter une strate sonore qui n'est pas présente dans le jeu de base, afin de mieux réagir en cas d'*AoE* et de combler ce qu'ils peuvent considérer comme un manque de jouabilité pour l'expérience optimale. Notons aussi que le signal sonore peut retentir « avant que [le joueur subisse] des dégâts » (notre traduction), ce qui nous ramène à l'anticipation grâce au son⁴⁷.

Les stratégies sonores réflexes sont multiples, mais elles restent omniprésentes. L'impact de certaines strates sonores est exacerbé en cas de conflit, lorsque le joueur doit réagir rapidement à une situation qui lui demande de récupérer de très nombreuses informations en un laps de temps très court. Grâce aux alertes sonores, le combat est plus fluide et le joueur peut faire appel à ses réflexes.

3.2.3 Indices musicaux et rythmes d'attaque

En plein combat, le joueur doit à la fois attaquer l'antagoniste, mais il doit aussi éviter d'être touché. Dans la plupart des jeux, le déséquilibre est grand, lors des combats de *boss*, entre les dégâts donnés et les dégâts subis. Être victime d'une seule attaque d'un antagoniste puissant peut diviser la vie du personnage jouable par deux ou par trois. Dans ces conditions,

⁴⁶ <https://www.curseforge.com/wow/addons/gtfo>, consulté en ligne le 13 juin 2020.

⁴⁷ On trouve un exemple d'emploi de *GTFO* dans la vidéo illustrant l'exemple 3.2.1, que nous repropsons ici : <https://youtu.be/61Fb1wHRGyU?t=439>. Les sons « 4...3...2...1 » et « Beware » n'appartiennent pas au design sonore de base du jeu WoW.

les stratégies d'évitement et de défense doivent se mettre en place de la manière la plus effective possible. Il existe plusieurs stratégies sonores, employant à la fois le son général du combat mais aussi, notamment, la musique extradiégétique, pour permettre au joueur de réagir le plus rapidement possible, grâce aux attitudes réflexes étudiées précédemment. Dans certains cas, c'est même l'écoute qui va conditionner la réussite, ou du moins la compréhension, du combat vécu.

Nous parlons de musique de combat. Si celle-ci est un indice important pour différencier les situations de jeu, il ne faut pas pour autant limiter sa fonction au changement de ressenti. La musique peut proposer un usage tactique à l'intérieur du combat, et nous allons expliquer de quelle manière, à travers le rythme et les changements de composition, la musique peut aider le joueur dans l'affrontement direct.

Tout d'abord, les musiques de conflit sont en grande majorité plus rythmées que les autres musiques du jeu. La différence principale entre une musique d'ambiance et une musique de combat, comme nous l'avons déjà évoqué à plusieurs reprises, est l'omniprésence d'une pulsation claire et vigoureuse. Si la musique ne doit surtout pas gêner les alertes sonores de conflit, elle peut même les mettre en avant lorsque les attaques se déroulent en rythme. Un studio de jeux vidéo s'appuie particulièrement sur cette mécanique pour proposer des combats extrêmement difficiles et punitifs, qui peuvent être rendus un peu plus simples grâce à une écoute particulièrement attentive. Il s'agit de FromSoftware, à l'origine notamment de la saga des *Dark Souls* et plus récemment de *Sekiro: Shadows Die Twice* (2019, FromSoftware). Dans ces jeux d'aventure, le joueur incarne un chevalier ou un samouraï, bien moins puissant et bien moins résistant que les ennemis qu'il affronte. Les jeux sont construits entièrement autour du *die and retry*, la moindre erreur est très punitive. Les *boss* sont

particulièrement puissants et la mécanique de combat repose essentiellement sur l'évitement de la moindre agression, et l'attaque de l'antagoniste dans des laps de temps courts et précis. Les boss de *Dark Souls* ou de *Sekiro*⁴⁸ sont réputés pour leur imprévisibilité et la taille de leurs attaques, qui prennent souvent une bonne partie de l'espace de conflit et sont donc difficilement esquivables.

Le studio FromSoftware apporte toutefois une aide de taille au joueur : chaque boss possède sa propre musique de combat, et ces musiques permettent de prévoir à quel moment le boss va attaquer⁴⁹. En effet, les boss de *Dark Souls* attaquent en rythme. Nous allons entrer ici dans la sphère des perceptions auditives. La musique suit un rythme, qui est une manière de compter le déroulement du temps. Le tempo détermine la durée de chaque note. Il peut subir des accélérations ou des ralentissements. La base du rythme réside dans la pulsation, qui est le temps entre deux noires. Dans la musique occidentale, le rythme est soit binaire (décomposé par deux), soit ternaire (décomposé par trois). Puisque ces principes sont à la base de toute composition musicale, depuis des centaines, voire des milliers d'années, notre oreille est particulièrement efficace pour repérer le tempo d'un morceau lorsqu'il est simple et bien mis en avant. Dans les musiques de combat de FromSoftware, la lisibilité du tempo est la plus claire possible. Les temps sont marqués par un des instruments, et celui-ci possède un volume assez fort par rapport au reste de l'orchestration. Les attaques du boss sont systématiquement synchronisées avec la musique : l'antagoniste attaque sur les temps ou les demi-temps, et comme la pulsation est rendue explicite dans les morceaux (souvent par le

⁴⁸ La principale différence entre les deux franchises réside dans l'environnement jouable : un Moyen-Âge fantasmé dans les *Dark Souls*, contre un médiévalisme fantastique du Japon dans *Sekiro*. Les questions de jouabilité étant les mêmes, nous préférons évoquer les jeux comme un ensemble plutôt que de les distinguer lorsque ça n'est pas nécessaire.

⁴⁹ <https://youtu.be/rk6mzf240Ew?t=43> (consulté le 23 août 2020).

bruit des attaques elles-mêmes), on peut repérer assez facilement le moment où le personnage jouable est susceptible de subir une attaque. La musique au premier abord extradiégétique amène ici une importante part de jouabilité dans l'expérience, et les combats deviennent des jeux de rythme : le joueur peut utiliser l'information transdiégétique qui lui est donnée, à savoir le temps entre chaque attaque, pour attaquer lui-même. Bien entendu, la musique ne fait pas tout, et le joueur doit rester attentif aux informations données par le visuel, qui vont déterminer plus précisément quel sera le type d'attaque, si le boss esquive ou non, etc. Mais l'omniprésence de la musique dans *Dark Souls 3* (on constate d'ailleurs que le volume de celle-ci est très élevé par rapport aux autres sons du jeu) va apporter une compréhension plus efficace du combat. À ce titre, les alertes sonores intradiégétiques, comme le bruit de chaque attaque, sont synchronisées sur la musique et deviennent presque de nouveaux instruments ajoutés à la partition – on pourrait alors se demander si ces alertes ne deviennent pas transdiégétiques, en informant le joueur sur les rythmes du combat. L'espace sonore devient très lisible et permet alors de se concentrer sur l'information la plus importante : le temps qui sépare chaque attaque et les moments entre ces attaques permettant au joueur de riposter.

Cette stratégie n'est pas propre aux jeux du studio FromSoftware. *Crypt of the Necrodancer* (Brace Yourself Games, 2015), par exemple, est un jeu *roguelike* avec une proposition surprenante : dans ce jeu en 2D vu de dessus, tous les monstres et personnages rencontrés se déplacent et attaquent au rythme de la musique⁵⁰. À chaque temps de la pulsation, l'antagoniste bouge d'une ou de deux cases, ou reste statique. Le joueur est fortement encouragé à s'aligner sur le rythme des personnages non jouables (il gagne plus de

⁵⁰ https://youtu.be/nmZoRXHmx_E?t=3099 (consulté le 23 août 2020).

points s'il se déplace en rythme) et la démarche de combat en devient alors simplifiée : il suffit de repérer le schéma d'attaque de l'antagoniste sur le tempo de la musique.

Dans la plupart des jeux, le moyen le plus clair de marquer la pulsation passe par l'utilisation des percussions, dont c'est la fonction première. On remarque dès lors que toutes les musiques de combat dont nous avons parlé jusqu'ici employaient fréquemment les percussions pour marquer le combat. Si FromSoftware construit ses boss en fonction de la trame musicale, cette mécanique est employée à différentes échelles dans d'autres jeux⁵¹.

Notons qu'un rythme volontairement bancal peut déstabiliser le joueur et complexifier le combat. Dans *Crypt of the Necrodancer*, lorsque la difficulté augmente au fil des niveaux, la pulsation est moins simple à repérer, ce qui complique la jouabilité volontairement⁵². Dans *Dark Souls 3*, le *Dancer of the Boreal Valley* fonctionne sur un rythme complètement différent puisqu'il est ternaire, là où les autres sont tous binaires⁵³. Les mouvements sont alors décomposés par trois et non plus par deux, ce qui va donner l'impression que les temps d'attaque sont décalés par rapport à ce que le joueur a déjà construit grâce à son cycle de connaissances du jeu. De plus, la pulsation est bien moins claire dans la musique de ce combat : des voix soprano chantent de longues notes tenues que quelques violons viennent soutenir, ce qui n'aide pas le marquage du tempo.

À la lecture de ces combats qui reposent essentiellement sur le rythme de la musique, il est important de rappeler une fois de plus que la distinction entre intra- et extradiégétique

⁵¹ Ainsi, les jeux *Rayman* ont une longue tradition de niveaux musicaux dans lesquels le personnage jouable avance constamment. Le joueur doit éviter ou détruire les antagonistes fixés sur son chemin sans pouvoir s'arrêter. Il doit donc caler ses attaques sur le rythme de la musique : <https://youtu.be/7m5YQrucis8?t=37> (consulté en ligne le 23 août 2020) (*Rayman Legends*, Ubisoft 2013).

⁵² https://youtu.be/nmZoRXHmx_E?t=4359 (consulté le 23 août 2020).

⁵³ <https://youtu.be/jZ21JRZKSK4?t=48> (consulté le 23 août 2020).

ne peut fonctionner de manière aussi binaire dans le jeu vidéo qu'au cinéma en prises de vue réelles. La musique donne des indices de jouabilité très importants, et à ce titre nous rappelons que les considérations de Collins sont un bon moyen de lire l'espace sonore à travers un prisme qui a plus de sens. Ces utilisations de la musique, intimement liées à l'action, renforcent l'importance du lien entre le cinéma d'animation (par la figure du *mickeymousing*) et le jeu vidéo.

Dans ce chapitre, nous avons vu que la concrétion sonore relie de nombreuses fonctions à l'intérieur d'une même situation de jeu. On peut noter la présence de strates sonores informatives, d'alerte, d'aide ou de contrainte à la jouabilité. D'autres strates quant à elles induisent plutôt des ressentis émotionnels, ou contribuent à un effet de présence de l'antagoniste dans la diégèse du jeu. À l'intérieur de cette concrétion sonore, le joueur est plus ou moins attentif à chaque information. En fonction de ses besoins, liés de près à la situation rencontrée par son personnage jouable, il est en mesure de hiérarchiser toutes ces strates pour se concentrer sur celles qui seront particulièrement pertinentes. Le système d'écoute, en reposant sur l'action, met en avant le lien de chaque strate sonore avec le caractère dynamique de l'expérience. Nous avons vu que les concepteurs peuvent avoir recours à diverses stratégies de conception simplifier ou complexifier le combat en passant par la structuration sonore de la situation.

Conclusion

Ce mémoire a commencé par une intuition. De nombreux mois plus tard, occupés par une recherche passionnante et un apprentissage constant, il faut maintenant dresser une conclusion. Celle-ci doit refléter l'étendue du travail accompli tout en rappelant les limites de

la recherche : nous ne pouvons affirmer que nous avons trouvé toutes les réponses aux questions que nous avons posées. Dressons ici le bilan de nos découvertes.

Nous avons ouvert ce mémoire en nous questionnant sur la manière dont le sonore fonctionne dans un jeu vidéo. En nous basant sur une expérience d'écoute, nous pensons que les signaux audio générés par les créateurs de jeu apportaient de nombreuses informations au joueur, et nous avons cherché des précisions sur ce ressenti. Afin de circonscrire le sujet, il a rapidement fallu faire un choix, et nous avons décidé de concentrer nos recherches sur une situation de jeu précise : la situation de conflit, qui oppose de manière directe un personnage jouable à un antagoniste, lui-même jouable ou non-jouable. Dans cette situation, les signaux sonores se multiplient pour apporter au joueur plusieurs informations, avertissements et ressentis. Entre la musique, le bruit associé à l'antagoniste, ceux qui sont associés au personnage jouable et à ses armes, ainsi que parfois les sons d'interface, il fallait désormais décortiquer un ensemble confus afin de mieux comprendre son fonctionnement. Nous nous sommes donc demandé : de quelle manière le sonore se construit-il dans une situation de conflit vidéoludique ? Qu'est-ce que cette construction apporte au joueur ?

Afin de cerner à la fois le sujet et les concepts qui y étaient reliés, nous sommes allés chercher du côté d'un champ d'études cousin du jeu vidéo et un peu plus développé car plus âgé : les études cinématographiques. Le parallèle pouvait être établi rapidement entre notre sujet et les textes théoriques qui décrivent l'audio au cinéma. Les deux médias sont tout d'abord considérés comme audiovisuels, ou plutôt audio-visuels comme Michel Chion nous l'apprendra. Une dynamique particulière naît entre les signaux audio et vidéo, qui semblent fonctionner de concert pour proposer une expérience plus cohérente. Le sonore contribue à la création d'un fil conducteur temporel, en reliant entre elles des images qui ne prennent pas

le même sens selon l'audio qui y est relié. Avec son point de vue et sa rigueur cognitivistes, Annabel J. Cohen apporte quant à elle un modèle de fonctionnement de la réception des différents signaux par l'être humain confronté à l'expérience cinématographique. Elle détermine que des signaux concordants permettent au spectateur-auditeur de comprendre ou d'assimiler plus rapidement une information. En apportant cette première notion de hiérarchisation, Cohen établit alors des fonctions de l'audio, et plus précisément de la musique, dans le cinéma. On prend conscience de l'importance du sonore, et de la complexité des catégories qui s'articulent sans être fixées, qui finalement ne concernent pas seulement la musique mais à peu près tous les sons et bruits présents dans une expérience audio-visuelle.

Pour nous rapprocher du jeu vidéo, nous sommes par la suite allés étudier plus précisément le sonore dans le cinéma d'animation. Ce dernier partage une caractéristique supplémentaire avec le média qui nous intéresse : en plus de son caractère audio-visuel, il intègre une création visuelle totale, l'invention d'un univers dont les référents réels sont décalés, reproduits, imités mais ne sont pas vraiment présents. Tout comme en jeu vidéo, l'image représentée sort de l'imagination d'un créateur, elle est une perception de l'esprit et devient ensuite un dessin, une représentation. Dans ce cadre, le sonore joue un rôle supplémentaire en apportant de la crédibilité à l'expérience, mais également une illusion de vivacité. Les personnages mis en scène, lorsque leurs gestes sont appuyés par des éléments sonores et musicaux mimétiques, ont l'air plus... vivants. Cette constatation est soulignée par le travail de Cohen sur la concordance de signaux issus de sources sensorielles différentes. Avec l'aide de Zach Whalen et de son travail sur le jeu vidéo *Super Mario Bros.*, nous avons pu faire un premier lien entre le cinéma et le jeu vidéo par l'entremise du cinéma d'animation. Il s'avère que les mécanismes sonores de mimétisme du visuel sont abondamment employés

dans le jeu vidéo, où l'image construite nécessite la crédulité volontaire du joueur pour que l'expérience soit vécue à son plein potentiel.

Toutefois, le jeu vidéo se distingue du cinéma par bien des aspects et notre deuxième chapitre s'est concentré sur ces distinctions. Nous avons intégré à notre analyse le concept de jouabilité, qui a apporté une différence fondamentale entre les deux médias audio-visuels. Là où le cinéma est prévisible et reproduit la même expérience pour des auditeurs-spectateurs différents, le joueur impliqué dans l'expérience de jeu la rend interactive. Les actions ne sont pas reproduites au même rythme, au même moment pour tous les joueurs. Ces derniers peuvent choisir de revenir sur leurs pas, d'avancer très rapidement ou plus lentement, et le sonore s'adapte forcément à ces contraintes. La bande sonore ne se déroule plus de manière linéaire, mais en superposant plusieurs signaux qui interviennent en fonction des décisions du joueur : elle s'adapte.

À l'intérieur de ce cadre en mouvement, comment le joueur peut-il repérer certains signaux ou diriger son attention pour obtenir une information, comme c'est le cas dans le cinéma ? Nous nous sommes alors intéressés à l'écoute elle-même. Grâce aux travaux de Tuuri et Eerola, qui dressent une synthèse des différentes écoutes mises en lumière par des chercheurs en études cinématographiques et en études cognitivistes, nous parvenons à la conclusion que l'écoute se décline selon les besoins de l'auditeur, qui passe d'un mode d'écoute à un autre pour se concentrer sur les signaux sonores qui lui semblent pertinents. Cette hiérarchisation ne s'effectue pas selon la nature du signal sonore, qu'il soit musical, sonore ou sémantique. Elle semble plutôt formée autour de l'action qui intéresse l'auditeur, souvent la sienne, mais aussi celle qui se déroule à l'écran. Dans le cadre de la situation de

conflit vidéoludique, les actions prioritaires sont celles reliées au combat, issue fatale de la confrontation entre le personnage jouable et un antagoniste.

Il était temps, au milieu de ce mémoire, de reporter notre attention sur la situation de conflit. Nous avons fait usage des connaissances acquises pour émettre l'hypothèse d'une concrétion sonore, un ensemble de strates sonores composées de différents signaux et qui forment une situation sonore complète et cohérente, en accord avec l'expérience visuelle qui se déroule. Le joueur hiérarchise, à l'intérieur de cette concrétion, les strates sonores selon le rapport direct qu'elles entretiennent avec la situation de conflit. Si une information sonore est cruciale pour la survie du personnage jouable, le joueur va naturellement se concentrer sur celle-ci et lui accorder une priorité de traitement. Toutefois, notre oreille n'est pas vraiment capable de supprimer les autres signaux sonores, et c'est pour cela que nous avons préféré parler de la constitution d'une concrétion plutôt que de séparer les signaux sonores en fonction de leur nature. Nous émettons à ce moment l'hypothèse que des signaux sonores de nature différente peuvent apporter la même information. À l'inverse, des signaux sonores provenant de la même source véhiculent parfois de nombreux sens. Le joueur, qui constitue en jouant un ensemble de connaissances et d'attentes, va avec l'aide des concepteurs (qui divisent, par exemple, la bande sonore en plusieurs pistes aux volumes différents) être en mesure d'identifier plus ou moins rapidement les strates sonores qui apportent des informations pertinentes pour la situation de conflit à laquelle il est confronté – à travers son personnage jouable. Nous avons ensuite exposé la multiplicité des strates sonores en utilisant l'exemple d'une situation de jeu dans *Assassin's Creed : Brotherhood*, afin de mettre en lumière notre hypothèse basée sur la perception du joueur.

Une fois cette concrétion sonore établie, nous avons choisi de mettre en avant un autre point fondamental issu de nos hypothèses de départ : le fonctionnement commun de la concrétion sonore dans des situations de conflit issues de genres de jeux différents. Nous pensions en effet que le joueur, grâce à ses expériences de jeu, constituait un ensemble de connaissances à invoquer au fur et à mesure de son avancée dans le jeu, connaissance qu'il pouvait également transférer d'un jeu à l'autre. Est-ce que la concrétion sonore change drastiquement dans un jeu de plateforme, un jeu multijoueur en ligne ou un RPG ? Nous avons présenté différentes concrétions sonores, en démontrant que des signaux sonores de nature différente contribuent à apporter une information similaire d'un jeu à un autre. La situation de conflit se construit autour d'un combat, et celui-ci est extrêmement alimenté par les strates sonores qui offrent à la fois une crédibilité à l'expérience vidéoludique, de l'aide pour le joueur ou au contraire une difficulté supplémentaire, ainsi qu'une emphase sur la situation de conflit par rapport à une autre situation présente dans le même jeu.

À travers différents exemples, nous avons donc présenté plusieurs fonctions issues de plusieurs strates sonores et hiérarchisées par le joueur selon ses besoins. Nous nous sommes intéressés à la fonction d'alerte, qui indique au joueur la nécessité urgente de se concentrer sur un signal précis pour modifier son attitude de jeu. Le signal d'alerte peut indiquer la transition entre une situation d'exploration et une situation de conflit, l'imminence d'une attaque ou le faible nombre de points de vie restants au personnage jouable. Nous nous sommes penchés sur les aides de navigation proposées par la concrétion sonore : les différences de volume, la spatialisation de certains bruits dans l'environnement de jeu ou encore la préparation à une situation de conflit à travers des données audio sont autant de moyens d'informer le joueur rapidement tout en lui fournissant de nombreuses informations différentes.

En étudiant le combat direct à l'intérieur d'une situation de conflit, nous avons pu montrer que le sonore joue un rôle d'intégration de l'antagoniste dans le reste du jeu, en créant un lien sonore entre les différentes situations rencontrées par le joueur. Dans cette perspective, la musique constitue à la fois ce lien tout en apportant directement des aides (ou des contraintes) à la jouabilité. Elle rajoute en effet dans la concrétion sonore le sens du rythme et la pulsation, qui peuvent tout à fait être utilisés de manière stratégique par les développeurs et ainsi créer du sens pour le joueur.

Ce qu'il faut retenir de ces dernières analyses, c'est que la nature de ce qu'on entend ne détermine pas vraiment la fonction qui en découle. Des signaux sonores aussi divers qu'un son d'interface, une musique de combat ou des bruits associés aux grognements d'une créature constituent tout autant des signaux d'alerte qui peuvent avertir le joueur de l'imminence de la situation de conflit. Finalement, ce qui va permettre au joueur de tirer parti de cette concrétion sonore, c'est plutôt sa stratégie d'écoute. En prêtant l'oreille aux différentes strates sonores, le joueur attentif peut se servir de la concrétion sonore pour mieux analyser et comprendre la situation de jeu dans laquelle se trouve son personnage jouable. Bien entendu, les informations audio vont de pair avec l'ensemble des informations visuelles. Sans créer de séparation factice entre les deux, nous croyons que le modèle développé par Annabel J. Cohen explicite très bien la communication qui opère entre plusieurs types de signaux reçus et interprétés pendant l'expérience audiovisuelle.

Cette recherche a également mis en lumière la difficulté de normaliser une forme de classement des signaux sonores. En ne séparant pas la musique du reste des signaux, nous avons déjà renoncé à une grande majorité de modèles émis par les ludomusicologues. Nous croyons toujours aussi fortement qu'il est nécessaire, pour mieux rendre compte de

l'expérience vécue par le joueur, d'envisager les signaux sonores dans leur ensemble, à la manière d'un joueur attentif qui écoute la concrétion sonore d'un jeu vidéo. L'autre manière de trier ces signaux passait pour beaucoup par leur lien avec la diégèse proposée. Ce modèle, qui fonctionne très bien pour l'expérience audio-visuelle cinématographique, montre des limites lorsqu'il est étendu au jeu vidéo, comme le démontre Kristine Jørgensen en théorisant un aspect transdiégétique, lorsque les sons font partie de la diégèse du jeu mais qu'ils existent uniquement pour avertir le joueur. Cette explication, dans le cadre de la situation de conflit où les réflexes, les actions et l'analyse priment sur la profondeur diégétique de l'expérience, ne nous satisfaisait pas. C'est pour cette raison que nous avons finalement choisi le modèle émis par Karen Collins, où les signaux sonores sont analysés selon leur rapport à l'action, à l'instar du modèle d'écoute mis en lumière dans notre recherche.

Par sa nature, cette recherche est loin d'être complète ou exhaustive. Nous pensons toutefois avoir mis en lumière plusieurs aspects essentiels du fonctionnement du sonore dans une situation de conflit vidéoludique. Nous espérons, à l'avenir, être en mesure de préciser certaines de nos hypothèses à l'aide de nouvelles recherches, dans une démarche universitaire cohérente et justifiée. Le fonctionnement du sonore est un aspect vidéoludique encore peu mis en avant, et nous souhaitons nous inscrire dans une dynamique d'exploration de ce système fascinant et peu accessible.

Références bibliographiques

Albersnagel, Franz A. 1988. « Velten and musical mood induction procedures : A comparison with accessibility of thought associations », *Behaviour Research and Therapy* 1 n°26, 79-95.

Arsenault, Dominic, et Bernard Perron. 2009. « In the Frame of the Magic Cycle: The Circle(s) of Gameplay ». Dans Bernard Perron et Mark J.P. Wolf (dir.), *The Video Game Theory Reader* 2, 109-132. New York : Routledge.

Boulez, Pierre, Jean-Pierre Changeux et Phillipe Manoury. 2014. *Les neurones enchantés. Le cerveau et la musique*. Paris : Odile Jacob.

Caillois, Roger. 1958. *Les Jeux et les Hommes (le masque et le vertige)*. Paris : Gallimard.

Chion, Michel. (1990) 2013. *L'audio-vision. Son et image au cinéma*. 3^e éd. Version numérique d'un livre déjà paru. Paris : Armand Colin, 2013.

Cohen, Annabel J. 1998. « The Functions of Music in Multimedia: a Cognitive Approach », *Proceedings of the Fifth International Conference on Music Perception and Cognition*. Seoul : Western Music Research Institute, Seoul National University.

Cohen, Annabel J., Roger A. Kendall, Scott D. Lipscomb et Siu-Lan Tan. 2013. *The Psychology of Music in Multimedia*. Version numérique d'un livre déjà paru. Oxford : Oxford University Press.

Collins, Karen. 2008. *Game Sound. An Introduction to the history, theory and practice of video game music and sound design*. Cambridge: The MIT Press.

Eco, Umberto. (1979) 1985. *Lector in fabula*. Paris : Grasset.

Eerola, Tuomas et Kai Tuuri. 2012. « Formulating a Revised Taxonomy for Modes of Listening ». *Journal of New Music Research* 41 n°2, 137-152.

Jørgensen, Kristine. 2009. *A Comprehensive Study of Sound in Computer Games. How audio affects player action*. Lewiston : The Edwin Mellen Press.

Juul, Jesper. 2005. *Half-Real. Video Games between Real Rules and Fictional Worlds*. Cambridge : The MIT Press.

Kamp, Michiel. 2013. « Musical Ecologies in Video Games ». *Philosophy and Technology* 2 n°27 (juin), 235-249.

Metz, Christian. 1974. *Film language: A semiotics of the cinema*. New York : Oxford University Press.

Perron, Bernard. « Sign of a Threat: The Effects of Warning Systems in Survival Horror Games ». *COSIGN 2004, 4th Conference on Computation Semiotics*. Croatie : University of Split.

Consulté en ligne :

http://www.cosignconference.org/downloads/papers/perron_cosign_2004.pdf

Pignatiello, Michael F., Cameron J. Camp et Lee Anna Rasar. 1986. « Musical mood induction: An alternative to the Velten technique ». *Journal of Abnormal Psychology* 95, n°3, 295-297.

Rossetti, Gregory James. 2020. « Overworlds, towns, and battles: how music develops the worlds of role-playing video games ». Thèse de doctorat, Rutgers Université du New Jersey.

Consulté en ligne : <https://doi.org/doi:10.7282/t3-ze9j-zx68>

Roux-Girard, Guillaume. 2009. « L'écoute de la peur : une étude du son dans les jeux vidéo d'horreur ». Mémoire de maîtrise, Université de Montréal.

Rouse, Richard. 2001. *Game Design : Theory and Practice*. Plano : Wordware.

Salen, Katie et Eric Zimmerman. 2004. *Rules of Play : Game Design Fundamentals*. Cambridge : The MIT Press.

Tuuri, Kai, Manne-Sakari Mustonen et Antti Pirhonen. 2007. « Same sound – different meanings: A novel scheme for modes of listening ». Dans *Audio Mostly 2007, 2nd Conference on Interaction with Sound*, sous la direction de Katarina Delsing, 13-18. Actes du colloque « *Proceedings of Audio Mostly 2007* » Röntgenbau, Ilmenau 27-28 septembre 2007.

Ilmenau : Fraunhofer Institute for Digital Media Technology IDMT.

Varela, Francisco, Eleanor Rosch et Evan Thompson. 1991. *The Embodied Mind*. Cambridge : The MIT Press.

Whalen, Zach. 2004. « Play Along –An Approach to Videogame Music ». *Game Studies* 4 n°1.

Consulté en ligne : <http://gamestudies.org/0401/whalen/>.

Ludographie

Assassin's Creed : Brotherhood (2011, Ubisoft Montréal)

Assassin's Creed Origins (2017, Ubisoft Montréal)

Assassin's Creed (franchise)

Counter Strike (2010 à 2012, Valve)

Dark Souls III (2016, FromSoftware)

Doom (2016, Bethesda Softworks)

Grim Fandango (1998, LucasArts)

Hellblade : Senua's Sacrifice (2017, Ninja Theory)

Hollow Knight (2017, TeamCherry)

Hitman: Contracts (2004, Eidos Interactive)

Metal Gear Solid (1998, Konami)

Minecraft (2009, Markus Persson, puis Mojang)

Overwatch (2015, Blizzard Entertainment)

Rayman Legends (2013, Ubisoft)

Sekiro: Shadows die Twice (2019, FromSoftware)

Silent Hill (1999, Konami)

Super Mario Bros. (Nintendo, 1985)

The Elder Scrolls V : Skyrim (2011, Bethesda Softworks)

The Legend of Zelda : Breath of the Wild (2017, Nintendo)

The Legend of Zelda : Ocarina of Time (1998, Nintendo)

The Witcher 3 (2015, CD Projekt)

World of Warcraft (2004 à 2020, Blizzard Entertainment)