École Nationale Supérieure Louis Lumière

Promotion Son 2015

Du mutisme au dialogue

Les interactions vocales dans le jeu vidéo

Partie pratique : *v0x*

Mémoire de fin d'étude

Rédacteur : Charles MEYER

Directeur interne: Thierry CODUYS

Directrice externe : Isabelle BALLET

Rapporteur : Claude GAZEAU

Année universitaire 2014-2015

Mémoire soutenu le 15 juin 2015

Remerciements :

Je tiens à remercier chaleureusement mes deux directeurs de mémoire pour leur implication, leur confiance et leur exigence.

Je remercie tout particulièrement Nicolas GIDON, sans qui la réalisation de la partie pratique de ce mémoire aurait été plus chronophage et complexe..

Je remercie et salue Nicolas FOURNIER et Baptiste PALACIN, dont les travaux et la gentillesse ont été une source d'inspiration et de détermination.

Je remercie également ma mère, ma tante, Jordy, Julien et Julien (n'en déplaise à Julien), Timothée et mes amis pour leur soutien indéfectible.

Merci à madame VALOUR, monsieur COLLET, monsieur FARBRÈGES

ainsi qu'à leurs élèves.

Enfin, merci à *From Software* et à *NetherRealm Studios* pour leur jeux, qui auront été un défouloir bienvenu.

Résumé

Ce mémoire de master a pour objet d'étude les interactions vocales dans le jeu vidéo.

Cependant, il ne se limite pas à une étude historique de l'évolution de la vocalité au sein des jeux vidéo mais en propose une formalisation théorique autour de trois concepts essentiels : **Mécanique**, **Narration** et **Immersion**. De ces trois concepts découlent trois types de voix : les **voix système**, les **voix narratives** (**linéaires** et **non-linéaires**) et les **voix d'ambiance**.

Dans le prolongement de cette étude et en s'appuyant sur les travaux menés dans le cadre des parties expérimentale et pratique de ce mémoire, ayant abouti à la réalisation d'un jeu vidéo basé sur l'analyse spectrale de la voix du joueur, ***v0x***, nous proposons une extension de cette théorie de la vocalité vidéo-ludique afin d'intégrer l'inclusion de la voix du joueur au sein de ce cadre d'étude.

Nous déclinons dans la dernière partie de ce mémoire deux types d'écritures vocales permettant de distinguer les jeux dans lesquels la voix du joueur est intégrée

au *gameplay,* des autres jeux : les écritures vocales **unidirectionnelles** et **bidirectionnelles**.

Enfin, nous étudions les différentes méthodes de détection et de reconnaissance vocale appliquées au jeu vidéo jusqu'à aujourd'hui pour proposer une nouvelle utilisation de l'analyse spectrale de la voix du joueur.

Mots-clés : Jeu vidéo, *gameplay*, voix, écriture vocale, reconnaissance vocale, analyse spectrale, Mécanique, Narration, Immersion.

Abstract

This master's thesis is meant to study vocal interactions in video games.

Thus, it does not solely consist in a study of *vocality* thoughout the history of video game but aims to propose a new theoretical set of concepts and tools to analyse the use of voice in video games. From three essential concepts - **Mechanics**, **Narration** and **Immersion** – we deduce three central types of voices : **system voices**, **narrative voice** (**linear** and **non-linear**) and **ambiance voices**.

As a mean to extend this formal analysis, this master's thesis also studies the many ways to include the player's voice into the *gameplay* of a game.

Based upon the experimental and practical parts of this thesis, including the making of an experimental game, ***v0x***, with a *gameplay* powered by the spectral analysis of the player's voice, we elaborate the concepts of **unidirectional** and **bidirectional vocal writing** to distinguish games including the player's voice into their *gameplays* from other games.

Finally, the last part of this thesis studies the application of speech recognition technologies to video games and their principles to propose a new way of considering vocal interactions and integrating the player's voice to a game's gameplay : spectral analysis and pitch detection.

Keywords : Video Games, gameplay, voice, vocal writing, vocal recognition, spectral analysis, Mechanics, Narration, Immersion.

Table des matières

INTRODUCTION

« Mais en chemin, le jouet s'est cassé, le jeu s'est ouvert. Maintenant, Il grésille et il dit la vérité. [···] Et moi alors ? "Nous ne contrôlons pas le contenu, nous créons le contexte", dit Rose qui parle maintenant à la place du progamme. »

Mathieu Triclot – Philosophie des jeux vidéo, 2011, p. 237-238.

Dans la conclusion de son ouvrage sur les jeux vidéo, Mathieu Triclot étudie le dernier acte du jeu ***Metal Gear Solid 2 : Sons of Liberty*** (Konami, 2001). Il y décrit la spécialisation de ce jeu d'infiltration et d'action dans « la rupture du pacte ludique, dans l'omniprésence de l'interface, dans une esthétique de la perte du contrôle qui n'hésite pas à provoquer le joueur[1]. »

Qu'est-ce que ce pacte ludique que ***MGS2*** entreprend de rompre ?

Le jeu en lui même, et plus particulièrement son dernier acte, est un excellent point de départ pour répondre à cette question. En effet, un basculement s'opère lors de la conversation entre Rose et Raiden, le soldat incarné par le joueur : soudainement, la jeune femme s'adresse directement à lui et non plus au personnage de pixels qu'il contrôle à l'écran. Comme au théâtre, le quatrième mur s'effondre. Comme dans une salle de cinéma, la toile de l'écran se déchire. Les personnages révèlent qu'ils ont conscience qu'un tiers, le joueur, participe à leurs existences.

Nous en déduisons que le pacte ludique, liant le joueur au(x) créateur(s) d'un jeu, est très proche de la *suspension consentie de l'incrédulité* théorisée par le poète

1 *Mathieu Triclot,* ***Philosophie des jeux Vidéo****, Paris, La Découverte, Collection Zones, 2011, p. 238.*

anglais Samuel Taylor Coleridge[2]. Autrement dit, le pacte ludique consiste en une adhésion totale du joueur à l'univers qui lui est proposé ainsi qu'à ses principes de fonctionnement, au point d'accepter le surnaturel.

Mais ce qui fait la spécificité et l'inventivité de la série des ***Metal Gear Solid,*** c'est l'utilisation de la voix comme élément principal de rupture. Dans ***MGS2***, c'est au cours de dialogues radio que le trouble s'installe. Dans ***MGS1***[3], Psycho Mantis, un des antagonistes doté de facultés de télépathie, explore et commente à voix haute les fichiers personnels du joueur contenus dans sa carte mémoire.

Ces scènes troublantes relevant, comme l'indique Mathieu Triclot, d'une « esthétique de la perte de contrôle, » s'insèrent, qui plus est, dans une intrigue dont les dialogues sont le moteur.

Et si le nombre de jeux pratiquant une forme de distanciation[4] du joueur a augmenté depuis la parution de ***MGS1***, l'emploi de la voix comme outil narratif principal, déjà très répandu à l'époque, est aujourd'hui devenu incontournable. Le fait que des acteurs célèbres prêtent leur voix à des personnages est même devenu un standard dans le cadre de productions vidéo-ludiques à budget conséquent : en 2014,

---

2 *« [...] il fut convenu que je concentrerais mes efforts sur des personnages surnaturels, ou au moins romantiques, afin de faire naître en chacun de nous un intérêt humain et un semblant de vérité suffisants pour accorder, pour un moment, à ces fruits de l'imagination cette "suspension consentie de l'incrédulité", qui constitue la foi poétique. », Samuel Taylor Coleridge,* ***Biographia Literaria****, op. cit., vol. 7.2, chap. XIV, 1817, p. 6.*

3 Ce sont plus précisément les fichiers de sauvegarde et les statistiques associées qui sont examinés par le jeu afin de déterminer la ludothèque du joueur. Il en résulte des dialogues s'adaptant de façon troublante à chaque joueur. Quelques exemples de lignes de dialogues : *« Donc tu aimes* ***Suikoden*** *? » (****Suikoden****, Konami, 1995) ; « Tu ne sauvegardes pas très souvent, tu es du genre imprudent. »****Metal Gear Solid****, Konami, 1998.*

4 *La distanciation, en allemand Verfremdungseffekt*, est un concept théorisé par le dramaturge Bertolt Brecht pour décrire la prise de distance du spectateur de théâtre avec des personnages auquel il lui est impossible de s'identifier, en général en étant ramené à sa condition de spectateur.

Peter Dinklage, célèbre depuis sa participation à la série ***Game of Thrones*** (HBO, 2011) incarne le Spectre, compagnon de jeu robotique du joueur tout au long de ***Destiny***[5] (Bungie, Activision, 2014) et Kevin Spacey offre sa voix et son visage à l'antagoniste principal de ***Call of Duty : Advanced Warfare***[6] (Sledgehammer Games, Activision, 2014).

Quand la voix a-t-elle pris une telle importance au sein des jeux vidéo ? Quelles interactions entre le joueur et le jeu permet-elle de créer ? Pourrions-nous concevoir de nouvelles interactions vocales afin d'offrir plus de possibilités créatives aux concepteurs de jeu vidéo ?

Pour répondre à ces questions, il s'agit avant tout de définir les deux concepts essentiels des problématiques ci-dessus : la voix et le jeu vidéo.

La **voix** est l'ensemble des sons résultant de l'excitation d'un milieu par le larynx et les cavités supra-glottiques d'un individu.

Lorsque ces sons sont émis dans le but de communiquer, la voix devient **parole**.

De ces définitions découlent deux remarques essentielles :

- Les modes de production de la voix et de la parole sont intrinsèques à leur

5 *À sa sortie,* ***Destiny*** *est le jeu vidéo au budget le plus élevé. (http://www.franceinfo.fr/emission/jeux-video/2014-2015/destiny-le-jeu-video-qui-valait-500-millions-de-dollars-09-09-2014-10-05)*

6 Il est d'ailleurs intéressant de remarquer que la présence de Kevin Spacey au casting de ***CoDAW*** a été un des arguments marketing principaux d'Activision. *(http://www.adweek.com/news/advertising-branding/inside-massive-campaign-behind-call-duty-advanced-warfare-160993)*

existence[7]. Toute voix suggère donc un corps, de la chair, des muscles, faisant vibrer un milieu de propagation. Ceci va induire plus tard dans ce texte l'apport de nuances dans la caractérisation de la voix d'un personnage de jeu vidéo.

- À la notion de voix ne sont pas nécessairement associées les notions de **langage**[8] et de **langue**[9]. Pour cette raison, nous ne décrirons que très brièvement les technologies de détection et de reconnaissance de la parole appliquées au jeu vidéo. En effet, la complexité des algorithmes utilisés, ainsi que le très haut niveau de compétence requis en traitement du signal, rendent ce sujet ardu à traiter dans le cadre d'un mémoire de master sans qu'il lui soit entièrement dédié. Nous adopterons plutôt une approche de vulgarisation au sujet de ces technologies.

Un **jeu vidéo** est un ensemble d'interactions, liant d'une part une interface électronique et audiovisuelle à un joueur, régies d'autre part par un système de règles et dont le résultat est une « expérience instrumentée[10]. »

Cette définition fait apparaître de nouveaux concepts dont l'explicitation est indispensable à la compréhension de notre objet d'étude.

---

7 *« La parole se distingue des autres sons par des caractéristiques acoustiques, qui ont leur origine dans les mécanismes de production. » Calliope,* ***La parole et son traitement automatique****, Paris, Masson et CNET-ENST, 1989, p. 2.*

8 *« Moyen de communication utilisé par une communauté humaine ou animale pour transmettre des messages. Un langage est constitué d'unités minimales appelées signes ou signaux. » R. Galisson et D. Coste,* ***Dictionnaire de didactique des langues****, Hachette, 1976 p. 305.*

9 *« Tout système spécifique de signes articulés, servant à transmettre des messages humains. » Saussure ;* ***Ibidem****.*
À ce titre, la parole a un aspect individuel : elle est la réalisation de la langue par un locuteur.

10 *« Le jeu vidéo, il me semble que nous le reconnaissons d'abord comme une certaine forme d'expérience, une "expérience instrumentée", bien sûr, qui a besoin de l'écran et de la machine de calcul pour se produire. » Mathieu Triclot,* ***Philosophie des jeux video****, Paris, La Découverte, Collection Zones, 2011, p. 11.*

Une **interaction** est l'action réciproque d'un élément sur un autre.

Dans le cadre de ce mémoire de master, questionner les interactions vocales dans le jeu vidéo revient à étudier ce que l'émission de sons vocaux par le jeu provoque pour le joueur, mais aussi comment le joueur pourrait agir, grâce à sa voix, sur le jeu.

Une **interface** est une « frontière conventionnelle entre deux systèmes ou deux unités, permettant des échanges d'informations suivant des règles déterminées[11]. »

Dans le cas du jeu vidéo, les deux systèmes, le joueur et le jeu, sont mis en relation par l'intermédiaire d'une machine, qu'elle soit un ordinateur, une console de salon, un téléphone, une tablette, où tout appareil électronique pouvant faire fonctionner un jeu.

A la notion d'interface sont également associées les notions d'entrées (en anglais *input*) et de sorties (*output*). Lors d'interactions avec un jeu vidéo, le joueur envoie des données en *input* qui vont être analysées puis traitées par le programme pour que de nouvelles données soient transmises au joueur en *output*.

Les « règles déterminées » mentionnées par la définition du Larousse sont celles qui régissent les possibilités et les modalités de liaison du joueur à l'interface.

Par exemple, une des règles du jeu ***Super Mario Bros***[12] décrit qu'en appuyant sur le bouton A de la manette de son ***NES***[13], le joueur fera sauter le personnage de

11 ***Dictionnaire étymologique de la langue française***, *Paris, Larousse Édition 2014.*

12 **Super Mario Bros**, *Nintendo, 1985.*

13 La ***Nintendo Entertainment System*** est une console conçue et commercialisée à partir de 1985 par la firme japonaise Nintendo. Dotée d'un processeur 8 bits offrant des capacités de calcul supérieures aux consoles précédentes, elle a participé au renouveau du jeu vidéo grâce à son catalogue très fourni, sa facilité d'utilisation ainsi qu'à ses graphismes et à ses

Mario à l'écran tandis qu'un son spécifique accompagnera ce bond. Nous déduisons de cet exemple que la manette permet au joueur d'envoyer des informations au programme tandis que l'écran et les haut-parleurs en fournissent d'autres en retour.

Nous appellerons par la suite *feedback* ces éléments d'*output*, qu'ils soient visuels ou sonores, qui sont émis conséquemment à une information envoyée par le joueur en *input*.

A ces règles, que nous appellerons **ergonomie**, et qui définissent les modalités d'interfaçage entre le joueur et le jeu, s'ajoute un ensemble de règles qui lui est intrinsèque.

Ce dernier, que nous appellerons l'***axiomatique*** du jeu, détermine non seulement les interactions entre les différents objets qui le constituent, mais aussi leurs conditions d'existence.

Pour reprendre l'exemple précédent de ***Super Mario Bros***, lorsque Mario saute sur l'ordre du joueur, l'**axiomatique** détermine la direction et la vitesse du personnage.

L'**axiomatique** d'un jeu décrit ses lois internes de fonctionnement. Elle définit un paradigme pour l'univers du jeu, quel qu'en soit le genre. Le succès du joueur dépendra de sa capacité à appréhender ces lois, pourtant a priori inaccessibles car inscrites dans le code du jeu. L'**axiomatique** est, pour chaque jeu, le pacte ludique devenu programme informatique. Elle est de fait ce qui va maintenir notre incrédulité suspendue jusqu'au triomphe du joueur sur le jeu.

---

fonctionnalités sonores.

L'ensemble formé par l'**ergonomie** et l'**axiomatique** constitue le *gameplay*.

Ce terme essentiel à l'étude d'un jeu peut être défini comme suit : le *gameplay* est un système de règles définissant toutes les interactions possibles entre le joueur et le jeu, qu'elles aient été prévues ou non par ses créateurs.

Étymologiquement, ce mot est décomposable en deux parties, issues de la langue anglaise : *game* et *play*. Pour Mathieu Triclot, elles sont associées à des concepts très distincts :

« On dira alors qu'il faut distinguer ce que le français confond alors que l'anglais l'autorise : les games *qui sont les dispositifs d'objets, les jeux avec leurs règles, et le* play *qui désigne l'activité protéiforme du jeu*[14]*. »*

Dans la continuité de cette citation, Mathieu Triclot s'appuie sur l'étude de jeux classiques par le philosophe Jacques Henriot et adapte aux jeux vidéo une des thèses de ce dernier : il n'y a pas de jeu sans joueur. L'activité vidéo-ludique, « l'expérience instrumentée » dont parle Mathieu Triclot, naît donc de la rencontre du joueur et du *gameplay*.

Pour Jacques Henriot, « *s'il y a jeu, le jeu n'est que dans l'attitude de l'acteur à l'égard de son acte*[15]. » C'est au regard de cette citation que s'éclaire notre définition de *gameplay*. En effet, par le détournement de l'*ergonomie o*u de l'*axiomatique* d'un jeu, le joueur peut en élargir l'horizon des possibles, et ce bien au-delà des prévisions de ses développeurs. En se réappropriant le *gameplay*, en s'y confrontant, en l'interprétant et en l'éprouvant, le joueur réinvente le jeu, donnant lieu à ce que nous appellerons le *gameplay émergent*.

---

*14 Mathieu Triclot, **Philosophie des jeux video**, Paris, La Découverte, Collection Zones, 2011, p. 24.*

*15 Jacques Henriot, **Le jeu**, Paris, Presses Universitaires de France, 1969, p. 73.*

Dans le but de proposer de nouvelles possibilités de *gameplay* basées sur la voix du joueur, *une* analyse préliminaire de la voix dans le jeu vidéo s'impose.

Nous étudierons donc dans un premier temps l'apparition de la voix au sein de la grammaire sonore vidéo-ludique.

Nous montrerons que si le texte a été le premier moyen de transmettre des messages au joueur, l'apparition de périphériques électroniques de synthèse vocale puis l'augmentation des capacités de stockage informatique vont permettre à la voix de devenir une source d'information incontournable.

Nous mettrons ensuite en évidence le rôle central de la voix dans l'évolution du jeu vidéo comme majoritairement narratif.

Ceci nous amènera à la formalisation de la grammaire vocale vidéo-ludique autour de trois pôles essentiels : Mécanique, Narration et Immersion. Nous les décrirons en nous appuyant sur l'analyse des écritures vocales de différents jeux vidéo.

Enfin, nous étudierons les premières applications au jeu vidéo de technologies de détection et de reconnaissance vocale. Nous aborderons les principes techniques et mathématiques de ces technologies pour ensuite focaliser notre attention sur les possibilités offertes par l'analyse du contenu spectral d'un signal vocal. Nous nous appuierons pour ce faire sur les parties pratique et expérimentale de ce mémoire de master. Les travaux menés dans ces deux parties ont pour objectif l'**élaboration d'un prototype de jeu vidéo se basant sur l'analyse spectrale de la voix du joueur comme mécanique principale de *gameplay***.

La partie pratique documente la conception puis la programmation de ce jeu, ***v0x***, tandis que la partie expérimentale vise à caractériser l'outil d'analyse utilisé.

Si des éléments de réflexion issus de ces deux démarches nourrissent le corps principal de ce mémoire de master, deux dossiers spécifiques décrivent en annexes la mise en œuvre de ces différents travaux.

CHAPITRE 1 : Histoire de la vocalité vidéoludique

DES CAMPUS AUX SALLES D'ARCADE

Le premier jeu vidéo est totalement silencieux

Illustration 1: Deux de ses développeurs jouent à **Spacewar***.*

Programmé sur un ordinateur PDP-1 par des étudiants du Massachusetts Institute of Technology en 1962, ***Spacewar*** est un jeu d'un genre nouveau. Appelée « sport » par un journaliste[16], faute d'un meilleur terme, la pratique de ce jeu se démocratise très rapidement au sein des campus universitaires américains. De plus en plus de PDP-1 sont utilisés et testés au sein des laboratoires d'électronique. Cet ordinateur à transistor, inspiré du TX-0, premier appareil de ce type, présente une interface de saisie du code accompagné par un écran permettant le contrôle du bon fonctionnement du programme. L'émission de sons par la machine n'est pas envisageable au regard de l'architecture de l'ordinateur et de son jeu d'instruction.

16 *D.J Edwards et J.M Graetz,* **« PDP-1 plays at Spacewar, » DECUSCOPE,** *vol. 1, n° 1, avril 1962, p.2*

Il faut attendre 1971 et l'assimilation du phénomène étudiant ***Spacewar*** par l'industrie des jeux d'arcade pour que le jeu vidéo devienne sonore. Plus particulièrement, *Nutting Associates*, spécialisée dans la conception et la commercialisation de flippers, va développer ***Computer Space***, le premier jeu vidéo d'arcade produit en série et massivement commercialisé. Des sons minimalistes accompagnant les mouvements du vaisseau sont générés par des associations de composants électroniques au sein de la deuxième carte mémoire de la machine. La façon de penser les flippers au sein de leur environnement de jeu contamine profondément ***Computer Space***. Dans une salle d'arcade, un jeu se doit d'être bruyant et accrocheur pour attirer l'attention plutôt qu'un autre jeu voisin. Le jeu de *Nutting Associates* hérite de cette esthétique spectaculaire, presque agressive.

En 1972, ***Pong*** atteint les salles de jeux. Conçu par Ralph Baer puis réadapté par Nolan Bushnell qui le commercialise sous la forme d'une borne d'arcade, ce jeu de ping pong minimaliste intègre la diffusion d'un premier son à chaque fois que la balle touche une raquette ou une paroi de l'écran et d'un second son lorsqu'un point est marqué. Peu avant le lancement de la borne, Nolan Bushnell demande à Al Alcorn, brillant ingénieur informatique en charge du projet, d'implémenter ces fonctions. Alcorn utilisera finalement des segments du circuit pour qu'ils émettent les fréquences désirées via des haut-parleurs, sur commande de l'horloge de synchronisation du programme[17].

17 *Karen Colins,* **Game Sound,** *MIT Press, 2008, p. 8-9.*

LES SALONS S'ANIMENT

Cinq ans plus tard, en 1977, Atari commercialise une console de salon, le *Video Computer System*, qui rencontre un succès foudroyant. En 1980, ***Space Invaders***, sa musique minimaliste et ses sons de tirs s'invitent dans plusieurs millions de foyers américains. Les Programmable Sound Generators, des processeurs dédiés à divers types de synthèse sonore, apparaissent au sein d'un nombre croissant de jeux d'arcade et sur les cartes électroniques du *VCS* et de ses principales concurrentes : l'*Intellivision* et la *Colecovision*.

Mais c'est en 1982, avec le module de synthèse vocale *Intellivoice*, conçu par Mattel pour sa console *Intellivision*, que les jeux vidéo se mettent à parler au joueur.

Illustration 2: L'Intellivision de Mattel

Illustration 3: L'Intellivoice. Le connecteur sur la gauche du périphérique permet son

L'*Intellivoice* prend la forme d'un adaptateur modifiant le port réservé aux cartouches de jeu classiques pour augmenter les capacités de calcul et de traitement

de la console. Grâce à une puce de synthèse vocale SP0256-012 de General Instruments, des jeux programmés spécifiquement pour ce matériel peuvent exploiter les nouvelles possibilités sonores de l'*Intellivision*.

Dans la mémoire morte de la puce SP0256-AL2, dont la SP0256-012 est inspirée, se trouvent des instructions pour émettre cinquante-neuf sons différents pouvant, par combinaison, former des mots de la langue anglaise. Le langage est ici décomposé en phonèmes, c'est à dire unités sonores minimales permettant de distinguer les différents mots d'une langue. Par exemple, le phonème [p] et le phonème [b] permettent de différencier les mots « pas » et « rat » en français. Mais si le phonème est une unité de son minimale, il n'indique pas nécessairement les méthodes de prononciation du son, ce qui aboutit à la production d'allophones, c'est-à-dire de différentes réalisations sonores d'un même phonème.

Les cinquante-neuf sons inscrits dans le jeu d'instruction du SP0256 sont des allophones de la langue anglaise permettant d'en prononcer les phonèmes de différentes façons en fonction du mot à synthétiser. En nous référant au manuel d'utilisation du processeur, on observe que le mot *alarm* est formé par la synthèse successive des allophones notés *AX*, *LL*, *AR* puis *MM*. Les sons sont un signal audionumérique généré par un circuit de modulation de largeur d'impulsion (en anglais Pulse Width Modulation) converti en un signal analogique grâce à une série de filtres passe-bas. La bande passante de l'appareil s'étend officiellement de 0 Hz à 5 kHz, avec une dynamique maximale de 42 dB et un rapport signal sur bruit d'approximativement 35 dB[18].

18 *Documentation du SP0256.*

Au sein de l'*Intellivoice*, le processeur SP0256-012 se base sur un même principe, mais supplante ces cinquante-neuf allophones par des mots spécifiques à Mattel et aux cinq jeux qui seront développés pour ce support : ***Space Spartans***, ***B-17 Bomber***, ***Bomb Squad***, ***Tron : Solar Sailer*** en 1982 et ***World Series Major League Baseball*** en 1983.

Grâce à l'utilisation de la mémoire morte des cartouches de jeu de l'Intellivision, le jeu d'instruction de synthèse du SP0256-012 peut être étendu : pour la première fois, des voix sont enregistrées pour un jeu vidéo puis converties en un programme qui permettra à la puce de les réinterpréter et de les synthétiser.

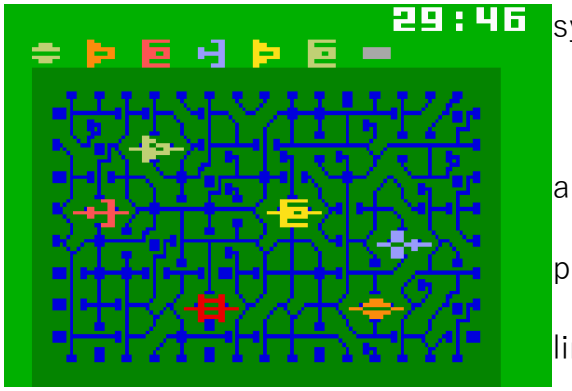


Illustration 4: Un des écrans de **Bomb**

Dans ***Bomb Squad***, un opérateur indique au joueur comment désamorcer un explosif puissant. Les différentes phrases sont en nombre limité et sont essentiellement des indications sur la méthode de désamorçage à adopter ou des éléments vocaux de *feedback* lors de la fin du jeu, qu'elle soit heureuse ou non.

Mais l'espace de stockage accordé à ces éléments vocaux étant extrêmement restreints, les développeurs doivent utiliser une fréquence d'échantillonnage la plus basse possible voire jongler au sein d'un même mot entre différentes fréquences. Les voix sont distordues et nécessitent un effort de compréhension. L'*Intellivoice* ne rencontre pas le succès commercial espéré par Mattel malgré le caractère visionnaire de l'appareil.

Il faut attendre l'avènement de Nintendo pour que le jeu vidéo prenne une toute autre ampleur et délaisse véritablement l'arcade au profit du salon.

Au départ société familiale d'édition de jeux de cartes et de jouets originaire de Kyoto, Nintendo choisit à partir des années 70 de diversifier ses activités et investit dans les jeux d'arcade et les consoles de salon.

En 1983, le *Famicom* (pour Family Computer), sort au Japon. Il y rencontre un grand succès grâce aux performances de son processeur 8-bits et à son prix accessible.

Deux ans plus tard, le *Nintendo Entertainment System*, version occidentale de la console japonaise frappe l'Amérique du Nord et l'Europe comme une déferlante, porté par un catalogue de jeux triés sur le volet par la firme japonaise. Après l'effondrement du marché de l'arcade en 1984[19], Mario et Link (***The Legend of Zelda***, Nintendo, 1986) viennent à la rescousse du jeu vidéo moribond.

19 William Audureau « 1984 : quand le jeu vidéo a évité le K.O. » JV, n° 3, janvier 2014, p. 86-91.

Illustration 5: Un NES.

Les deux machines embarquent un PSG : le Ricoh 2a03 pour les versions NTSC et le Ricoh 2a07 pour les versions PAL. Conçu par le compositeur Yukio Kaneoka, ce processeur est surtout utilisé pour la génération d'effets sonores et de la synthèse musicale mais dispose, comme le SP0256 de capacités de synthèse vocale[20]. L'un des cinq canaux sonores du Ricoh 2a0X, le canal de Modulation Delta est un échantillonneur autorisant deux méthodes de synthèse. Si la première , appelée *direct memory access* utilise des sons quantifiés sur un *bit* unique pour générer des effets sonores ponctuels, la deuxième, la modulation d'impulsion codée (PCM), permet la création de signaux vocaux utilisés par exemple pour les sons de foule du jeu de boxe ***Mike Tyson's Punch Out !*** (Nintendo, 1987).

20 *Documentation du Ricoh 2a07.*

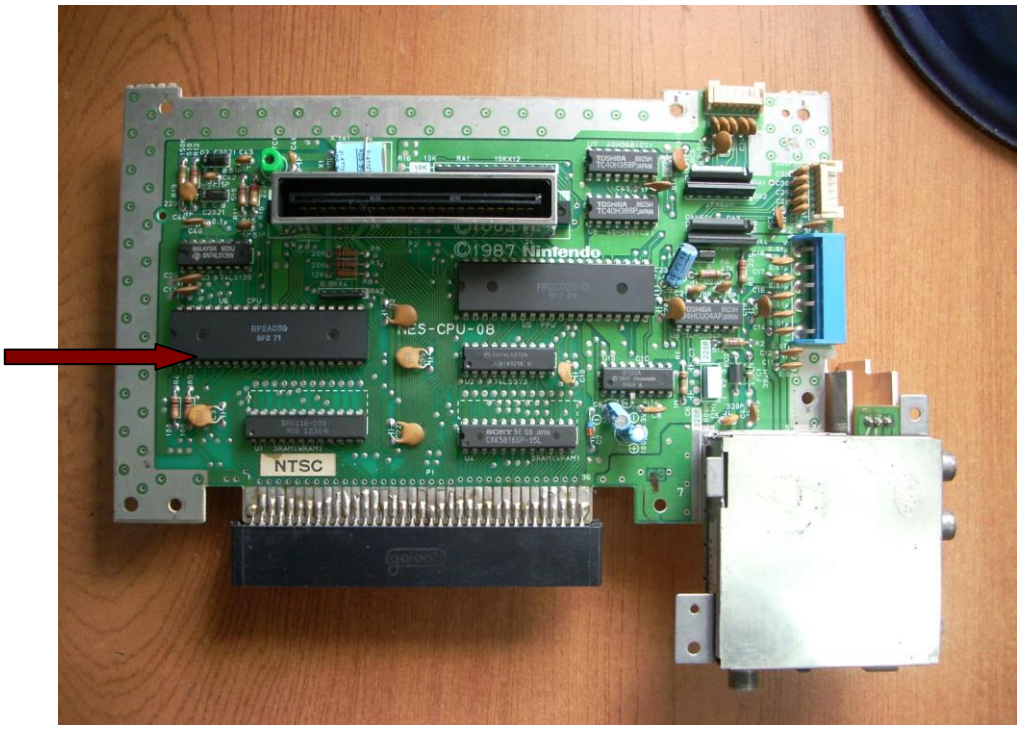


Illustration 6: Une carte mère de NES dans sa version NTSC. La position du Ricoh 2a03 est indiquée par une flèche rouge.

Malgré ces évolutions technologiques, la voix est à cette époque le parent pauvre du son de jeu vidéo. Héritée de l'arcade, l'esthétique sonore vidéo-ludique vise avant tout à attirer le joueur et à le stimuler continuellement pendant le jeu pour l'encourager à poursuivre sa partie et introduire une pièce supplémentaire dans la machine. La composition de mélodies énergiques et aisément mémorisables, la reprise de thèmes musicaux célèbres sont les principaux objectifs des développeurs en charge du *sound design* des jeux d'alors[21]. La série de jeux ***Megaman***, initiée sur *NES* par Capcom en 1987 est symptomatique de cette tendance. Dès leur premier opus, ces jeux s'imposent à la fois comme pionniers du jeu de plate-forme mais font également chanter le *NES* comme jamais auparavant. Néanmoins, ***Megaman*** restera mutique jusqu'au huitième épisode de la série, publié en 1998 sur la *Playstation* de Sony et sur

21 *Karen Collins, op. cit., p. 26.*

la *Saturn* de Sega.

Avec leurs distorsions très audibles et leur intelligibilité limitée, les signaux vocaux générés par les *PSG* risquent de faire sortir le joueur du jeu, rompant de fait le pacte ludique. Les voix se font donc discrètes, voire symbolistes. Une interjection, un gémissement ou un cri sont remplacés par d'autres sons dont les modes de production sont identifiés comme électroniques. Mais dans la mesure où ceux-ci ne sont pas produits par la mise en vibration d'un milieu par l'appareil phonatoire d'un individu, peuvent-ils toujours être appelés voix ?

Pour Bruno Bossis, « rien ne s'oppose [⋯] à ce qu'un son puisse être perçu comme vocal bien que son origine ne soit pas liée à l'organe de la phonation[22]. » Cependant, il insiste sur la nécessité de différencier ces sons produits par des méthodes différentes :

« Par ailleurs, lorsqu'un son est produit à l'aide d'une machine, quelle que soit sa technologie, il doit être considéré comme artificiel, même s'il s'agit d'un son naturel transformé par cette machine[23]. »

Pour résoudre ce problème de terminologie, Rossis propose d'associer à ces sons artificiels le concept de vocalité artificielle, en décrivant la vocalité comme la « qualité vocale d'un événement sonore[24]. »

Dès lors, il apparaît que le jeu vidéo est éminemment illusionniste : il veut attacher à des corps fictifs des vocalités que nous ne leur imaginons pas.

Néanmoins, aux yeux de Bruno Bossis, cette artificialité est loin d'être un

22 *Bruno Bossis,* **La voix et la machine : la vocalité artificielle dans la musique contemporaine,** *Presses Universitaires de Rennes, 2005, p. 8.*

23 *Bruno Bossis, op. cit., p 8.*

24 *Bruno Bossis, op.cit., p 8.*

défaut :

« La vocalité artificielle interpelle ainsi l'imaginaire, d'autant plus qu'elle élargit le champ de la vocalité naturelle et détourne les limites de la voix musicale[25]. »

Cette remarque, bien qu'émise à propos de la vocalité artificielle dans la musique contemporaine, peut s'appliquer au jeu vidéo. Les sonorités synthétiques, les distorsions métalliques liées à une fréquence d'échantillonnage trop faible, le nombre réduit de variations de volume possibles… Tous ces artefacts liés aux caractéristiques électroniques des premières consoles de salon vont peu à peu faire partie de leur identité sonore, au point d'être mis en œuvre plus tard pour caractériser la voix de robots ou d'intelligences artificielles, des personnages récurrents dans la science-fiction notamment.

Les limitations techniques des consoles empêchent les personnages de se faire entendre. Malgré les tentatives que nous avons décrites, le texte reste donc le vecteur privilégié d'informations, qu'elles soient liées aux règles du jeu ou à son histoire.

Pour autant, des animations font bouger les lèvres de pixels des personnages. Le *NES* voit également la multiplication du nombre de jeux d'aventure inspirés de ***The Legend of Zelda*** mais aussi de jeux de rôle, comme ***Final Fantasy*** (Square, Nintendo, 1987) ou ***Dragon Quest*** (Enix, Nintendo, 1986). Portés par des scénarii denses et complexes, ces jeux font preuve d'un intérêt particulier pour l'écriture de dialogues et sont peuplés d'interlocuteurs potentiels. Comme dans le cinéma sourd de Michel

25 Bruno Bossis, op. cit., p 185.

Chion[26], les voix sont présentes en creux dans les premiers jeux vidéo, de ceux du *VCS* à ceux du *NES* mais les personnages en sont réduits à balbutier.

Illustration 7: Capture d'écran de **The Legend of Zelda**, *sur NES*

DE L'USAGE DES PREFIXES *SUPER* ET *MEGA*

L'arrivée des consoles à processeurs 16-bits change considérablement la donne.

Inspirés par les dernières inventions du marché mourant de l'arcade, les ingénieurs de SEGA intègrent à partir de 1985 une puce de synthèse FM dans une nouvelle version de leur console 8-bits, la *Master System*. Si cet appareil n'est pas commercialisé hors du Japon et que cette puce est optionnelle sur les autres modèles de *Master System*, il préfigure la *Megadrive* (ou *Genesis*), première console 16-bits remarquable disponible au Japon dès 1988.

En terme de performances, la machine de SEGA est loin devant le *NES*. Un catalogue de jeux prestigieux, valorisé par une campagne publicitaire intense, parfois

26 *Michel Chion,* **Un art sonore, le cinéma**, *Cahiers du cinéma, 2010, p. 11-25.*

agressive[27], permet à la *Megadrive* de s'installer chez de nombreux joueurs qui découvrent ***Ghouls n' Ghosts*** (Capcom, 1988), ***Golden Axe*** (Sega, 1989) ou ***Sonic the Hedgehog*** (Sega, 1991). Sans concurrence pendant trois ans, la console de SEGA rencontre un succès foudroyant.

Illustration 8: La Megadrive, nommée Genesis en Amérique du Nord

Autour du processeur 16-bits de la *Megadrive*, un 68000 CPU de Motorola, se trouve un processeur moins puissant : le Zilog Z80. Ce dernier permet avant tout le pilotage électronique de deux circuits totalement dédiés au son : la puce de synthèse FM YM2612FM de Yamaha et le PSG SN76489A de Texas instruments[28].

Les éléments vocaux des jeux *Megadrive* sont plus nombreux et d'une bien meilleure qualité que ceux de la génération de console précédente. La puissance de calcul considérablement plus élevée de la machine permet une grande diversité de sons. Néanmoins, leur production passe par de fastidieuses phases de programmation en assembleur, un langage de programmation de très bas niveau.

Le sixième canal FM du module YM2612FM habituellement dédié à de la

27 « Genesis does what Nintendon't. »

28 Documentation du CPU 68000.

synthèse sonore peut être temporairement dédié à la lecture de sons enregistrés dans un format numérique PCM 8-bits. Les possibilités sonores offertes par l'électronique de la *Megadrive* amènent un véritable renouveau dans la façon de composer de la musique de jeu vidéo et permet l'implémentation toujours plus fréquente d'éléments vocaux.

Malgré la réussite du *NES*, Nintendo redoute un monopole de SEGA. Dès 1991, le *Super Nintendo Entertainment System* est commercialisé et s'affirme comme un concurrent de la *Megadrive*. La concurrence est d'autant plus féroce que les capacités techniques du *SNES* sont supérieures à celles de la console de SEGA. Pour les fonctionnalités sonores en particulier, Nintendo opte pour des technologies radicalement différentes. A la synthèse FM, le *SNES* préfère la synthèse par tables d'ondes[29].

Illustration 9: Une version japonaise du SNES, *le* Super Famicom.

Ce procédé, créé par Wolfgang Palm en 1979, permet la génération de multiples sons grâce à des instruments synthétiques pré-programmés et inscrits dans la

29 Karen Collins, op. cit., p. 45-47.

mémoire de l'appareil. La synthèse par table d'ondes permet la simulation de nombreux instruments, y compris des modules de synthèse classiques.

Le *SNES* se distingue de sa concurrente par la présence au sein de sa carte mère d'une *Audio Processing Unit* autonome. Autrement dit, cette APU fonctionne comme un second processeur, presque indépendant de l'unité centrale de la console, le Ricoh 5A22.

La synthèse par table d'ondes est mise en œuvre par l'articulation de trois composants électroniques créés par Sony spécialement pour la console : le SPC700, le S-DSP ainsi qu'un convertisseur numérique-analogique 16-bits.

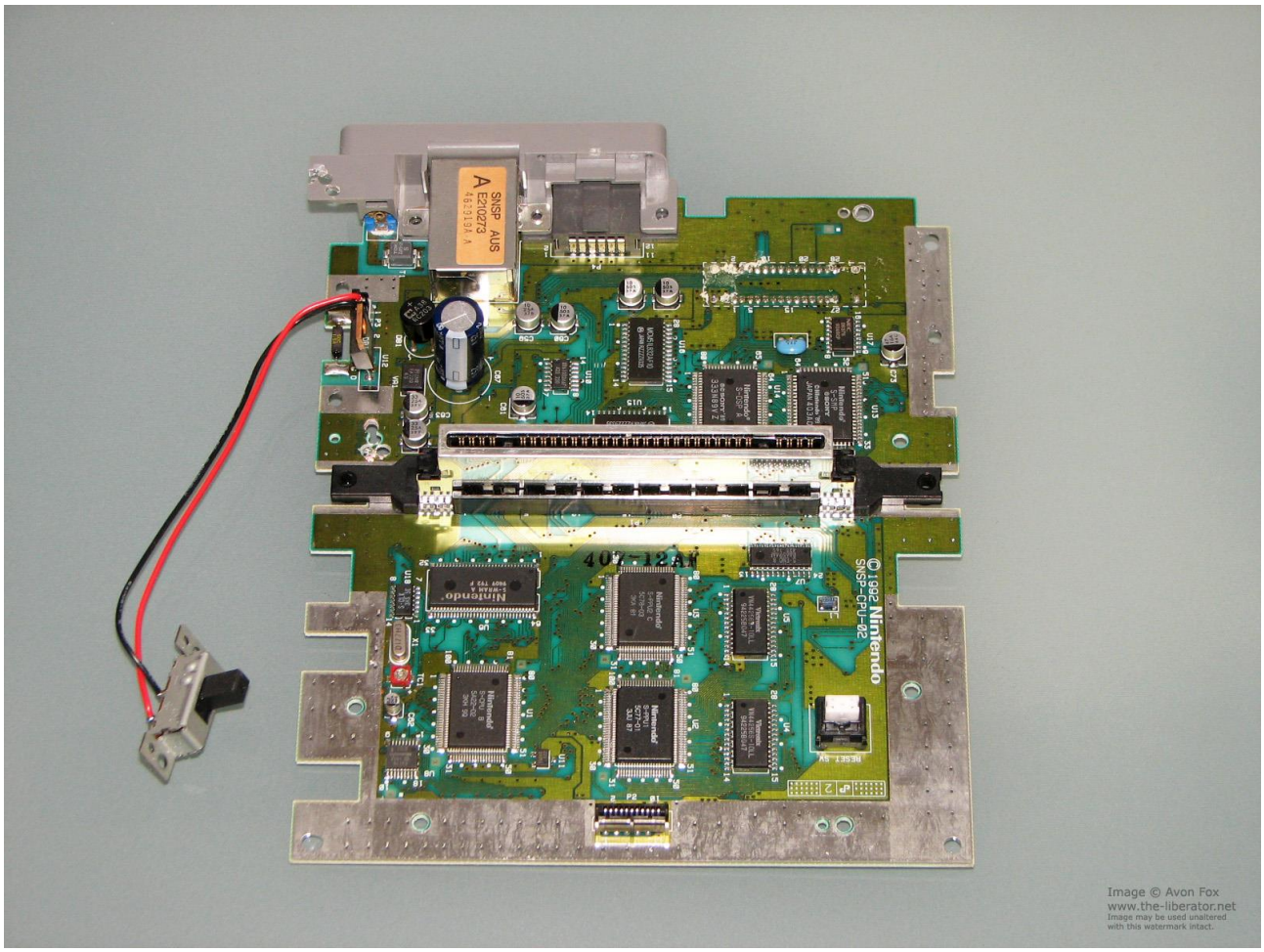


Illustration 10: La carte mère du SNES. *En rouge, le S-DSP. En jaune, le SPC700. En bleu, le RIcoh 5A22.*

Si ce dernier permet la lecture de sons enregistrés puis numérisés au préalable, il se distingue du convertisseur de la *Megadrive* en utilisant un format d'écriture de fichier propriétaire. Afin d'optimiser l'exploitation des performance du *SNES*, Nintendo développe les fichiers .SPC qui permettent d'enregistrer différents états de l'APU et donc de gérer dans le temps les adressages mémoire des données sonores. Cette innovation est rendue possible par l'adoption d'un système de compression des données sonores. Plutôt que de la modulation d'impulsion codée, les ingénieurs de Nintendo créent une méthode de compression avec perte : la *Bit Rate Reduction* ou réduction du débit binaire.

Le concept fondamental de la BRR est la modulation par impulsion et codage différentiel adaptatif, ou ADPCM, développée pour le domaine de la téléphonie par les laboratoires Bell dans les années 1970. Conçue pour l'encodage et le transport numériques de données vocales, l'ADPCM utilise des pas de quantification variables pour que seules les informations essentielles bénéficient d'une résolution maximale.

Dans un esprit similaire, la BRR transforme une séquence de 16 échantillons sonores PCM quantifiés sur 16 bits, soit 32 octets de données non compressées pour des informations stéréo en 9 octets de données compressées. Un premier octet d'en-tête comporte quatre bits indiquant le pas de quantification adopté pour cette séquence de données, suivis de deux bits déterminant les filtres à utiliser lors du décodage puis deux bits de contrôle à destination du SPC700. Les huit octets suivants sont 16 paquets de quatre bits qui décrivent les données sonores.

L'articulation du Ricoh 5A22, du SPC700 et du S-DSP est la réponse des

développeurs de Nintendo et des ingénieurs de Sony à la contrainte écrasante que constitue la faible capacité de stockage des consoles et des cartouches de jeu.

Pour la première fois, une console de salon génère des musiques et des vocalités réalistes. La qualité technique, par rapport à l'époque, des instruments simulés ainsi que des vocalités artificielles diffusées, permet au *SNES* de se démarquer des consoles précédentes. Pour Karen Collins, c'est durant l'ère 16-bits que le son de jeu vidéo gagne ses lettres de noblesse et devient un véritable enjeu esthétique et narratif[30].

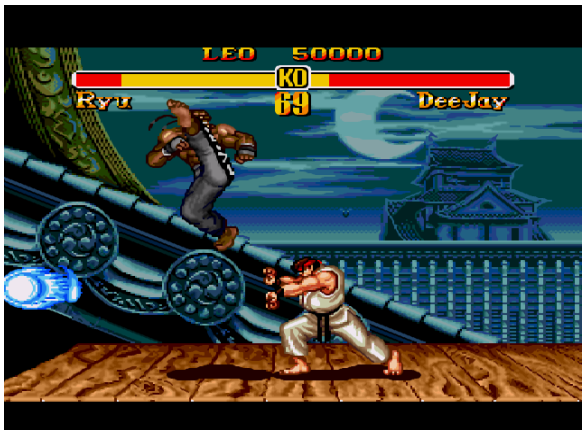


Illustration 11: Capture d'écran de **Super Street Fighter II** *sur SNES*

Illustration 12: Capture d'écran de **Mortal Kombat**, *sur SNES*

Les personnages des jeux de combat les plus célèbres de cette époque, ***Super Street Fighter II*** (Capcom, 1991) puis ***Mortal Kombat*** (Midway, 1992), ont chacun un *gimmick* vocal qui complète leur panoplie de coups pour étoffer leur identité.

Mortal Kombat s'illustre également par la présence d'un commentateur qui encourage le joueur à achever son adversaire d'une façon particulièrement sanglante au terme d'un combat très violent. La phrase « Finish him !», devient une des marques de fabrique de cette série de jeux en symbolisant le plaisir transgressif et cathartique que ces affrontements virtuels peuvent procurer.

30 *Karen Collin, op. cit., p. 59-61.*

La grammaire sonore des jeux de *versus fighting* est intégralement inventée par les développeurs de Capcom et Midway. Elle est encore aujourd'hui la même.

Néanmoins, les genres éminemment narratifs qui apparaissent ou s'affirment à l'ère 16-bits préfèrent des dialogues écrits à des dialogues enregistrés pour véhiculer l'intrigue du jeu. Les jeux de rôle et d'aventure d'inspiration nipponne se multiplient. Les séries ***Zelda***, ***Final Fantasy*** et ***Dragon Quest*** s'étoffent tandis que de nouvelles licences comme ***Secret of Mana*** ou ***Chrono Trigger*** apparaissent. Néanmoins, les personnages de ces jeux sont au mieux inaudibles lorsqu'ils ne sont pas complètement mutiques.

Il faudra attendre l'intégration des lecteurs optiques et l'édition de jeux vidéo sur CD-ROM pour assister à une nouvelle évolution majeure de la vocalité vidéo-ludique.

L'avènement des supports optiques

LA REVOLUTION COMPACT DISC

En 1979, Sony et Philips présentent un nouveau support de stockage de données numériques : le disque compact.

Désormais, des données peuvent être stockées sur un disque de polycarbonate puis lues par un lecteur laser. En 1982, le format de fichier *Compact Disc for Digital Audio* (*CD-DA*) est décrit dans le *Red Book,* un document technique de Philips, ce qui aboutira à l'établissement de la norme IEC60908. Dès sa création, le CD est destiné à

accueillir des données audio. Les 700 Mo de stockage permettent d'accueillir jusqu'à 80 minutes de son encodées sur deux canaux PCM échantillonnés à 44,1 kHZ et quantifiés sur 16 bits. Très rapidement, le CD est adopté par les leaders de l'industrie musicale qui délaissent peu à peu le disque vinyle.

Conçu au départ pour stocker des oeuvres musicales numérisées, le nombre de types d'information différents que le CD est en mesure d'accueillir va augmenter avec la création d'extensions du format de fichier CD-DA.

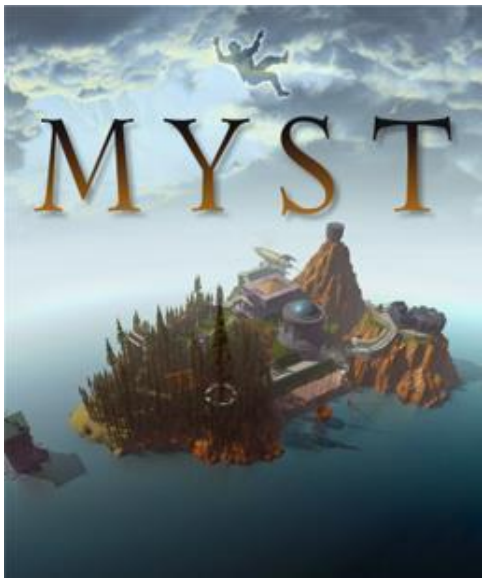
En 1988, Philips et Sony publient le *Yellow Book*. Ce document technique décrit les caractéristiques du CD-ROM. Dès lors, tous types de données peuvent être stockées sur un disque compact. Cet apport technologique va propulser le jeu vidéo sur console de salon, mais aussi sur ordinateur personnel.

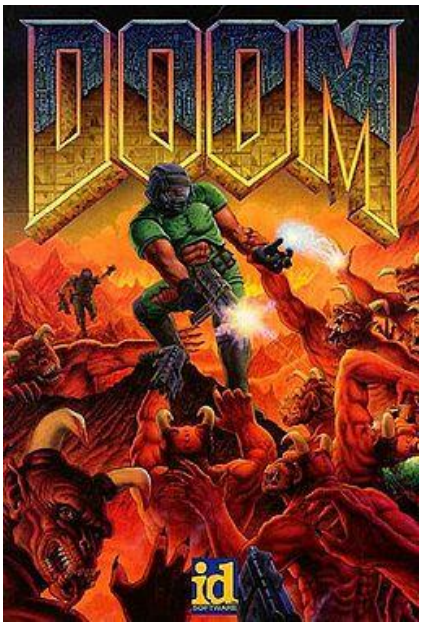
Dès décembre 1988, le *PC-Engine* de NEC Corporation est la première console pouvant communiquer avec un lecteur optique. Mais la puissance de calcul de la machine de NEC Corporation ne tire pas réellement profit de ce périphérique.

Peu après, à partir de 1991 au Japon, la *Megadrive* reçoit elle aussi son lecteur optique externe. Se branchant sur le port cartouche, le *Mega-CD* augmente considérablement les capacités de la console 16-bits de Sega. Mais son prix exorbitant limite son succès à un public de niche fortuné ou particulièrement passionné.

D

EUX NOUVEAUX ETENDARDS : ***MYST*** ET ***DOOM***





Il faut se tourner vers les ordinateurs personnels pour comprendre ce que l'arrivée du CD-ROM apporte au jeu vidéo, et en particulier au travail des voix.

Pour Karen Collins, la démocratisation des CD-ROM sonne le glas des autres systèmes de synthèse et lecture de sons. Plus particulièrement, les systèmes de synthèse par table d'ondes ou par le biais d'instruments MIDI sont délaissés[31]. Cela présente de nombreux avantages aussi bien pour les développeurs que pour les joueurs. Les jeux deviennent de fait plus accessibles financièrement.

Pour les joueurs, il n'est plus nécessaire de se procurer une carte d'extension dédiée et coûteuse afin d'augmenter les capacités de calcul de son ordinateur personnel.

Pour les développeurs, le besoin de mettre au point des solutions matérielles

31 *Karen Collins, op. cit., p. 69.*

prenant la forme de périphériques supplémentaires ou de dépendre du *hardware* développé par d'autres entreprises ne se fait plus sentir, ce qui allège considérablement les coûts de production.

Enfin, le fait que de plus en plus d'ordinateurs soient capables de lire les CD-ROM permet aux créateurs de jeux de prévoir comment va être restitué le son de leur jeu, quasiment indépendamment de la machine du joueur, grâce aux normes spécifiques d'encodage et de décodage des informations sur un CD-ROM. Dès lors, le *voice acting* pour le jeu vidéo, c'est-à-dire l'enregistrement d'acteurs vocaux incarnant des personnages d'un jeu, devient également de plus en plus répandu.

Le début des années 90 voit l'apparition de jeux matriciels.

En 1993 les ordinateurs personnels accueillent deux jeux qui vont, à bien des égards, changer la façon de penser le jeu vidéo : ***Myst*** (Cyan Inc.) et ***Doom*** (id Software).

Pour Lev Manovich, ces deux jeux incarnent comme nul autre jeu avant eux une des tendances des médias à partir de l'apparition des technologies numériques : la création d'un espace navigable dans lequel l'utilisateur - ici le joueur - va évoluer[32].

Le CD-ROM n'est cependant pas la seule évolution. Si les capacités de stockage des machines augmentent considérablement grâce à ce support, les performances de celles-ci sont aussi en perpétuelle amélioration. En particulier, le début des années 90 est marqué par le début de l'intégration de cartes graphiques autorisant l'affichage de mondes virtuels en trois dimensions. De nombreux studios de développement

32 Lev Manovich, **Le langage des nouveaux médias**, *Les Presses du Réel, 2010, p. 432-443.*

s'engouffrent dans la brèche et délaissent les *sprites* 2D au profit de modèles 3D constitués de polygones.

Avec l'ouverture de l'espace à une nouvelle dimension, le son devient une source abondante d'informations : la localisation des sons qui composent le paysage sonore d'un niveau indique au joueur la position de ses objectifs, de ses cibles, des sources potentielles de danger.

Si ***Wolfenstein 3D*** (id Software, 1992) était un pionnier des jeux de tir en 3D à la première personne, son moteur son était très peu performant : il n'autorisait pas la lecture simultanée de plusieurs sons. ***Doom*** est, à tout point de vue, un parachèvement du jeu précédent d'id Software.

Illustration 13: Capture d'écran de **Doom**.

Le scénario de ***Doom*** abandonne l'humour noir qui traverse celui de

Wolfenstein 3D. Plutôt que des soldats nazis, le joueur affronte une horde de démons arrivés sur une station scientifique martienne suite à une expérience catastrophique sur la téléportation.

Doom conserve néanmoins la nervosité du *gameplay* de son aîné. Le joueur déambule à grande vitesse dans des niveaux labyrinthiques. Les armes qu'il utilise sont de plus en plus dévastatrices et les ennemis de plus en plus puissants et monstrueux. Le joueur, bien qu'acteur de sa montée en puissance tout au long du jeu reste dans un état de vigilance constante. En effet, les démons les plus puissants représentent une menace terrible. Leurs attaques peuvent venir au bout du joueur très rapidement. À la nervosité du jeu s'ajoute une exigence vis-à-vis du joueur : ses faux pas sont sanctionnés par des pièges mortels et par la férocité des ennemis.

Pour déjouer ces nombreux dangers, le joueur doit être tout ouïe : les sons qu'émettent les ennemis indiquent non seulement leur position dans l'espace, mais aussi leur type, donc leur niveau de dangerosité.

Dans une interview de 2006 citée par Karen Collins[33], Bobby Prince, compositeur et *lead sound designer* de ***Doom***, décrit la structure du moteur son du jeu :

> « There were several classes of sounds in ***Doom***. One was general active sounds that were not attached to any one demon. These were more or less ambient sounds, but they didn't play until demons close to the player ''woke up'' (usually based upon the player making some noise in the area). Then there were demon active sounds that were attached to individual demons. These sounds let the player know what class of demon was around the corner. Each type of demon had a sight sound

33 Karen Collins, op. cit., p. 65.

that played when the demon ‘‘saw’’ the player. There were also attack, hurt and death sounds particular to each type of demon. Another helpful thing about the sound driver was that the volume of sounds depended upon the distance from the player to the source of the sound. This helped keep the overall volume down during non- combat. It also stood to help scare the pants off the player when a demon in a dark niche woke up and immediately screamed his attack sound[34]. »

La lecture de cette interview révèle l'inventivité des développeurs de ***Doom***. L'intégration de techniques de spatialisation du son dans le moteur du jeu permet l'apport de nouvelles d'informations pour le joueur. Celui-ci est alors à la fois plus alerte, poussé à une vigilance de tous les instants, mais aussi par conséquent beaucoup plus impliqué dans le jeu. A l'affût de tous les éléments sonores qui pourrait lui permettre de survivre à la prochaine salle, le joueur progresse à pas de loup, les yeux rivés sur l'écran, une arme à la main.

Wolfenstein 3D et ***Doom*** sont deux jeux essentiels pour un genre qui apparaît avec eux : le jeu de tir à la première personne (ou *First Person Shooter*). Avec ce genre arrive un nouveau paradigme pour les jeux vidéo, celui de l'immersion. Les jeux à la première personne ou plus largement à focalisation interne deviennent légion suite au succès international de ***Doom***.

34 [Traduction] Il y avait différentes classes de sons dans ***Doom***. Une d'entre elles rassemblait les sons généralement actifs qui n'étaient associés à aucun démon. Ces sons étaient plus ou moins ambiants mais n'étaient pas lancés avant que des démons se 'réveillent' à proximité du joueur (en général lorsque ce dernier les attire en faisant du bruit dans la zone). Puis une autre classe de sons rassemble des sons directement rattachés à un démon en particulier. Ces sons indiquaient au joueur la classe du démon qui l'attendait au tournant. Chaque type de démon avait un son d'alerte spécifique qui se lançait une fois le joueur repéré par le démon. Il y avait également des sons d'attaque, de douleur ou de mort pour chaque type de démon. Un autre mécanisme du moteur son constituait une source d'indices pour le joueur. Le volume des sons dépendait de la distance entre le joueur et la source du son. Ceci aidait d'une part à maintenir le volume global du jeu à un niveau raisonnable hors des combats. Cela permettait d'autre part de faire bondir de peur le joueur lorsqu'un démon se réveille dans un recoin obscur puis hurle immédiatement son son d'attaque.

Aux cartes graphiques 3D s'ajoute ***DirectX***, une suite de librairies informatiques publiées par Microsoft à destination des développeurs. Dans une volonté de rendre la programmation de jeux de plus en plus accessible, ***DirectX*** rassemble différentes interfaces de programmation permettant d'optimiser les liens entre logiciel et matériel, et notamment entre logiciel et *hardware* audio.

Dès lors, les innovations techniques et les trouvailles de *gameplay* d'id Software ne tardent pas à faire école et à devenir des standards pour la plupart des FPS.

Avec ***Myst***, un nouveau genre voit le jour également. Ou plutôt, une nouvelle esthétique.

Myst est un jeu d'exploration à la première personne dans lequel le joueur incarne un voyageur hagard perdu dans un monde onirique. Pour progresser, il doit résoudre des énigmes en interagissant avec des éléments du décor, en assemblant des objets récupérés au gré de ses déambulations ou en récupérant des pages de différents livres. La seule chose susceptible de rapprocher le jeu de Cyan de ceux d'id Software est qu'ils sont tous trois à la première personne. Au-delà de ce point commun, ***Myst*** est l'antithèse de ***Doom***.

Pour Lev Manovich, « ***Doom*** et ***Myst*** diffèrent à maints égards. Le rythme du premier est rapide ; celui du second est lent. Dans ***Doom***, le joueur court dans les couloirs pour essayer de franchir chaque niveau le plus tôt possible et ainsi passer rapidement au suivant. Dans ***Myst***, le joueur se déplace dans le monde virtuel littéralement pas à pas, dénouant l'intrigue en cours de route. ***Doom*** est peuplé de

démons tapis dans tous les coins qui attendent le moment d'attaquer ; ***Myst*** est complètement vide[35]. »

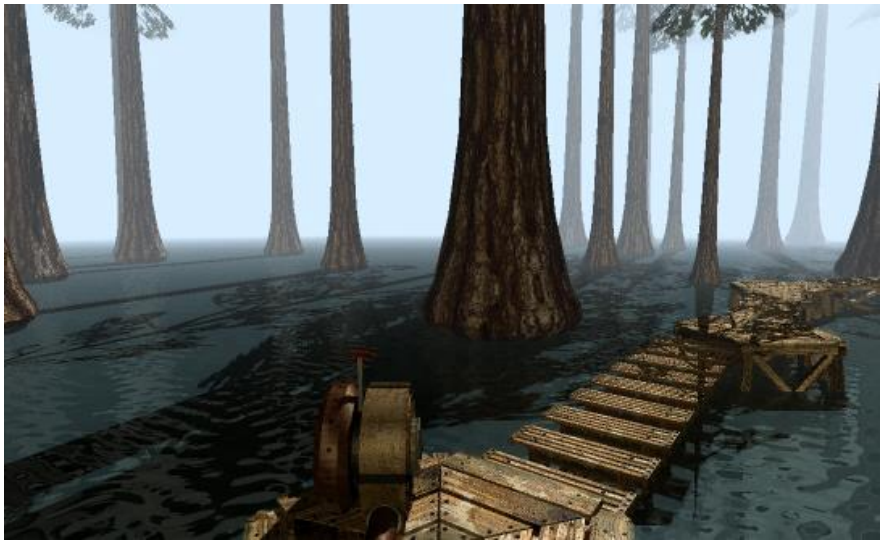


Illustration 14: Capture d'écran de **Myst**.

Si Manovich poursuit cette comparaison par la suite, sa dernière remarque est celle qui retient le plus notre attention. ***Myst*** est un monde sans habitants, que le joueur explore seul. Seules les voix des trois principaux personnages non-joueur, Atrus l'artisan et ses deux fils, Sirrus et Achenar, accompagnent occasionnellement le joueur dans son périple.

Si les objectifs de ***Doom*** sont immédiatement compréhensibles ; tuer les démons qui s'opposent à nous ; ceux de ***Myst*** se dérobent très souvent à notre regard. L'intrigue est diffuse et s'articule autour de la création de livres-mondes dans lesquels des individus peuvent être emprisonnés. Il appartient au joueur de plonger dans ces mondes pour en découvrir, ou plutôt décrypter, l'histoire d'Atrus et de ses fils.

Myst propose une nouvelle esthétique vidéo-ludique basée sur de l'exploration pure, sans autre enjeu, au point de relever de la contemplation. Les voix y résonnent

35 *Lev Manovich, op. cit., p. 433.*

comme autant d'invitations à la perte, à l'égarement, à l'errance.

Pour Michel Chion, cette invitation à la perte repose, dans le cinéma de Marguerite Duras et de Kenji Mizoguchi, sur le traitement des voix. Elles y sont souvent fuyantes, fugaces, éthérées.

Selon Michel Chion, « il semble [⋯] que la voix tienne ses effets de mystère, de pouvoir, de transcendance et jusqu'à son supposé tout-voir, au cinéma, de ce lieu du pas-tout-voir qu'est l'écran, qui la laisse jouer avec le leurre d'une puissance qu'il pourrait à tout moment lui reprendre[36]. »

La voix apparaît alors comme une instance quasi mystique, que Michel Chion n'hésite pas à rapprocher du Verbe de la philosophie ou de la religion, dont les pouvoirs ne se limitent pas à la simple énonciation de faits susceptibles de développer une intrigue.

Michel Chion poursuit : « La voix qui perd le cinéma, c'est une belle histoire, mais une histoire de cinéma. Le cinéma de la perte n'est pas une perte du cinéma, au contraire. N'est-ce pas dans les salles obscures qu'on aime venir à se perdre ? La voix, dans cette invitation à la perte, ne serait-elle que la plus séduisante des ouvreuses ?[37] »

La tentation de remplacer *cinéma* par *jeu vidéo* dans les deux premières phrases de cette citation est grande au regard de la proposition esthétique de ***Myst***.

Les possibilités techniques offertes par CD-ROM, savamment exploitées par Cyan, font de ***Myst*** le pionnier des jeux dans lesquels la contemplation est reine. C'est

---

36 Michel Chion, **La voix au cinéma**, *Cahiers du Cinéma, 1993, p. 113.*

37 Ibidem, p. 114.

également avec ***Myst*** que des voix s'élèvent pour suggérer que le jeu vidéo pourrait être une nouvelle forme d'art[38].

PLEASE INSERT DISC ONE

Les deux révolutions esthétiques que sont ***Myst*** et ***Doom*** ne tardent pas à se propager des ordinateurs personnels aux consoles de salon.

Après les premières tentatives de machines de salon pouvant lire des CD-ROM, la *Playstation* de Sony et la *Saturn* de Sega, commercialisées toutes les deux à partir de 1994, sont les fers de lance des consoles 32-bits. L'augmentation des capacités de stockage est accompagnée par la montée en puissance des machines de jeu.

Illustration 15: La SEGA Saturn.

Illustration 16: La Playstation.

La ***Saturn*** embarque huit processeurs dont deux processeurs 32 bits centraux Hitachi SH2 à jeu d'instruction réduits et un processeur audio fabriqué par Yamaha pour Sega, le SCSP (Saturn Custom Sound Processor), piloté par un processeur 8 à 16 bits Motorola 68EC000. Cet ensemble d'unités centrales de calcul permet à la *Saturn* de générer une grande variété de sons, qu'ils soient pré-enregistrés sur le CD-ROM du

38 http://www.nytimes.com/1994/12/04/arts/a-new-art-form-may-arise-from-the-myst.html

jeu ou synthétisés au sein de la carte mère de la console.

La carte mère de la *Playstation*, elle, est articulée autour d'un processeur central R3000 de MIPS Computer System. L'unité de calcul sonore est un processeur 16 bits conçu par Sony sobrement nommé SPU, dispose globalement des mêmes fonctionnalités que son équivalent au sein de la carte mère de la *Saturn*. Néanmoins, la *Playstation* se distingue de sa concurrente par sa capacité à décoder des données compressées, ce qui permet d'optimiser toujours plus l'utilisation des espaces de stockage sur le disque du jeu. A partir de sa publication en 1995, le codec de compression avec perte MPEG-2 Layer 3, plus connu sous le nom de MP3, devient un incontournable et représente un aboutissement de la logique d'optimisation à l'oeuvre ici. Bien que la qualité audio des données en soit amoindrie, l'introduction de ces techniques de compression permet aux jeux de présenter toujours plus de contenu audio.

LE CAS DE LA *NINTENDO 64*

Illustration 17: La Nintendo 64.

Sony et SEGA s'imposent rapidement sur le marché des consoles 32-bits, sans véritable concurrent. Nintendo ne commercialise l'héritière du *SNES*, la *Nintendo 64*, qu'à partir de 1996, mais se démarque très nettement des autres constructeurs. La firme japonaise propose d'une part une console 64 bits bénéficiant des performances du processeur VR4300 de NEC mais incapable de lire des CD-ROM d'autre part. Pour Nintendo, les cartouches présentent plusieurs avantages :

- L'intégration d'un lecteur de cartouche plutôt qu'un lecteur optique permet de limiter le prix total de la console pour le consommateur.
- Les cartouches de jeu, constituées d'unités de mémoire morte (*Read-Only Memory*) contenant les différents fichiers du jeu, permettent un accès aux données bien plus rapide que les CD-ROM.

La *Nintendo 64* se distingue également par l'absence de processeurs dédiés au son au sein de sa carte mère. L'intégralité des opérations sonores est effectuée par le VR4300 qui génère des sons à partir de données inscrites dans la cartouche et grâce à des systèmes de synthèse similaires à ceux du *SNES.*

Néanmoins, les choix matériels effectués par Nintendo sont paradoxaux. Si l'intégration d'un processeur 64 bits plus puissant que ses homologues permet l'affichage d'environnements en trois dimensions d'une grande qualité, l'absence d'unité de calcul dédiée au son et la faible capacité de stockage des cartouches apportent peu de nouveautés par rapport au *SNES* pour ce qui est des fonctionnalités sonores de la machine. Par ailleurs, les coûts de développement et de fabrication de jeux sur cartouches sont bien plus élevés que pour des jeux sur CD-ROM. Du fait du

nombre limité de studios de développement pouvant se permettre un tel investissement, le catalogue de jeux de la *N64* est beaucoup plus restreint que celui de la *Playstation* et par conséquent, la console repose sur des licences pré-existantes représentant un faible risque pour les investisseurs. Les plus grands succès commerciaux de la console mettent en scène les héros Nintendo des ères 8 et 16 bits - Mario, Link, Donkey Kong et Fox McCloud en tête - ou prolongent les univers d'oeuvres artistiques à grand succès (***Star Wars*** et ***James Bond*** pour ne citer qu'eux). Seule la nouvelle licence ***Pokémon***, responsable de l'explosion des ventes de la console portable de Nintendo, la *Gameboy*, s'impose comme un nouvel incontournable du catalogue de Nintendo.

La firme japonaise se mettra au pas par la suite, avec la nouvelle génération de consoles. Tout comme ses concurrentes, la *Playstation 2* (2000) de Sony et la *XBox* (2001) de Microsoft, la *GameCube* (2001) embarque un lecteur optique.

Dès lors, le principe de fonctionnement global des consoles et leur gestion des données audio n'évolueront plus. Seules les capacités de stockage des supports optiques et les performances des machines augmenteront au fil du temps, du CD-ROM au DVD puis du DVD au Blu-Ray.

La voix au cœur de la narration

L'ESTHETIQUE NINTENDO

Les jeux vidéo auraient-il eu quelque chose à dire au joueur si les supports de stockage optique ne s'étaient pas imposés de la sorte ? L'exemple de Nintendo et des cartouches à faible capacité de stockage de la *N64* nous laisse penser que ce n'est pas le cas.

Pour surmonter cette contrainte, Nintendo a néanmoins développé une esthétique vocale immédiatement reconnaissable pour ses trois licences phares : ***Mario, Legend of Zelda*** et ***Pokémon***. Ces séries de jeux, conçues pour toucher un public extrêmement large, se doivent d'être accessibles pour les jeunes joueurs. De fait, si les développements scénaristiques peuvent y être complexes et riches en rebondissements, les dialogues s'illustrent par leur clarté et leur accessibilité.

Ces choix ont une répercussion évidente sur la conception des éléments vocaux de ces jeux. En général colorés, joyeux et très dynamiques, les jeux Nintendo préfèrent à de longues et sérieuses phases de dialogues une esthétique proche du cinéma d'animation, de la bande dessinée ou du manga. Les onomatopées et les interjections sont reines, les sonorités vocales sont lumineuses et énergiques sans être clinquantes. La musique est l'élément sonore qui apporte généralement le plus de richesse aux bandes-son de ces jeux, avec une volonté de créer des mélodies mémorables dont l'évolution suit la progression du joueur.

Ces choix esthétiques, mis en œuvre dès ***Super Mario 64*** (1996) et ***Legend of Zelda : Ocarina of Time*** (1998), tous deux sur *Nintendo 64*, permettent de contourner les limitations techniques de la console et des supports de jeu.

Certains jeux édités par Nintendo se démarquent cependant de leurs formes

canoniques, avec un travail d'écriture de dialogues et des performances atypiques d'acteurs vocaux. ***Conker's Bad Fur Day*** (Rare, 2001), pour ne citer que lui, détourne les codes des jeux de plate-forme « mignons » ayant pour personnage principal un animal servant de mascotte au jeu (voire à une entreprise entière) et a pour marque de fabrique l'humour scatologique et décalé, la violence gratuite ainsi qu'un discours autoréflexif sur le jeu vidéo et ses codes.

Illustration 18: L'évolution graphique des jeux **Super Mario : Super Mario 64**, *sur N64.*

Illustration 19: **Super Mario Sunshine**, *sur GameCube.*

Illustration 20: Et **Super Mario Galaxy**, *sur Wii.*

Néanmoins, ***Conker*** arrive très tard, en fin de vie de la *Nintendo 64*. Les codes qu'il détourne ont déjà été établis tout au long de l'ère des consoles 32 bits.

DES FILMS DONT VOUS ETES LE HEROS ?

En effet, sur *Saturn* et *Playstation*, une révolution s'est opérée : l'abandon quasi-systématique des textes au profit de la voix afin de transmettre au joueur les éléments essentiels de l'intrigue. Dans la continuité des jeux d'aventure sur ordinateur dont ***Myst*** est lui-même un parachèvement, les personnages de jeu vidéo se mettent à parler sans retenue et deviennent les conteurs d'innombrables histoires.

A la base de ce phénomène se trouve un rapprochement de plus en plus affirmé et assumé entre le jeu vidéo et le cinéma. Les jeux ***Another World*** (Delphine Software, 1990) et ***Alone in the Dark*** (Infogrames, 1992), sortis tout d'abord sur ordinateur puis adaptés sur console, représentent l'avant-garde de ce rapprochement d'un point de vue visuel. Dans ces deux jeux, le joueur navigue au sein de différents plans d'ensemble composés et structurés comme autant de plans cinématographiques. Dans ces deux jeux, la progression du joueur est entrecoupée de séquences de transition mises en scène et découpées en différents plans. Ces séquences, appelées cinématiques (ou *cutscenes*), permettent, lorsqu'elles sont disposées dans l'introduction d'un jeu, d'en installer l'univers sans avoir recours à des séries de textes descriptifs ou d'images fixes.

Mais ***Another World*** et ***Alone in the Dark*** restent sourds[39]. La narration y est visuelle, textuelle ou sonore, mais en aucun cas vocale.

« SNAKE, TALK TO ME ! »

La métamorphose s'achève finalement sur l'impulsion d'un concepteur de jeu japonais : Hideo Kojima.

Passionné du septième art, Kojima rêve de devenir réalisateur mais cultive un intérêt pour les jeux vidéo. En 1986, il intègre le studio de développement et d'édition de jeux Konami. Dès 1987, la direction d'un projet de jeu partiellement inspiré par le

39 Michel Chion, **Un art sonore, le cinéma***, op. cit, p. 11-25.*

film ***La grande évasion*** (John Sturges 1963) lui est confiée. Appelé ***Metal Gear***, le jeu présente pour la première fois le personnage de Solid Snake, un soldat d'élite appelé à un grand avenir vidéo-ludique.

En 1998, la ***Playstation*** accueille ***Metal Gear Solid*** développé par une équipe de Konami sous la direction d'Hideo Kojima. De nombreuses *cutscenes* rythment la progression du joueur et permettent au concepteur japonais de développer un scénario dense et adulte. A priori sans surprise, celui-ci va se révéler extrêmement riche.

Au début du jeu, le soldat Solid Snake est envoyé pour libérer une installation nucléaire militaire tombée aux mains de l'escouade Fox Hound, composée de soldats génétiquement modifiés. Si le joueur s'attend, à en croire les éléments de scénario qui lui sont donnés en préambule, à une progression résolument linéaire dans la base militaire, ponctuée de duels contre les membres de l'escouade Fox Hound, il voit peu à peu ses attentes contrariées.

Car pour Kojima, pacifiste convaincu, il est inenvisageable de représenter la guerre ou d'évoquer la puissance destructrice de l'arme nucléaire sans interpeller le joueur.

Si Psycho Mantis, personnage mentionné précisément, s'adresse directement à l'humain manette en mains, les autres membres de l'escouade Fox Hound et Solid Snake lui-même ne sont pas en reste. Lors de joutes verbales méticuleusement écrites et incarnées par de talentueux acteurs vocaux, les personnages font part de leurs états d'âme et plus particulièrement du sentiment irrépressible de révolte qui les

anime. Génétiquement modifiés pour être les soldats ultimes, ils n'ont pas d'autre choix que celui de combattre. Ils choisissent donc de répandre guerre, violence et mort pour façonner dans les cendres et le sang un monde auquel ils seront parfaitement adaptés.

Le jeu fait preuve d'un refus radical du manichéisme dans sa narration et surprend le joueur par ses rebondissements et révélations successifs. Kojima revendique avec ses jeux vidéo une posture d'auteur comme aucun autre développeur avant lui et participe à bouleverser ce loisir de plus en plus régulièrement considéré

Illustration 21: Solid Snake et Liquid Snake, leader de l'escouade Fox Hound et antagoniste principal

comme un art.

Metal Gear Solid fait date et devient une référence pour les développeurs suivants. L'influence cinématographique devient omniprésente. La narration devient plus que jamais un enjeu majeur. À partir de 1998, la voix en devient le vecteur

principal.

L'utilisation de vocalités artificielles devient précisément codifiée. La voix est plus que jamais envisagée comme un outil à la disposition des concepteurs de jeux et aux fonctions multiples. Quels sont ces codes et ces fonctions ? Ont-ils été érigés pour répondre à des besoins, contourner des contraintes ou au contraire brouiller les pistes et perdre le joueur ?

Dès lors, de nouveaux codes sont établis. Et de mutique, le jeu vidéo devient bavard.

CHAPITRE 2 : Proposition de formalisation de la grammaire vocale vidéo-ludique

Grammaire et écriture vocales vidéo-ludiques

Nous avons montré précédemment la trajectoire historique de la voix au sein des jeux vidéo. Au cours des années 90 s'opèrent différents tournants qui vont modeler les modes de production des vocalités artificielles vidéo-ludiques. De nouvelles contraintes vont apparaître avec la diversification des jeux et de leurs discours.

Pour répondre à ces différentes contraintes, que nous allons étudier, les développeurs vont établir des codes toujours plus nombreux, constituant avec le temps une **grammaire vocale**.

Dans le cadre de l'étude d'une langue, une grammaire est un ensemble de structures linguistiques propre à la langue étudiée. Maîtrisée par les concepteurs de jeu ainsi que par les joueurs, la grammaire vocale vidéo-ludique régit les méthodes de communication vocales entre le jeu et le joueur.

Cette grammaire adaptative évolue en fonction du genre de jeu de la même façon qu'un montage et des compositions de cadre spécifiques permettent de distinguer un film d'épouvante d'un western.

Nous appellerons par la suite **écriture vocale** l'adaptation de cette grammaire à un jeu par ses concepteurs. Il y a donc potentiellement autant d'écritures vocales qu'il y a de jeux !

Cependant, la **grammaire vocale** ainsi que les autres grammaires vidéo-ludiques ou même les emprunts aux grammaires d'autres arts et médias permettent, comme nous l'avons évoqué ci-dessus, une classification par genre. Nous en déduisons d'une part que les choix des développeurs de jeux différents peuvent répondre à des contraintes communes. D'autre part, il nous semble pertinent d'envisager, dans la mesure où la grammaire vocale vidéo-ludique a été apprise par les joueurs au gré des heures de jeu, que des développeurs en choisissent différents éléments pour permettre à leur jeu de louvoyer d'un genre à l'autre.

Pour le joueur ou le concepteur, cette grammaire vocale a été assimilée de façon empirique. L'association des différentes composantes d'une voix, c'est-à-dire de son contenu discursif mais aussi de sa manifestation acoustique, à une fonction, est une opération mentale de tous les instants pour l'individu en situation de jeu.

Nous proposons donc, dans une optique de formalisation théorique, de décrire les éléments principaux de la grammaire vocale vidéo-ludique à partir de l'analyse des écritures vocales de quatre jeux : ***Diablo III*** (Activision Blizzard, 2012), ***Halo IV*** (343 Industries, Microsoft Games, 2012), ***Dark Souls 2 : Scholar of the First Sin*** (From Software, Namco Bandai, 2015) et ***Bioshock*** (Irrational Games, 2K Games, 2007).

À partir de l'étude de ces jeux, nous mettrons en évidence trois concepts autour

desquels s'articule la grammaire vocale vidéo-ludique. Ces notions, **Immersion**, **Narration** et **Mécanique**, rassemblent différentes réponses aux contraintes que peuvent rencontrer les développeurs d'un jeu. En les analysant, nous montrerons qu'ils naissent d'impératifs différents, eux-mêmes apparus à des époques successives. Puis nous expliquerons quelles répercussions ces impératifs peuvent avoir sur la conception des vocalités artificielles d'un jeu vidéo.

UN NOUVEAU PARADIGME : L'IMMERSION

L'avènement du genre des FPS et plus largement des jeux en vue subjective témoigne de la volonté de nombreux développeurs de créer des univers dans lesquels les joueurs plongent entièrement. Le succès commercial de ces jeux prouve de plus la réceptivité des joueurs à ces propositions nouvelles.

À la perception du joueur doit se substituer un Nouveau Monde à pénétrer , à appréhender, celui du personnage incarné. Si cette substitution repose au départ sur la conception graphique des jeux à la première personne, elle est rendue plus prégnante par l'adaptation de techniques de spatialisation sonore aux jeux vidéo ainsi que la démocratisation des dispositifs de diffusion multicanale.

Très rapidement, cette nouvelle façon d'envisager et de concevoir l'univers sonore d'un jeu s'impose comme un nouveau standard, y compris pour les jeux à la troisième personne. Le maître mot devient l'**immersion,** défini stricto sensu comme

l'action d'immerger, de plonger dans un liquide. Ce terme décrit aussi, et surtout, d'un point de vue sociologique ou ethnologique, le fait de se retrouver dans un milieu tout à fait étranger sans garder de contact avec notre milieu d'origine.

Pour les joueurs, cela se manifeste par l'apparition de mondes toujours plus foisonnants de détails. Le pacte ludique, basé sur la suspension consentie par le joueur de son incrédulité, est de plus en plus solide. Les univers créés par les développeurs sont de plus en plus crédibles et cohérents, à l'épreuve d'un examen attentif et exigeant du joueur.

L'**ergonomie** des jeux vise par ailleurs une grande fluidité dans les enchaînements d'actions du joueur. L'interface doit se faire oublier et éviter à tout prix d'être un obstacle entre le joueur et l'univers dans lequel on souhaite l'immerger.

L'**axiomatique**, elle, se base de plus en plus sur l'adoption de modèles physiques pour décrire l'évolution de phénomènes particuliers au sein du jeu. Des interactions reposant sur l'utilisation d'une gravité reproduite avec une grande finesse font leur apparition. Des modèles de propagation du son sont adaptés aux différentes sources sonores virtuelles présentes dans le jeu si bien que l'éloignement du joueur à ces différentes sources a un impact puissant sur la perception sonore du joueur.

L'immersion devient avec ***Doom*** et ***Myst*** un idéal à atteindre.

A CHAQUE JEU SES CONTEURS

Mais l'immersion est essentiellement pensée pour être le support de la narration. Comme nous l'avons vu, la narration vocale devient dominante par rapport aux narrations textuelle ou visuelle à partir de la démocratisation des supports de stockage optique.

Néanmoins, la narration est en elle-même un des enjeux majeurs du jeu vidéo. Très tôt, il est proposé au joueur de devenir l'acteur d'une histoire, même s'il doit la reconstituer partiellement.

Au départ, les jeux d'aventure comme ***Zork*** (Infocom, 1979) ou ***Rogue*** (Epyx, 1980), très marqués par le jeu de rôle papier, alors très en vogue, proposent au joueur d'imaginer une histoire riche prenant place dans un univers complexe à partir d'une interface de jeu rudimentaire essentiellement textuelle.

A partir des années 1980, les jeux d'aventure et les jeux de rôle se diversifient respectivement avec l'apparition de jeux comme ***Legend of Zelda*** et ***Dragon Quest***. La part laissée à l'imagination du joueur devient mineure. La narration se complexifie. Les trames scénaristiques s'épaississent. Les histoires des jeux deviennent néanmoins plus linéaires en contrepartie[40]. Le joueur incarne un personnage identifié ayant un rôle très précis à jouer dans l'intrigue.

En général appelé à un grand avenir, comme sauver le monde d'une fin certaine ou délivrer une princesse des griffes d'un adversaire maléfique, le personnage principal est de plus en plus écrit pour que le joueur se sente l'acteur de la progression

40 Damien Mecheri, « **Qu'est-ce qui nous faisait vibrer dans un J-RPG ?, » Level Up,** *Niveau 1, Third Editions, 2015, p. 34-39.*

de l'intrigue.

De fait, les enjeux narratifs doivent lui être transmis d'une façon ou d'une autre. Le texte est un outil de choix dans un premier temps. Des paragraphes présentent souvent l'univers du jeu en introduction, puis l'essentiel de la narration est relaté par des dialogues écrits dont le joueur est lecteur passif. Ces différents dialogues réunissent aussi bien les personnages moteurs de l'histoire que les personnages plus mineurs et permettent au joueur de comprendre l'intrigue peu à peu et d'en éclaircir les tenants et les aboutissants.

Avec l'augmentation des possibilités offertes par les machines de jeu mais aussi l'apparition du CD-ROM, les développeurs de jeu vidéo vont considérablement faire évoluer les modes narratifs vidéo-ludiques.

Leur expérience leur permet tout d'abord de développer des intrigues bien plus complexes qu'auparavant, avec des trames scénaristiques secondaires au déploiement tentaculaire.

À partir du début des années 1990, le rapprochement assumé entre le cinéma et le jeu vidéo amène deux bouleversements principaux.

Tout d'abord la narration visuelle vidéo-ludique se développe en intégrant les nouvelles possibilités de la synthèse d'images en trois dimensions, forte également de divers emprunts à la grammaire cinématographique.

Ensuite, le jeu video devient, sous l'inspiration du cinéma, **verbo-centré**.

Le **verbocentrisme** est un concept créé par Michel Chion pour l'étude du cinéma. En particulier, ce terme décrit l'importance de la voix et de la parole dans le

cinéma classique à partir de la fin des années 1930. Pour Michel Chion, ce cinéma est « un cinéma où la parole est le centre de l'attention, mais sans en avoir l'air, parce qu'elle se crée des relais dans des actions visuelles parallèles[41]. »

Dans le cas du jeu vidéo, nous observons, à partir du début des années 1990, les mêmes phénomènes. Y compris au sein de jeux très immersifs dans lesquels le joueur ne perd jamais le contrôle des déplacements de son avatar, la narration vocale est dominante. L'intrigue ne progresse que par discussions et monologues successifs. Le joueur débloque, d'une certaine manière, ces différentes lignes de dialogues en résolvant les énigmes qui constituent le jeu, en traversant un niveau ou en terrassant un adversaire surpuissant. Faire progresser l'intrigue d'un jeu revient à accomplir des tâches particulières dont les conséquences seront étudiées par les personnages dans une sorte de débriefing. Dans cette perspective, les emplois de la voix et du langage oral relèvent d'un **verbo-centrisme** sensiblement similaire à celui qui est à l'oeuvre au cinéma.

Ce développement est entièrement cohérent avec le paradigme de l'immersion qui s'est imposé avec ***Doom*** et ***Myst***. En effet, une des conséquences de l'adoption de ce paradigme est la recherche de l'oubli total de l'interface. Comment faire oublier à un joueur qu'il joue ? De plus en plus, les *cutscenes* sont intégrées au jeu et ne se concrétisent plus autant qu'avant comme des séquences de non-jeu pendant lesquelles le joueur devient un simple spectateur. D'autre part, le texte est relégué aux

41 *Michel Chion,* **Un art sonore, le cinéma**, *op. cit., p. 70*

menus du jeu et majoritairement abandonné pour transmettre les informations principales.

Ces deux observations témoignent d'une volonté commune à de nombreux développeurs de créer des expériences instrumentées fluides et proposant une trame narrative ininterrompue. Rien ne doit venir s'opposer au joueur dans son exploration de l'univers façonné par les concepteurs du jeu, ni une cinématique provoquant une rupture de rythme indésirable, ni une succession de menus explicatifs ou de cartons descriptifs.

Il s'agit donc pour le développeur de structurer ses voix dans le temps et surtout par rapport aux actions du joueur. C'est ce dernier aspect, dynamique, qui va véritablement différencier les méthodes de production des vocalités vidéo-ludiques de celles qui ont cours au cinéma. Il est ici important de bien comprendre la logique d'un jeu vidéo et du déroulement d'une partie en fonction des actions du joueur.

Bien souvent, la narration d'un jeu vidéo progresse suivant un arbre de conditions logiques. Par exemple, SI le joueur atteint la caverne d'un vieil homme, ALORS ce vieil homme lui dira qu'il est dangereux de voyager seul et lui offrira une épée. Ou alors SI le joueur possède l'Anneau des Murmures, ALORS il pourra parler à l'Homme-Scorpion.

Illustration 22: **The Legend of Zelda** *sur NES, un vieil homme donne au héros sa première épée*

Illustration 23: **Dark Souls 2**, *Tark l'Homme-Scorpion.*

La spécificité de la narration vidéo-ludique apparaît ici. Les concepteurs d'un jeu créent une trame narrative dans laquelle le joueur évolue. Une arborescence de choix et de conditions à remplir limite ses excursions. Un jeu à la trame monolithique, sans digression ni aparté, sera considéré linéaire. À l'inverse, plus un jeu aura une arborescence narrative riche et fournie, plus il pourra être considéré ouvert.

La linéarité ou l'ouverture d'un jeu peuvent parfois se mesurer grâce à des indices vocaux comme le nombre de lignes de dialogues enregistrées.

En effet, le rapprochement entre jeu vidéo et cinéma a également bouleversé les méthodes d'enregistrement des voix d'un jeu. Si les développeurs étaient auparavant les acteurs vocaux de leur propre jeu, comme ce fut le cas pour ***Myst***, la production vocale s'est depuis considérablement professionnalisée et diversifiée.

Actuellement, il est courant que des acteurs vocaux, parfois célèbres[42], incarnent les personnages d'un jeu. Les dialogues sont segmentés en lignes que les acteurs vont enregistrer séparément et qui seront identifiées chacune comme un

42 *Se référer à l'introduction de ce mémoire.*

fichier distinct par le moteur son du jeu. Leur lecture sera alors soumise à des conditions fixées par les développeurs et que le joueur devra remplir.

Si la narration est un des trois pôles qui sous-tend la production des vocalités vocales vidéo-ludiques, il apparaît également que la narration vidéo-ludique a des principes qui lui sont propres, du fait de sa non-linéarité.

DEUX SOURCES DE CONTRAINTES MECANIQUES : L'**ERGONOMIE** ET L'**AXIOMATIQUE**

Qu'est-ce qui, au sein d'un jeu vidéo, définit la linéarité de sa narration ? Comment les différents choix de mise en scène des développeurs se concrétisent-ils ?

Nous avons défini le ***gameplay*** d'un jeu comme l'ensemble constitué par son **ergonomie** et son **axiomatique**. Pour concevoir le *gameplay* d'un jeu, il suffit au départ d'un crayon, de papier et d'imagination. Mais lorsqu'il s'agit de concrétiser ces esquisses et de réellement programmer le jeu, il existe trois options.

Historiquement, la première solution consiste en la programmation du jeu dans un langage de programmation informatique dont le choix est à l'appréciation du concepteur du jeu. Les premiers jeux vidéo furent par exemple codés en BASIC, un des premiers langages de programmation de haut niveau réputé pour être très accessible.

A partir des années 1990 apparaît un compromis. Avec les groupes de bibliothèques comme DirectX, les développeurs voient l'opportunité de tirer profit de

l'optimisation des liens entre *hardware* et *software* que permet DirectX. Ils vont donc programmer des interfaces de programmation et d'édition dédiées à leur jeu. Encore une fois, id Software est à l'avant-garde de cette évolution.

À ***Doom*** succède ***Quake*** (id Software, GT Interactive)***.***

Commercialisé dès juin 1996, ***Quake*** est un FPS extrêmement nerveux dans lequel le personnage principal est envoyé dans différentes dimensions pour détruire une entité cauchemardesque issue de l'oeuvre littéraire de Howard Philip Lovecraft.

Le caractère anecdotique du scénario laisse place à une expérience de jeu fluide et riche en sensations diverses. Les ennemis sont puissants et les différents niveaux ont une architecture pleine d'alcôves secrètes qui poussent le joueur à l'inventivité et à l'exigence. Les joueurs découvrent notamment qu'en contrepartie de dégâts subis, ils peuvent utiliser le souffle d'une explosion pour se projeter en l'air et atteindre des plateformes en hauteur jusqu'alors inaccessibles. Le *rocket jump* devient très rapidement un des emblèmes de ***Quake***, étant à la fois caractéristique de l'exigence requise pour découvrir tous les secrets du jeu, mais aussi de sa nervosité et de sa sophistication. Car la possibilité de réaliser des *rocket jumps* suppose l'existence d'une gravité et de différents modèles physiques décrivant les forces produites lors d'une explosion.

Illustration 24: Capture d'écran de **Quake***.*

D'autre part, ce sont les graphismes du jeu qui lui permettent de rencontrer un tel succès. En effet, à l'utilisation de fichiers image, appelés

sprites, pré-calculés et pré-rendus, les développeurs d'id Software préfèrent une technique novatrice : le rendu en temps réel de modèles en trois dimensions constitués de polygones.

Ces différents éléments, qui ont fait le renom de ***Quake***, sont rendus possibles par son moteur, le *Quake Engine*. Programmé en accord avec le cahier des charges fixé par les concepteurs du jeu, ce moteur est totalement dédié au jeu. Il permet aux développeurs de mettre en place les différents éléments du jeu, de les assembler et de visualiser le résultat. Qui plus est, le moteur gère différents modèles physiques pour régir les simulations d'interactions mécaniques entre les différents objets en jeu.

Quake rencontre un public constitué de joueurs de ***Doom*** déjà conquis par la patte d'id Software mais aussi de nombreux nouveaux joueurs. Car au-delà de la partie du jeu pouvant être arpentée par un joueur seul, ***Quake*** propose une expérience multi-joueur en réseau local (*Local Area Network* en anglais) ou sur Internet.

Si bien que très rapidement, de nombreux studios de développement veulent créer leur *Quake-like,* de la même façon que la sortie de ***Doom*** avait donné lieu à l'apparition d'un grand nombre de jeux plus ou moins ouvertement inspirés par le jeu d'id Software.

Le développement de tels jeux est permis par la publication du code source du Q*uake engine* sur Internet, faisant de ***Quake*** un logiciel libre pouvant être réutilisé ou modifié gratuitement. Il devient dès lors accessible à un très grand nombre de personnes de coder de nouveaux jeux d'une sophistication technique au moins égale à celle du jeu dont il utilise le code source.

D'autres moteurs connaîtront la même trajectoire comme les différentes versions de l'*Unreal Engine,* d'Epic Games*,* suite de moteurs dont est issue la série des ***Unreal Tournament***.

Illustration 25: Trois jeux réalisés avec l'Unreal Engine 3 : **Gears of War 3**

Illustration 26: **Batman Arkham Asylum**

Illustration 27: Et **Dishonored**.

Puis, des interfaces entièrement dédiées à la programmation de nouveaux jeux sont commercialisées ou publiées gratuitement, toujours dans une logique d'accessibilité au plus grand nombre.

Cet aperçu de l'histoire des moteurs de jeu est révélateur du fonctionnement d'un moteur de jeu. Un moteur de jeu permet de concevoir un jeu par la création puis la mise en interaction d'axiomes initiaux définis par les développeurs. Ces multiples axiomes vont orienter le jeu en posant les bases de son axiomatique puis en les développant. Par la suite, des fonctions et des outils internes seront développés pour paramétrer les interactions entre les différents axiomes initiaux. Élaborer l'**axiomatique** d'un jeu revient donc à en formuler l'ordre logique.

Mais dans la dynamique de recherche de crédibilité voire de réalisme qui est celle des développeurs actuellement, l'ordre logique des jeux est très souvent une imitation plus ou moins fidèle de modèles physiques réels.

De plus, au-delà de ces paramètres, un moteur de jeu permet de définir et contrôler les conditions d'existence de tous les éléments mis en jeu, l'affichage visuel

ainsi que la diffusion de sons.

Il en découle des séries de contraintes auxquelles il faut répondre pour éviter l'apparition d'interactions incompatibles avec un ou plusieurs principes déjà établis dont les conséquences pourraient être susceptibles de faire dysfonctionner le programme du jeu.

Mais le moteur d'un jeu en définit également l'**ergonomie**.

Les conditions de génération ou de lecture des sons du jeu, mais aussi et surtout leurs conditions de diffusion, sont paramétrées par les développeurs grâce au moteur du jeu. Notamment, des outils internes permettent un mixage dynamique des sons. Ce mixage est d'autant plus dynamique qu'il s'adapte et évolue en réaction aux faits et gestes du joueur. Pour répondre à quelles contraintes ces outils ont-ils été développés ? L'analyse de l'écriture vocale de ***Diablo III*** va nous permettre de proposer des éléments de réponse à cette question.

Diablo III *: de la nécessité de* feedbacks *vocaux*

La série de jeux ***Diablo*** débute en 1996 sur l'impulsion du studio californien Blizzard Entertainment. Dès son premier épisode, la série propose un jeu de rôle qui plonge le joueur dans un univers médiéval fantastique dont l'équilibre dépend de l'issue d'une lutte acharnée entre anges et démons.

Encore une fois inspiré par les jeux de rôle papier, ***Diablo*** fait cependant abstraction des aspects littéraires, contemplatifs ou centrés sur l'élaboration d'un personnage complexe à incarner avec implication. La part belle est donnée à des combats nerveux mais non moins tactiques. Le jeu se démarque de fait des autres jeux de rôle au rythme parfois plus pesant et aux systèmes de combats plus lourds.

Les affrontements ont lieu en temps réel dans des donjons générés aléatoirement par le programme du jeu une fois la porte franchie par le joueur. Les ennemis apparaissent également en nombre aléatoire, avec un placement et des capacités spéciales déterminées par le hasard également. Si bien que chaque partie est potentiellement différente à la fois de la précédente, mais aussi de la suivante,

renouvelant par là même l'intérêt du joueur. La narration est en retrait, avec des axes scénaristiques assez évidents et avares en réelles surprises. Les choix de mise en scène de Blizzard et la mise en avant du dynamisme des combats placent ***Diablo*** dans la catégorie des *action-RPG*, au rythme généralement plus enlevé que les RPG classiques. Un système de récompense, que nous décrirons par la suite par le terme *looting* est également mis en place par le jeu. Chaque monstre vaincu est susceptible de laisser tomber au sol un objet qui permettra au joueur d'en équiper son personnage pour le rendre plus puissant et apte à poursuivre son exploration. Le *loot* conditionne d'autant plus la progression que plus un adversaire sera dur à vaincre, plus la récompense qu'il pourra laisser tomber sera intéressante. Ce système, également une fois basé sur des lois de probabilités, encourage le joueur à explorer plusieurs fois une même zone, à affronter un même adversaire redoutable jusqu'à l'obtention d'objets particulièrement puissants. Cette progression, dépendante de l'équipement du personnage autant que de l'expérience et de la dextérité du joueur qui l'incarne, permet de classer ***Diablo*** dans une autre sous catégorie du jeu de rôle, le *hack & slash* (littéralement « entailler et trancher »).

Attendu comme le messie par une génération entière de joueurs après un deuxième épisode devenu culte, ***Diablo III*** sort le 15 mai 2012 après une très longue phase de développement et de promotion commerciale. Le scénario s'organise en quatre actes et revisite la mythologie des deux premiers jeux en mettant en scène la lutte perpétuelle entre les anges et les démons.

Comme ses deux prédécesseurs, ***Diablo III*** voit son contenu modifié et élargi

par la commercialisation d'une extension ***Diablo III : Reaper of Souls*** à partir de mars 2014. Le scénario du jeu est rallongé d'un cinquième acte dont l'ambiance morbide rappelle les deux premiers opus de la série.

Nous analyserons l'écriture vocale de ***Diablo III.*** Le jeu, accompagné de son extension, ***Reaper of Souls,*** présente en effet une diversité de personnages et de situations de jeu bien plus riche que le jeu seul. Qui plus est, le système de jeu est arrivé, suite à un certain nombre de mises à jour d'équilibrage, à un stade de perfectionnement propice à son analyse. Bien qu'il soit sorti sur consoles, nous parlerons plus particulièrement de la version sortie sur ordinateur personnel.

Pour débuter une nouvelle partie, le joueur doit créer un personnage en choisissant son nom, son genre et sa classe. Ce sont ces deux derniers éléments qui auront un véritable impact sur l'expérience du jeu par le joueur.

La classe d'un personnage détermine son style de combat et une partie de sa personnalité, indifféremment du genre choisi par le joueur. Ces classes sont au nombre de six : Barbare, Croisé, Moine, Féticheur, Chasseur de démons et Sorcier. A chaque classe est associé un ensemble de compétences, ou de sorts, pouvant être choisis et utilisés par le joueur. Leur utilisation est conditionnée par la gestion de ressources spéciales.

Illustration 28: Sur ce concept art de **Diablo III***, on peut observer l'évolution de l'armure de la chasseuse de démons. Plus le joueur progresse, plus son avatar pourra s'équiper d'armures solides et stylisées.*

Par exemple, un chasseur de démons génère, à chaque coup porté à un ennemi, de la Haine qu'il peut utiliser pour déclencher des mouvements dévastateurs. Il peut aussi utiliser de la Discipline, une autre ressource produite au fil du temps et permettant des actions encore plus élaborées.

Parmi toutes les compétences disponibles pour son personnage, le joueur doit en choisir six qu'il répartira sur différents boutons de son interface de jeu. Sur son ordinateur personnel, le joueur associera généralement une compétence à chacun des clics de sa souris (le gauche et le droit) et quatre compétences à des boutons proches les uns des autres sur son clavier. Cette association de compétences détermine le style de jeu du personnage, définissant de la sorte d'une part des affinités avec d'autres classes dans le cadre d'une partie coopérative et d'autre part des forces et des faiblesses à l'égard de certains types d'ennemis.

Traditionnellement, les jeux de rôle permettent deux styles généraux de combat, à savoir au corps à corps (en anglais *melee*) ou à distance (*ranged*). L'adoption d'une classe va souvent conditionner le choix de l'un ou l'autre de ces styles de combat puis

déterminer l'orientation du joueur vers l'un des trois rôles principaux de ce type de jeux : source de dégâts (en anglais *DPS* pour *Damage Per Second*), soigneur (en anglais *healer*) ou bouclier humain (en anglais *Tank*). Ces trois rôles sont complémentaires et les forces de l'un permettent de pallier les faiblesses des deux autres pour augmenter les chances de survie d'un groupe de guerriers face à des vagues successives d'ennemis.

Car la voie d'un joueur, qu'il joue seul ou en groupe, est parsemée d'embûches et d'adversaires de plus en plus puissants et retors, demandant une maîtrise toujours plus poussée du jeu et de ses mécaniques.

En dehors de ces situations de combat intenses, le joueur évolue dans une zone faisant office d'avant-poste ou de refuge dans laquelle il peut converser avec des personnages non-joueur afin de faire progresser l'intrigue du jeu, de développer des trames scénaristiques secondaires liées à d'autres personnages gravitant autour de celui du joueur ou de perfectionner son équipement auprès de marchands ou d'artisans.

Au cours de ces scènes va se développer la plus grande partie du scénario, portée par des dialogues entre les différents protagonistes ou par des *cutscenes* mettant en scène les évènements les plus marquants à grand renfort de spectaculaire. Mais c'est par l'aisance ergonomique qu'elle met en place que l'écriture vocale de ***Diablo III*** s'illustre.

Comme nous l'avons expliqué, ***Diablo III*** a un *gameplay* centré essentiellement

sur les combats, qu'ils opposent un joueur seul ou un groupe de joueurs à un groupe d'ennemis ou à un *boss.* Dans tous les cas, l'écran est surchargé d'informations. Le joueur doit tantôt se concentrer sur les ennemis présents à l'écran et leurs différentes attaques, mais aussi sur le positionnement de ses alliés et l'assistance qu'ils pourraient potentiellement requérir. Mais ces deux éléments visuels ne concernent que le combat en lui-même. Le joueur doit également contrôler en permanence une interface utilisateur très dense divisée en différentes parties réparties autour de l'écran. Chacune des sections de l'interface apporte des informations précises :

- En bas se trouvent toutes les informations sur le joueur. À gauche, une fiole se vidant et se remplissant d'un liquide rouge indique la quantité de points de vie restante. À droite, un récipient au fonctionnement similaire indique la disponibilité de la ressource propre à la classe du personnage du joueur. Entre ces deux fioles se trouvent une série de boutons indiquant d'une part les différentes compétences du personnage et sa capacité à les lancer à un instant donné mais aussi d'autre part la disponibilité d'une potion de soin et des boutons permettant de se téléporter rapidement jusqu'au refuge le plus proche ou d'ouvrir des menus de contrôle. Au-dessus de cette série de boutons se trouve une jauge orange se remplissant au fil des combats et indiquant l'expérience du personnage.
- En haut à gauche figurent les noms et les portraits des différents membres du groupe, qu'ils soient de réels joueurs ou des personnages non-joueur.
- En haut à droite est inscrit le nom de la zone en cours d'exploration.
- En haut au centre se trouve le nom de l'ennemi actuellement ciblé ainsi qu'une

barre rouge indiquant ses points de vie restants.

- Sur les bords droit et gauche, le joueur peut trouver différentes informations écrites indiquant les objectifs qu'il doit atteindre pour progresser mais aussi une fenêtre de discussion accueillant des échanges écrits entre les membres du

1. Portrait du joueur et de ses coéquipiers.
2. Santé du joueur.
3. Compétences du joueur.
4. Barre d'expériece du joueur.
5. Potion de santé.
6. Boutons d'accès aux menus.
7. Ressource du joueur.
8. Nom et carte de la zone.
9. Objectifs de la quête en cours.
10. Personnage du joueur.
11. Compagnon.
12. Personnage-non joueur avec lequel il est possible d'interagir.

Illustration 29: Interface visuelle de Diablo III.

groupe.

Même si ces informations sont très rapidement identifiables et compréhensibles, leur grand nombre rend l'interface visuelle complexe. C'est alors que de nombreux indices sonores vont compléter cette interface visuellement surchargée.

Plus particulièrement, les différents personnages vont prendre la parole pour décrire dynamiquement les enchaînements de situations au cours d'un combat.

Toutes les informations mentionnées précédemment sont susceptibles d'être données à voix haute par un des personnages à l'écran.

Plus précisément, le personnage incarné par le joueur indiquera qu'une compétence n'est pas disponible, que sa santé est faible, qu'il vient au contraire d'être soigné, qu'il vient de passer au niveau d'expérience supérieur, que ses ressources sont épuisées, qu'il ne dispose plus de potion de soin ou qu'il vient d'abattre très rapidement un très grand nombre d'ennemis.

De même, les personnages accompagnant le joueur dans sa quête pourront, qu'ils soient contrôlés par un autre joueur ou non, s'exprimer pour donner les mêmes informations les concernant.

Ces différentes vocalités pouvant être énoncées dynamiquement sont ce que nous appellerons par la suite des **voix système**. Ces vocalités sont des réponses à des contraintes ergonomiques. Elles ont pour fonction de former une interface sonore dynamique pour informer le joueur en cours de partie. Dans le cadre de certains jeux, le très grand nombre d'informations visuelles à analyser et à prioriser amènent les voix systèmes à guider le joueur dans le choix des informations à prendre en compte.

C'est précisément pour cette raison que les voix système de ***Diablo III*** complètent et transcendent l'interface utilisateur visuelle. En effet, ces interfaces, que l'on appelle généralement ATH, pour Affichage Tête Haute, sont avec le temps

devenues de plus en plus complexes. Particulièrement, les ATH de jeux de stratégie, de gestion ou de rôle, dans un souhait d'exhaustivité et de richesse de *gameplay*, témoignent d'une complexité croissante. De plus, elle sont souvent constituées d'éléments statiques en surimpression par rapport à l'action et s'imposent au joueur comme un filtre entre son jeu et lui. Qui plus est, même si la sophistication graphique de ces interfaces participe à l'établissement d'une atmosphère de jeu, elles restent désincarnées et distinctes des personnages et des mondes qu'ils habitent. Il est cependant à noter que différentes expériences d'intégration totale de l'interface visuelle dans l'environnement de jeu, et notamment la série des ***Dead Space*** (Visceral Games, EA Games, 2008), s'opposent à cette tendance dominante.

Illustration 30: Dans Dead Space, la barre de vie du joueur est intégrée à la combinaison du personnage et est située au niveau de sa colonne vertébrale

Dans le cas de ***Diablo III,*** les voix système, ainsi que différents sons non-vocaux viennent donner une incarnation à l'interface visuelle en donnant une voix aux personnages à l'écran.

La grande qualité de l'écriture vocale du jeu devient alors évidente. L'ensemble

des voix systèmes offrent un réel confort ergonomique au joueur, et sont de fait une réponse à une série de contraintes liées à des informations jugées nécessaires par les développeurs, et devant de fait parvenir au joueur.

Mais les voix système de ***Diablo III*** transcendent ces contraintes en utilisant ces différentes situations d'énonciation pour donner vie à l'histoire principale, mais aussi en faisant naître des trames scénaristiques secondaires ou en décrivant l'évolution des relations entre les différents protagonistes. La personnalité du personnage incarné par le joueur est de plus déterminée par sa classe et est révélée progressivement par les différents commentaires que ce dernier est susceptible de faire lors d'un combat. L'analyse comparative de différentes situations de jeu permet de comprendre la construction des *feedbacks* vocaux au sein de l'écriture vocale de ***Diablo III***.

Les chasseurs de démons forment une classe de guerriers agiles mais fragiles se battant à distance grâce à des arcs et des arbalètes pour infliger d'importants dégâts à leurs cibles. Assez proche d'un assassin dans son style de combat, le chasseur de démons que peut incarner le joueur est un combattant solitaire mû par une soif inassouvie de vengeance. Il est à la fois déterminé, impassible et vif. Si sa personnalité est développée au cours de dialogues narratifs, étudions comment les différents éléments de *feedback* vocal complètent le portrait du personnage établi jusqu'alors.

Une fois à cours de ressources, le chasseur de démons est susceptible, si le joueur tente de lancer une compétence dépendant de la ressource en question, de

prononcer six phrases différentes pour décrire sa situation.

De même, si le joueur tente d'utiliser une compétence avant qu'elle ne soit disponible, le personnage pourra encourager le joueur à patienter par trois phrases différentes.

Dans ces deux différents cas, le personnage reste calme malgré l'urgence de la situation. Le chasseur de démons ne fera part de son inquiétude qu'une fois sa fiole de vie quasiment vide. Dans les quatre phrases témoignant de sa vulnérabilité, le chasseur de démons conserve un niveau de langage soutenu et une élocution ferme, mais l'affolement et l'inquiétude transparaissent dans sa voix.

Enfin, lorsqu'un grand nombre de monstres sont vaincus en un seul assaut, le personnage va célébrer cette victoire en apostrophant les monstres restants.

En version originale anglaise, le chasseur de démons est susceptible dans une telle situation de prononcer différentes phrases en accord avec sa personnalité et son histoire personnelle, notamment :

- « Darkness awaits you ! »
- « Vengeance ! »
- « You won't survive that ! »
- « Your fear betrays you ! »
- « Feel my wrath ! »
- « Do you want more ? »[43]

Illustration 31: Un concept art du chasseur de démons de Diablo III.

L'impulsivité et le désir profond de vengeance qui animent le chasseur de

43 *[Traduction] : Les Ténèbres vous attendent ! Vengeance ! Vous ne survivrez pas à ça ! Votre peur vous trahit ! Craignez ma fureur ! Vous en voulez encore ?*

démons transpercent ces différentes phrases, donnant un aperçu de ses émotions pendant un combat.

A titre comparatif, le personnage jouable du croisé (ou de la croisée), disponible depuis l'ajout de l'extension ***Reaper of Souls***, fonctionne de la même manière. Néanmoins, le style de jeu du croisé est à l'opposé même de celui du chasseur de démons. Le croisé est un mastodonte en armure, mû par une foi indéfectible. Maints champs de bataille ont été arpentés par ce chevalier aux armes sacrées.

De même que le chasseur de démons, le croisé se dévoile lors de la célébration d'une victoire. Mais ce sont sa foi, ainsi que sa discipline et sa retenue, acquises au cours de longues d'entraînement et de combats, qui émanent de ses phrases.

Par exemple, toujours en version originale anglaise :

- « By the Light be damned ! »
- « Bad luck for you, friend. »
- « Make your peace quickly. »
- « The crusade marches on. »
- « Cower before Zakarum. »
- « Had enough ? »[44]

Illustration 32: Le croisé est l'antithèse du chasseur de démons, aussi bien morphologiquement que pour ce qui est

Tout en ponctuant le jeu et en continuant de donner des informations au joueur sur le déroulement de sa partie ou de le congratuler, ces voix système apportent de la profondeur à l'intrigue et aux personnages.

44 *[Traduction] : Soyez damnés par la Lumière ! Pas de chance pour vous, l'ami. Trouvez le repos, et vite. La Croisade continue. Prosternez vous devant Zakarum! Vous en avez eu assez ?*

En effet, au-delà de ces phrases aléatoires, des dialogues aléatoires peuvent également débuter après une phase de combat. L'exploration de la zone actuelle n'est pas interrompue et les personnages à l'écran discutent brièvement. Le templier, le brigand ou l'enchanteresse, trois personnages non-jouables pouvant accompagner le joueur dans sa quête et dont les compétences peuvent évoluer selon ses besoins, se lient d'amitié avec le personnage incarné par le joueur lors de ces conversations. Ces échanges verbaux ont trois buts : continuer de fournir des éléments de *feedback* vocal, apporter de la profondeur aux différents personnages et donner à entendre différentes réflexions sur l'intrigue principale.

Lorsque des ennemis puissants, repérables par leur aura lumineuse de couleur, apparaissent dans le champ de vision, chaque compagnon est susceptible d'annoncer ce groupe d'adversaires. Encore une fois, cet élément de *feedback* vocal dépend dans son contenu et dans sa forme de la classe du personnage principal et de l'identité du compagnon. Ainsi, Eirena, l'enchanteresse ingénue et peu habituée aux combats aura des lignes de dialogue différentes de celles de Kormac, un templier sans crainte.

Les voix système de ***Diablo III*** sont d'autant plus remarquables qu'elles ont été réalisées avec intelligence. Les acteurs vocaux ont été choisis avec soin et incarnent avec pertinence les différents personnages. Enfin, elles apportent un supplément d'incarnation à des personnages qui seraient bien plus ternes s'ils ne pouvaient s'exprimer en plein combat.

Néanmoins, il faut bien comprendre que ces **voix système** dépassent avec brio

leur fonction principale qui est, rappelons le, de donner une alternative à l'interface visuelle. En général, les voix système ne sont pas pensées pour apporter ce type d'informations supplémentaires. Bien au contraire, dans le cadre de nombreux jeux, elles constituent une sorte d'assistant personnel pour le joueur. Elles lui indiquent la direction à suivre ou l'informent sur les armes à sa disposition d'une voix généralement neutre voire désincarnée. Ces voix répondent essentiellement à des contraintes ergonomiques et ne représentent que très rarement un enjeu narratif ou immersif.

Halo 4 *: Le cas Cortana*

A l'inverse, nous définirons par la suite **voix narratives,** les voix ayant pour objectif principal de véhiculer la narration, devenue dominante dans l'écriture vidéo-ludique.

Pour décrire ces vocalités artificielles auxquelles nous sommes désormais habitués en tant que joueurs, nous étudierons l'écriture vocale de ***Halo 4*** (Bungie,

Microsoft Games, 2012) en focalisant plus particulièrement notre attention sur le personnage de Cortana.

La série de jeux ***Halo*** débute sur *Xbox* en 2001 lorsque Bungie collabore avec Microsoft pour leur fournir un jeu disponible au lancement de la console et dont le but affiché est de devenir un incontournable.

Le premier épisode de la série pose les bases de la franchise. ***Halo : Combat Evolved*** est un *First Person Shooter* futuriste dans lequel le joueur incarne John-117, un super-soldat né d'expériences scientifiques et confronté à une terrible menace extraterrestre. Depuis des décennies, les Covenants, une alliance interstellaire de plusieurs espèces intelligentes a pris pour cible l'humanité et entrepris de vitrifier l'ensemble des planètes sur lesquelles des humains se sont installés.

Tandis que l'étau se resserre autour des dernières planètes habitées par les humains, une flotte humaine, menée par le vaisseau *Pillar of Autumn* avec à son bord John-117, découvre un corps céleste artificiel, le Halo. Farouchement gardé par les covenants, le Halo est un anneau, capable de détruire toute forme de vie organique dans la galaxie, dont le diamètre excède celui d'une planète.

L'intrigue de ***Halo*** premier du nom est centrée sur la découverte de cet anneau et sa destruction par le joueur, après de multiples rebondissements et révélations.

Pour ce qui est de son *gameplay*, ***Halo*** est un FPS dans la continuité logique de ***Quake***. Il se distingue cependant de ses homologues par deux éléments de *gameplay*.

Tout d'abord le jeu surprend par la sophistication, par rapport à sa date de sortie, de l'intelligence artificielle des covenants, capables d'encercler spontanément

le joueur pour ensuite le prendre à revers ou au contraire pouvant céder à la panique la plus totale une fois leur chef d'escouade abattu.

Ensuite, le jeu marque les joueurs par l'immensité des zones qu'ils peuvent explorer à l'envi. Celles-ci sont si vastes que des véhicules de conception humaine ou covenant peuvent être pilotés par le joueur pour progresser plus rapidement en bénéficiant d'un armement lourd et d'une protection supérieure.

Pour l'aider à se repérer et lui indiquer ses différents objectifs, John-117 peut alors compter sur Cortana, une intelligence artificielle intégrée à son armure de combat. Mais très rapidement, Cortana fait preuve d'humour et de personnalité, devenant un personnage à part entière, agissant en symbiose totale avec John-117. D'un ensemble de **voix système**, Cortana a subtilement mué en un ensemble de **voix narratives**.

La série s'étoffera par la suite avec ***Halo 2*** en 2004, ***Halo 3*** en 2007, ***Halo 3 : ODST*** en 2009, ***Halo Reach*** en 2010 et ***Halo 4*** en 2012.

Si le *gameplay* évolue très peu d'un jeu à l'autre, l'intrigue, elle, s'étoffe d'épisode en épisode, en particulier grâce à ***ODST*** et ***Reach*** qui délaissent John-117 au profit d'autres personnages principaux et se déroulent tous deux plusieurs années avant les évènements des jeux principaux.

A la fin de ***Halo 3***, après un combat intense à la surface d'une installation stellaire supposée contrôler l'ensemble des Halo existants, John-117 se place en sommeil cryogénique dans un vaisseau spatial disloqué et à la dérive tandis qu'il est

célébré sur Terre comme un héros tombé au combat. Il est veillé par Cortana qui prend la place de l'intelligence artificielle défaillante du vaisseau en déroute.

Développé par 343 Industries suite à la fin de la collaboration entre Microsoft et Bungie, ***Halo 4*** donne au joueur la possibilité d'incarner une fois de plus John-117 et reprend donc l'histoire là où le troisième opus l'avait laissée : dans une moitié de vaisseau à la dérive.

Après une courte série de cinématiques puis une brève phase de découvertes des contrôles, le joueur fait sortir John-177 de sa capsule de survie et Cortana intègre de nouveau son armure.

Les premiers combats du jeu débutent peu après, les restes du vaisseau étant attaqués par une flotte de vaisseaux covenants.

Dès lors, le jeu va alterner entre séquences de jeu et *cutscenes*. Ces dernières segmentent les différents niveaux et mettent en relief les éléments cruciaux de l'intrigue en les mettant en scène.

Nous en déduisons qu'il y a deux types de voix narratives : les voix entendues par le joueur lorsqu'il contrôle son avatar et celles qu'il entend au cours d'une cinématique.

Dans le cas de ***Halo 4***, il faut distinguer, au cours des premières minutes de jeu, deux postures d'écoute liées aux deux types de voix mentionnés ci-dessus.

La cinématique d'introduction du jeu ne demande aucune intervention du joueur. De fait, la posture d'écoute du joueur est alors extrêmement similaire à celle

d'un spectateur de cinéma. De même, les voix de cette séquence d'introduction ont été enregistrées, et post-produites suivant des méthodes similaires à celles qui sont employées dans le cadre de l'élaboration d'un film. Elles répondent donc d'une grammaire cinématographique plutôt que vidéo-ludique.

Nous appellerons ce type de **voix narratives**, employées dans le cadre de cinématique, des **voix linéaires**, du fait de leur lecture totalement indépendante du joueur.

Les **voix narratives** que le joueur entend lorsqu'il joue activement relèvent, elle, d'une grammaire purement vidéo-ludique. En effet, si leur enregistrement et leurs traitements les rapprochent des **voix linéaires**, leur diffusion est soumise à des conditions extrêmement précises définies par l'axiomatique du jeu. Pour ce faire, les développeurs, intègrent ces voix au sein du moteur du jeu puis spécifient les évènements aboutissant à leur diffusion ainsi que les caractéristiques de cette diffusion.

Afin de comprendre comment ces décisions prennent forme dans le moteur d'un jeu, étudions les **voix non linéaires** de ***v0x***, le jeu réalisé dans le cadre de la partie pratique de ce mémoire grâce au moteur Unity 5.

Si l'objectif de ce jeu est avant tout, comme nous le préciserons plus tard, de mettre en avant les nouvelles interactions vocales rendues possibles par l'utilisation de technologies de détection vocale, une structure narrative simple a cependant été conçue pour rendre l'expérience de jeu moins austère.

Une fois l'écran de titre passé, une cinématique accompagnée d'une **voix linéaire** débute. Dans le moteur du jeu, cette voix est un fichier son dont la lecture débute en même temps que le fichier vidéo correspondant à la cinématique, indépendamment d'une action du joueur.

A l'inverse, au cours du cinquième tableau du jeu, le personnage principal, l'agent 02 découvre une tête immense. Tandis qu'il traverse le tableau de gauche à droite, la tête reste inanimée et silencieuse. Soudain, une fois arrivé à son niveau, le joueur perd le contrôle de son avatar et une voix nouvelle se fait entendre. Ce que l'on identifie alors comme le gardien des lieux s'est mis à parler au cours de ce que nous appellerons un **dialogue scripté**.

Dans le moteur, nous avons défini une zone en dessous du personnage du gardien. Cette zone fonctionne comme un interrupteur déclenché par les mouvements du joueur, pour lequel une zone similaire a été définie. Lorsque ces deux zones entrent en collision, le fichier audio est lu. A l'inverse, il ne pourra l'être si le joueur reste immobile ou se déplace entre le bord gauche de l'écran et celui de la zone.

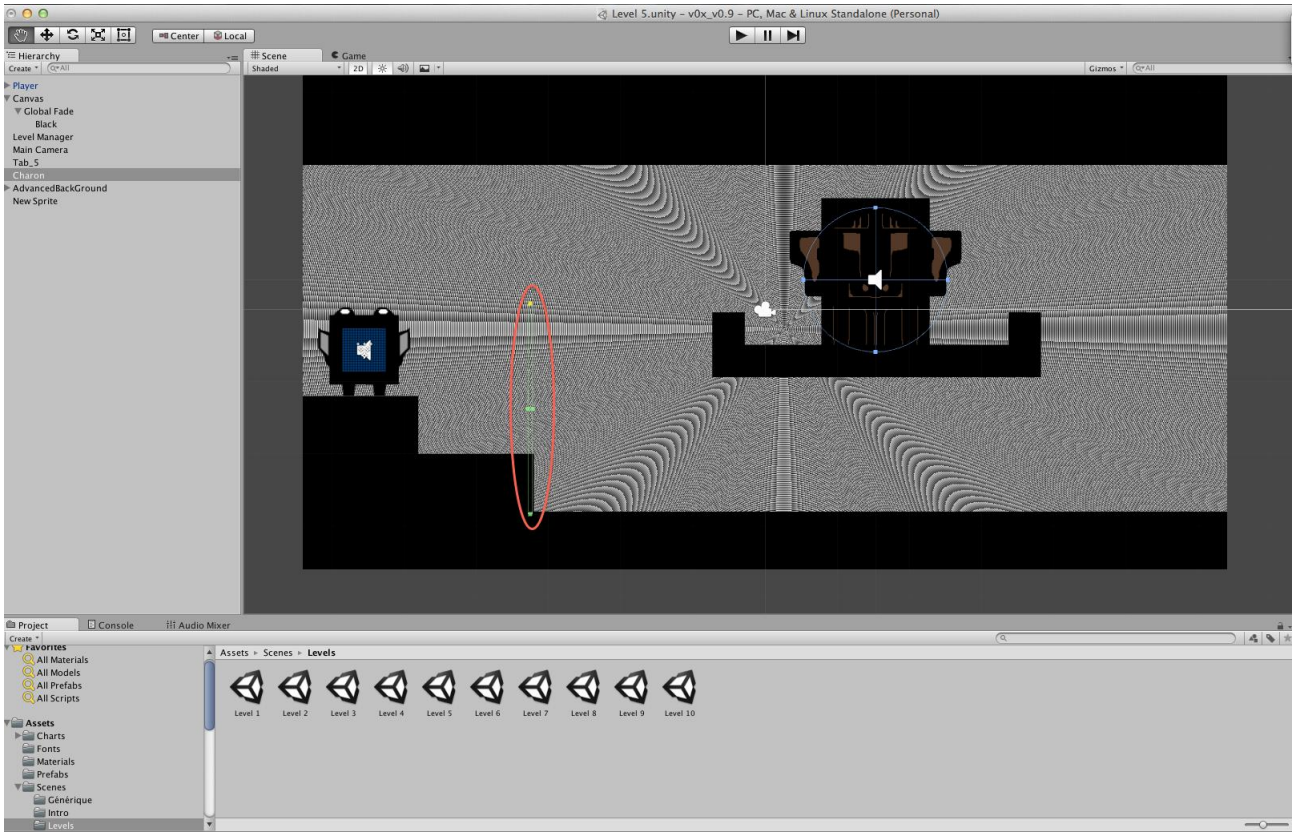


Illustration 33: Sur cette capture d'écran du logiciel Unity 5, nous avons entouré en rouge la zone que le joueur doit franchir pour déclencher le dialogue. Cette zone, matérialisée par un rectangle vert, est associée, ainsi que le fichier audio devant être lu, à l'objet nommé "Charon" dans la hiérarchie à gauche.

Cette explication apporte des notions supplémentaires indispensables à la compréhension de la distinction entre les **voix linéaires** et ce nouveau type de **voix narratives** dont l'évolution dépend intimement des actions du joueur.

Nous appellerons par opposition ce type de **voix narratives** des **voix non linéaires**.

Dans le cas de la série ***Halo***, de moins en moins de *cutscenes* viennent interrompre la partie du joueur, si bien que la grande majorité des informations scénaristiques sont relayées par des voix non-linéaires. Pour les développeurs, ces éléments vocaux doivent être entendus à tout prix par le joueur et donc se démarquer

nettement du reste de l'univers sonore du jeu.

Cependant, ***Halo 4*** est un jeu avec d'importantes ambitions immersives. Il serait donc impensable que l'intervention vocale d'un personnage dénote de celles de ses homologues ou qu'elle se noie dans d'autres éléments sonores.

Les développeurs de 343 Industries ont donc su exploiter une des possibilités offertes par les moteurs de jeu modernes : le **mixage dynamique**.

Dans la continuité logique de la distinction entre les voix narratives linéaires et non-linéaires, la nécessité d'établir une hiérarchie entre les différents sons que le joueur est amené à entendre est devenue peu à peu évidente.

Il serait fâcheux que la voix d'un personnage indiquant la destination suivante à atteindre ou une cible précise à abattre au cours d'un combat soit masquée par un son d'explosion ou par la voix d'un personnage mineur. Dès lors, pour parer à ces différentes éventualités, de nouvelles fonctions ont été ajoutées aux moteurs de jeu puis de nouveaux logiciels dédiés à la gestion dynamique des sons ont été développés.

Actuellement, les deux logiciels de ce type faisant office de référence sont ***FMOD Studio***, de la société FMOD et ***Wwise***, de la société Audiokinetic. C'est ce dernier logiciel qui a été utilisé pour la mise en oeuvre du moteur audio de **Halo 4**.

La logique de fonctionnement de ***Wwise*** est articulée autour du concept d'*Event* (en français évènement). Les *events* se divisent en deux types, les *action events* et les *dialogue events*, qui retiendront plus particulièrement notre attention. D'après la documentation de Wwise, les *dialogue events « utilisent un type d'arbre de décisions*

accompagné d'arguments permettant de déterminer dynamiquement quel objet sonore est joué[45]*. »* Nous retrouvons ici l'idée d'un raisonnement logique par arborescence qui nous indique que le mixage dynamique relève de l'*axiomatique* du jeu.

Par la suite, ces événements seront « intégrés au moteur du jeu afin qu'ils puissent être appelés en jeu en temps voulu.[46] » En effet, à ces *events* pourront être associés différents états correspondant à des ensembles de propriétés définissant des caractéristiques acoustiques de chaque fichier audio lu par le jeu. Par exemple, des changements d'espace de l'extérieur à l'intérieur d'une salle pourront être associés à une adaptation des effets de réverbération appliqués aux fichiers lus à cet instant. C'est ce qui est expliqué dans l'exemple suivant, issu de la documentation de Wwise.

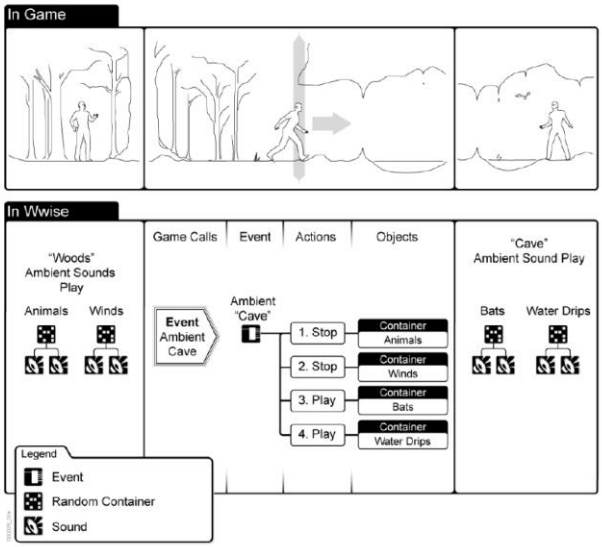
45 *Traduit de l'anglais : « These events use a type of decision tree with arguments to dynamically determine what object is played. »*

46 *Traduit de l'anglais : « [After Events are created in Wwise, they can be] integrated into the game engine so that they are called at the appropriate times in the game.*

Example 4.1. Using Action Events - Example

Let's say the character in your game must enter a cave to retrieve some hidden documents. When the character enters the cave from the woods, the ambient sounds in the game should change. To trigger this change, you must create an Event that will contain a series of actions that will stop the ambient "Woods" sounds and play the ambient "Cave" sounds. This Event will be integrated into the game engine and at the moment the character enters the cave, the game engine calls the specific Event that you created in Wwise.

The following illustration demonstrates how the game engine triggers an Event to change the ambient sounds playing in a game:



To deal with the transitions that occur between sound, music, or motion objects, each Event action also has a set of parameters that you can use to delay, or fade in and fade out incoming and outgoing objects.

Dans l'exemple ci-contre, les deux illustrations indiquent les différents éléments de Wwise mis en jeu lorsque les développeurs veulent effectuer une transition d'un premier

Enfin, des relations de priorité entre les différents évènements pourront être établies par les développeurs pour que la lecture d'un élément vocal narratif provoque une combinaison d'atténuations voire de modifications particulières d'autres éléments sonores.

Le confort ergonomique proposé par ce type de systèmes permet au développeur de s'assurer que le joueur pourra, en jeu, choisir plus facilement quels éléments vocaux requièrent son attention et donc en prioriser l'écoute.

La spécificité de la série de jeux ***Halo*** réside dans le fait que la quasi-totalité

des voix systèmes a été remplacée par des **voix narratives non-linéaires**, notamment grâce au personnage de Cortana, incarnée par l'actrice vocale Jen Taylor.

Au cours de la première mission de John-117, appelé Major par Cortana, l'intrication des deux fonctions de cette dernière est exposée. En sa qualité d'intelligence artificielle, elle est un guide pour le joueur, lui indiquant où diriger ses pas et quelles menaces sont à proximité. Cette partie, systémique, du personnage de Cortana, passe par une mise en relation des interfaces visuelle et sonore du jeu. Les ennemis apparaissent sous la forme de points rouges sur un radar en bas à gauche de l'écran. Au terme d'un combat avec un large groupe d'ennemis, Cortana indiquera au joueur que le covenant qu'il vient d'abattre était le dernier de la salle, l'autorisant par là-même à baisser sa garde de nouveau. Après un bref dialogue, elle indiquera le prochain objectif à atteindre et l'affichera sur l'ATH du joueur.

D'autre part, elle est, en raison de sa relation intime avec John-117, un des deux personnages principaux de cette histoire, ce qui explique la composante narrative dominante au sein de ses interventions vocales.

Au terme de la première mission, le vaisseau s'écrase sur une planète inconnue. John-117 reprend connaissance dans une vaste zone vallonnée défigurée par les débris. Une *cutscene* apporte alors une nouvelle information : Cortana est mourante. Elle révèle que les intelligences artificielles ont une espérance de vie fixe qu'elle a excédée d'au moins une année, ce qui risque de nuire à son intégrité et causer des dysfonctionnements.

L'assistance apportée par Cortana est alors mise en péril. Tandis qu'il navigue

d'un objectif à l'autre ou d'un combat à l'autre, le joueur franchit des seuils et déclenche par là-même différents dialogues scriptés. Au cours de certains d'entre eux, la déchéance de Cortana est mise peu à peu en évidence. Elle entre de plus en plus régulièrement dans un état qu'elle appelle la Frénésie (en anglais *Rampancy*) pendant laquelle sa voix devient distordue, métallique et agressive. Ses propos deviennent également moins compréhensibles et logiques. La diction de Cortana est souvent altérée voire erratique au cours de ces séquences, ce qui trouble généralement le Major.

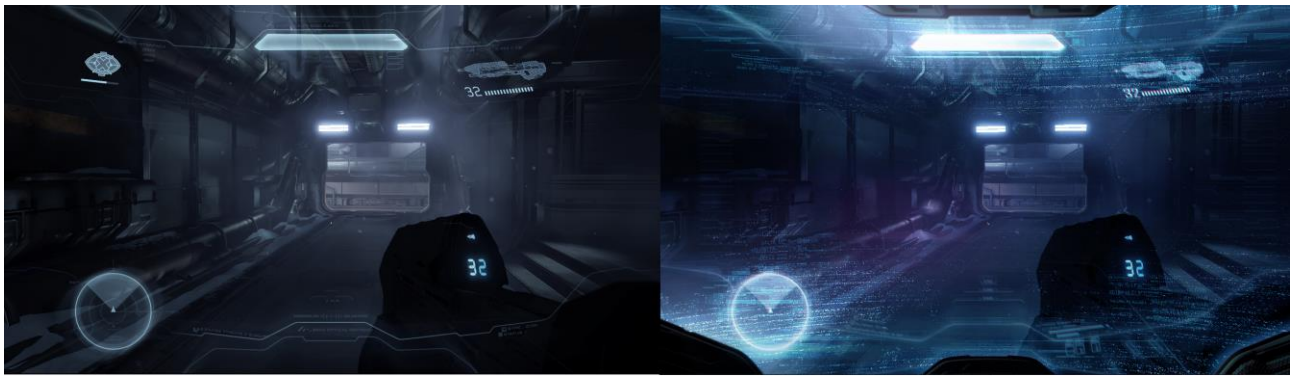


Illustration 34: La Frénésie de Cortana se manifeste également visuellement. À gauche, l'ATH normal de ***Halo 4****, à droite, l'ATH au cours d'une phase de Frénésie de Cortana.*

Mais qu'en est-il du joueur ? Comment le joueur perçoit-il cette perte de repère, a fortiori s'il a déjà joué aux précédents opus et qu'il a l'habitude de considérer Cortana comme un guide de confiance, inaltérable et inébranlable ?

C'est là toute la force de l'écriture vocale de ***Halo 4*** : faire de la métamorphose de Cortana, **voix système** par excellence (du fait de son statut d'IA) mais aussi **voix narrative** du fait de sa relation à John-117, un élément moteur du scénario.

Bien plus tard dans le jeu, le Major et Cortana retrouvent d'autres humains venus à bord d'un autre vaisseau, l'*Infinity*, pour les secourir. Mais une phase de

Frénésie particulièrement violente de Cortana les met en danger, ainsi que l'équipage de l'*Infinity*amenant l'officier en charge à ordonner son démantèlement. Le Major défie les ordres et décide de quitter les autres soldats humains pour protéger Cortana et poursuivre la mission à leur façon.

Le joueur incarne alors un Major déboussolé, accompagné par une intelligence artificielle de moins en moins stable et fiable. Sur un coup de colère, Cortana fait s'écraser le vaisseau covenant à bord duquel ils se sont lancés à la poursuite de leurs ennemis. La déroute est alors totale. Chaque séquence pendant laquelle Cortana perd le contrôle ou disparaît purement et simplement met le joueur dans une situation de grande faiblesse, révélant de fait sa dépendance vis-à-vis de Cortana.

Illustration 35: Cortana au cours d'une crise de Frénésie.

En offrant au joueur la présence d'un ensemble de voix système et de voix narratives sur lequel il se repose pour se déplacer et combattre d'une part, mais pour lequel il développe d'autre part une empathie croissante, les développeurs de ***Halo 4*** utilisent intelligemment le *feedback* et la narration vocaux. En le privant par la suite de

cet ensemble, matérialisé par le personnage de Cortana, les repères du joueur sont bouleversés, et l'urgence dramatique, voire le caractère désespéré de la situation, le frappe de plein fouet.

La désorientation du joueur est orchestrée par un enchevêtrement de **voix narratives linéaires** et **non-linéaires**. Si les premières définissent les enjeux qui sous-tendent le comportement de Cortana, l'exploitation du potentiel effet de surprise associé aux secondes installent définitivement un trouble en concrétisant la Frénésie par une perturbation du *gameplay.*

Toutefois, on peut remarquer que cette perturbation, ces effets de trouble que peut éprouver le joueur dans ***Halo 4*** tendent tous à faire progresser le joueur dans sa maîtrise du jeu et surtout sa compréhension de l'intrigue.

Il existe toutefois des jeux dans lesquels les personnages s'expriment, sans apport déterminant à la narration. Des jeux dans lesquels le joueur est condamné à errer dans un brouillard d'incertitude sans réel interlocuteur ni guide.

L'invitation à la perte : la série **Dark Souls**

Dans ces jeux, dont ***Dark Souls*** (From Software, Namco Bandai, 2011) et ***Dark Souls 2 : Scholar of the First Sin*** (Idem, 2015) font partie, l'absence totale de **voix système** et le nombre infime de **voix narratives linéaires** comme **non-linéaires** mettent à jour le troisième type de voix, les **voix d'ambiance**.

Ces éléments vocaux n'apportent aucune information supplémentaire à l'intrigue mais étoffent l'univers du jeu et ont de fait pour but de le rendre toujours plus profond pour le joueur, dans une logique d'immersion.

Ce type de voix est le plus fréquemment rencontré dans les jeux les plus immersifs notamment dans les jeux de rôle. En effet, si les autres jeux, comme par exemple ***Halo 4***, précédemment étudié, peuvent être immersifs, ils le sont généralement en proposant un jeu plutôt linéaire. L'incrédulité du joueur reste suspendue parce qu'on le place sur des rails dont le trajet garantit une expérience particulière et convaincante.

Les jeux de rôle, comme nous l'avons évoqué précédemment, laissent, eux, la part belle à l'exploration, et l'intègrent totalement à la narration pour la fragmenter et la disperser.

Cela peut impliquer de devoir traverser de vastes espaces pour aller d'un coin à l'autre de la carte du monde et pouvoir converser avec différents personnages non-joueur afin de faire progresser l'intrigue.

Cela peut impliquer qu'un garde d'une cité creusée dans les flancs d'une crevasse sans fond nous raconte qu'avant de prendre une flèche dans le genou, il était aventurier, comme nous.

L'exploration révèle l'existence de personnages mineurs, de lieux secrets ou d'intrigues secondaires apportant chacun un peu plus de cohérence et de subtilité à l'univers du jeu.

Les jeux de rôle à l'occidentale (par opposition aux jeux de rôle japonais), dont les séries ***Dragon Age*** (Bioware, EA Games, 2009), ***The Elder Scrolls*** (Bethesda, 1994) ou ***Divinity*** (Larian Studios, 2002) sont représentatifs, fourmillent de détails et de personnages ayant chacun une anecdote, une rumeur ou une légende à raconter.

Dans ces jeux, de nombreuses vocalités artificielles sont tout à fait différentes de **voix système** ou **narratives** car n'alimentant aucune intrigue. Ces voix, appelées **voix d'ambiance**, ont pour objectif de donner toujours plus d'éléments favorisant l'immersion.

En effet, le joueur découvre au cours de ses quêtes une multitude de lieux accueillant chacun de nouveaux interlocuteurs potentiels avec des informations très spécifiques sur la région dans laquelle ils habitent ou sur les différentes figures locales.

Dark Souls ainsi que ***Dark Souls 2 : Scholar of the First Sin*** se démarquent des jeux de rôle à l'occidentale habituels. Tout d'abord parce qu'ils ont été développés par les équipes du studio japonais From Software mais aussi pour leur *gameplay* et notamment leur écriture vocale.

Aucune voix système ne vient guider le joueur dans son périple et les voix narratives du jeu ne permettent pas à elle seule de reconstituer une trame narrative. L'univers de ***Dark Souls*** est austère et désespéré. L'environnement de jeu laisse une place minuscule au joueur voire lui est ouvertement hostile. Les pièges sont légion et

la plupart des ennemis, y compris les plus faibles, sont capables de le vaincre en un seul coup, en particulier en début de partie. Le joueur est donc très vite encouragé à progresser avec précaution en éliminant méthodiquement les ennemis d'une zone avant d'en poursuivre l'exploration. Car les deux ***Dark Souls*** sont des jeux qui ne se livrent pas aussi facilement. Le personnage principal reste mutique face aux différentes embûches qui entravent son chemin et l'omniprésence de la mort.

Dans ***Dark Souls***, la cinématique d'introduction installe une cosmogonie plus qu'un scénario. La genèse de quatre entités originelles, dont une qui engendrera l'Humanité, nous y est contée. Vient alors ce que le jeu nomme l'Âge du Feu, période faste de splendeur, bientôt ternie par l'apparition d'une malédiction transformant les humains en mort-vivants. Voyant leur règne mis en péril par le nombre toujours croissant de victimes de la malédiction, les entités originelles encore au pouvoir décident de déporter et d'enfermer les morts-vivants dans des asiles, loin de leur royaume de Lordran.

Au terme de la cinématique, le joueur prend le contrôle d'un mort-vivant, dont il a au préalable défini les caractéristiques initiales, dans les geôles d'un de ces asiles. Oscar, un chevalier en armure lui transmet, par un soupirail et sans un mot, les clés de sa cellule. Le joueur commence dès lors à chercher la sortie de l'asile. Il retrouve Oscar peu après, mourant. Ce dernier encouragera le joueur dans un râle d'agonie à partir pour la terre des Seigneurs Anciens pour y faire sonner la Cloche de l'Éveil. Il donnera enfin une potion de soin et une autre clé qui permettra au joueur de s'échapper enfin et de rejoindre Lordran.

Immédiatement après cette phase d'introduction, le joueur rallie le sanctuaire de Ligefeu, qui deviendra rapidement le seul véritable havre de paix du joueur.

Illustration 36: Le joueur se repose auprès du feu du sanctuaire de Ligefeu. Face à lui, le fantôme d'un autre joueur.

Au départ, le sanctuaire n'accueille aucun autre habitant qu'une femme muette emprisonnée derrière des barreaux de fer, un clerc qui refusera dans un premier temps d'adresser la parole au joueur ainsi qu'un homme assis près d'un feu, le guerrier déconfit.

Ce dernier ne donnera pas plus de clés de compréhension au joueur qui sera susceptible, en insistant et en essayant à maintes reprises de lui parler, d'obtenir de ce personnage des informations sur les mécaniques de *gameplay* du jeu ou d'indiquer vaguement la direction de la dite Cloche. Mais il le fera d'un ton extrêmement désobligeant, insistant sur l'impossibilité de la quête du joueur et raillant ce dernier en permanence.

Par la suite, le joueur pourra rencontrer ou même sauver différents personnages non-joueur puis les voir rejoindre le sanctuaire de Ligefeu. Une fois la cloche sonnée (les cloches en réalité), le personnage de Frampt fait son apparition au sanctuaire.

Frampt est un serpent géant millénaire dont la tête seule émerge d'un gouffre abyssal. Il apparaît pour donner enfin au joueur des éléments narratifs

supplémentaires. Il est attendu du joueur qu'il se rende à Anor Londo, la cité des Dieux, puis qu'il triomphe de quatre ennemis ayant hérité des âmes ou de portions d'âmes des quatre entités originelles. Il lui faudra alors détrôner le seigneur Gwyn, gardien de la Première Flamme et prendre sa place pour la raviver comme Gwyn le fit auparavant, au prix de son âme, pour instaurer l'Âge du Feu.

Néanmoins, Frampt s'exprime par énigmes et paraboles. Le joueur est pour lui un mort-vivant élu dont la destinée est d'hériter des pouvoirs de Gwyn et délivrer l'Humanité de la malédiction.

Illustration 37: Un concept art d'un des deux serpents primordiaux, Frampt et Kaathe

La difficulté à déterminer les informations fiables et essentielles dans les élucubrations de Frampt ne sera pas, malgré tout, un obstacle pour le joueur qui pourra, avec persévérance et après des heures d'exploration et de combats, accomplir la destinée que le serpent lui aura dictée. Le joueur pourra alors avoir le sentiment d'avoir terminé le jeu et d'avoir saisi les enjeux de la quête menée.

Cependant, ***Dark Souls*** est un jeu qui refuse le manichéisme et l'évidence. Le joueur explore Lordran tandis que la flamme a perdu de sa vigueur et que l'Âge du Feu, cet ancien âge d'or, approche de sa fin. Plus qu'un explorateur, le jeu de From Software propose de devenir l'archéologue de Lordran. En passant toutes les zones du jeu au peigne fin, le joueur trouvera des objets, voire des reliques ayant appartenu aux héros d'un autre temps et dont les descriptions, dans l'inventaire, révèlent leur histoire

et celle de Lordran à leur époque. Peu à peu, le joueur pourra, par curiosité, confronter les paroles cryptiques de Frampt à la chronologie ainsi reconstituée.

Sous certaines conditions, il pourra, en un lieu différent et bien moins accessible, rencontrer Kaathe, un autre serpent primordial. Ce dernier réside en effet dans une zone envahie par une obscurité tenace, requérant du joueur qui souhaite s'y rendre qu'il possède un anneau très particulier. Des deux serpents, il est celui qui parle le plus clairement. Sans détour, il raconte au joueur la décadence de Gwyn, liée à la perte d'intensité de la Première Flamme et sa crainte viscérale des humains frappés par la malédiction. Il révèle également au cours d'un dialogue que les morts-vivants ne sont apparus qu'après le sacrifice de Gwyn, ce qui laisserait penser qu'il est en réalité à l'origine du grand mal qu'il a depuis tenté de bannir de son royaume, en vain.

La rencontre avec l'un ou l'autre des deux serpents aboutira, une fois Gwyn défait, à un choix. À la fin de ***Dark Souls***, le joueur pourra, selon le désir de Frampt et d'autres personnages non-joueur, raviver la Première Flamme et relancer un nouvel Âge du Feu, ou quitter le Kiln de la Première Flamme pour devenir le Seigneur Sombre tant attendu par Kaathe et ainsi faire débuter l'Âge des Ténèbres.

Illustration 38: Le joueur prête serment devant Kaathe dans les Abysses.

La subtilité et la richesse de ***Dark Souls*** résident dans la posture proche de celle d'un archéologue nécessaire à la compréhension de l'univers du jeu et des lois qui le sous-tendent. Car si l'opposition entre le Feu, source de lumière et de chaleur, et les Ténèbres, semble au premier abord simpliste et manichéenne, l'exploration vigilante des différentes zones de jeu et la lecture attentive des descriptions d'objets apportent une multitude d'informations supplémentaires.

Dark Souls tisse l'intrigue qui relie les différents concepts de sa mythologie par l'association ou la comparaison d'éléments d'ambiance, qu'ils soient textuels dans le cas des descriptions d'objets, ou sonores, du fait du son écriture vocale. Sans aucune **voix système** et en utilisant un minimum de **voix narratives**, le jeu de From Software exploite les possibilités évocatoires des **voix d'ambiance** qui permettent de reconstituer l'histoire de Lordran en écoutant les contes, les légendes, les mythes ou les anecdotes personnelles transmis par les personnages non-joueur. Ce mode de

narration est un pied de nez à la narration classique des jeux linéaires que nous avons pu étudier. Il épouse le discours du jeu en préférant le faire découvrir au joueur par des méthodes plus proches de la tradition orale de transmission de savoirs anciens que de la narration cinématographique.

Dark Souls 2 : Scholar of the First Sin est la version définitive de ***Dark Souls 2***, constituée du jeu, sorti en 2014, et des différentes extensions prolongeant le jeu.

Comme son prédécesseur, ce jeu peut être fini par le joueur sans que celui ci n'en comprenne l'histoire pour autant. Pourtant, l'environnement de jeu de ***Dark Souls 2 : Scholar of the First Sin***, le royaume de Drangleic, est plus vaste et sa population plus nombreuse que celle de Lordran. Les personnages auxquels le joueur peut adresser la parole sont beaucoup plus nombreux que dans le premier opus. Comment l'intrigue du jeu peut-elle rester aussi opaque ? Que disent ces personnages non-joueur pour que le joueur soit dans une telle situation de déficit d'information ? Comment s'expriment-ils pour donner aussi peu d'informations au joueur ? Ont-ils seulement quelque chose à dire ?

L'écriture vocale de ***Dark Souls 2 : Scholar of the First Sin*** est un aboutissement de celle de ***Dark Souls*** premier du nom. Il en reprend ainsi de nombreux éléments, dont l'idée d'un havre de paix pour le joueur, Majula, dans lequel habitent ou viennent s'installer des personnages non-joueur amicaux proposant des services mais détenant également des informations sur la quête du joueur et sur le système de jeu. Le personnage de Saulden est notamment une réactualisation du

guerrier déconfit. Tout comme le joueur, il est un mort-vivant venu à Drangleic dans l'espoir de trouver une remède à la malédiction et retrouver son humanité et sa mémoire, mais il se sera résigné et aura abandonné cette quête dans un élan d'abattement total. Pourtant le joueur ne peut se résigner. S'il se résigne, le jeu prend

Illustration 39: Le Soleil couchant de Majula. Au premier plan, la tête entre les mains, Saulden.

fin.

Si ***Myst*** était une expérience immersive et contemplative dans laquelle la perte et la désorientation étaient sources de plaisir et de découvertes, ***Dark Souls 2 : Scholar of the First Sin*** invite le joueur à une toute autre errance. Plutôt qu'à l'égarement, le jeu exhorte le joueur à sa propre perte, soit la mort de son avatar.

Dans le film ***L'Intendant Sansho*** (Kenji Mizoguchi, 1954), la voix fantasmée d'une mère attire sa fille vers un étang dans lequel elle pénétrera pas à pas, jusqu'à sa disparition sous la surface ondoyante de l'eau. Pour Michel Chion, cette voix, incantatoire, quasiment magique, qui attire la jeune femme vers sa mort est au cinéma ce que le chant des sirènes est aux mythes grecs. Elle est, par ailleurs, d'autant plus remarquable que le film tisse autour d'elle un univers symbolique riche et signifiant, en les associant à des figures de la spiritualité japonaise. Parmi elles, l'eau, dans laquelle

viendra s'immerger la fille, Anju, et l'arbre dont la branche rompue augurera blessures et séparations :

> « L'eau est féminine, l'arbre masculin. L'eau et la voix sont deux images de ce qui n'a ni lieu ni limites si on ne lui en signifie une. L'arbre en revanche est un lieu, et marque une limite.
>
> Le motif des limites est important dans ***L'intendant Sansho***. [⋯] L'île de Sado est pour la Mère une prison qui la coupe de ses enfants. Le tendon qu'on lui coupe (l'arbre cassé) l'enferme encore plus dans les limites de son corps, et la plainte qui sort d'elle, et qui s'envole au loin, est d'autant plus déchirante.
>
> L'eau, dans ce film, est séparation, danger, mort. La voix de la Mère est ce qui subvertit les limites, ce qui traverse le temps et l'espace, mais pour la jeune fille, qui ne peut rejoindre sa Mère que dans la mort, la fusion avec l'eau, elle est *invitation à la perte*[47]. »

Nous remarquerons que la mère gagne sous la plume de Michel Chion une majuscule à son initiale. Par les pouvoirs de sa voix sur l'espace et le temps, qui lui sont conférés par son écriture vocale, ce personnage cinématographique semble transcender son rôle et devenir une Mère allégorique.

Si cette transcendance est provoquée par l'association de cette voix à des symboles spirituels, la voix en elle-même, dans sa forme tantôt incantatoire, tantôt plaintive, participe à l'établissement de ce personnage en tant que figure fantasmagorique.

Autrement dit, les caractéristiques acoustiques de cette vocalité véhiculent

---

47 Michel Chion, **La voix au cinéma,** *op. cit., p. 106.*

également des affects, des symboles et du sens.

Cette capacité que peut avoir une voix de se transcender elle-même et devenir un symbole plutôt que la simple incarnation vocale d'un personnage, Roland Barthes la décrit et l'explique grâce au concept de *grain* qu'il développe dans son texte ***Le Grain de la Voix***.

Il y formule ce concept à partir de réflexions sur la musique lyrique, en observant le rapport de certains chanteurs à la langue de l'oeuvre mise en voix. Roland Barthes s'appuie sur un exemple fictif pour éclairer son propos :

> « Listen to a Russian bass [...] : something is there, manifest and stubborn (one hears only *that*), beyond (or before) the meaning of the words, their form (the litany), the melisma, and even the style of execution : something which is directly the cantor's body, brought to your ears in one and the same movement from deep down in the cavities, the muscles, the membranes, the cartilages, and from deep down in the Slavonic language, as though a single skin lined the inner flesh of the performer and the music he sings. The voice is not personal, it expresses nothing of the cantor, of his soul ; it is not original (all Russian cantors have roughly the same voice), and at the same time it is individual : it has us hear a body which has no civil identity, no 'personality', but which is nevertheless a separate body. Above all, this voice bears along *directly* the symbolic, over the intelligible, the expressive : here, thrown in front of us like a packet, is the Father, his phallic stature. The 'grain' is that : the materiality of a body speaking its mother tongue ; perhaps the letter, almost certainly *signifiance*[48]. »

48 *Roland Barthes, «* **Le grain de la Voix***, »* **Image, Music, Text,** *Fontana Press, traduction de Stephen Heath, 1977, p. 181-182. [Traduction] : Écoutez une basse russe [···] : quelque chose est là, manifeste et borné (d'aucuns n'entendent que cela), au-*

Ce *grain*, définit par Barthes comme « la rencontre d'une voix et d'un langage[49], » est à l'origine de la puissance évocatoire d'une voix. D'un volume d'air inspiré puis expiré sous l'action des organes phonatoires et des poumons, elle devient, en rencontrant le langage du locuteur, une manifestation physique et signifiante de celui-ci.

Dans ***Dark Souls 2 : Scholar of the First Sin***, les différents acteurs vocaux qui offrent leurs voix aux différents personnages du jeu ont été choisis avec la plus grande attention, certes pour leurs qualités d'interprétation, mais aussi pour leurs *grains* de voix.

En effet, de la même manière que la voix de la mère de ***L'Intendant Sansho*** évoque la Mère ou que le chanteur russe imaginé par Roland Barthes donne à entendre le Père, l'acteur anglais William Houston, en doublant Vendrick, roi de Drangleic, incarne un monarque allégorique. En parlant à différents personnages et en croisant leurs dires, le joueur apprend que Vendrick a connu, comme de nombreux rois au cours d'autant de cycles, le même parcours que le joueur dans ***Dark Souls***, premier du nom. Mais, au moment de monter sur le trône et de raviver la Première Flamme,

---

delà (ou bien avant) le sens des mots, leur forme (la litanie), le mélisme et même le sytle d'exécution : quelque chose qui renvoie directement au corps du chanteur, parvenant à vos oreilles dans un unique mouvement issu des cavités, des muscles des membranes, des cartilages, et des profondeurs de la langue slavonique, comme si une simple peau distinguait la chair de l'artiste de la musique qu'il chante. La voix n'est pas personnelle, elle n'exprime rien du chanteur, de son âme ; elle n'est pas originale (tous les chanteurs russe ont peu ou prou la même voix), mais est malgré tout individuelle : elle nous donne à entendre un corps sans identité civile ni personnalité mais qui est un corps distinct. Par dessus tout, cette voix véhiculent avec elle le symbolique, plus que l'intelligible ou l'expressif : là, jeté à notre visage comme un paquet, se trouvent le Père, sa stature phallique. Tel est le grain *: la matérialité d'un corps parlant sa langue natale ; peut-être la lettre, plus certainement la* signifiance*.*

49 *Ibidem, p. 181.*

comme son devoir de roi le préconisait, il fut frappé de remords et n'eut pas la force d'entrer dans la salle du trône, laissant la Flamme faiblir.

En conversant avec Vendrick, le joueur apprend la source de son doute. Le roi a hésité à raviver la flamme car il avait été tenté de la laisser s'éteindre et de laisser débuter un nouvel âge de Ténèbres. Mais il ne parvint jamais à trouver la force de faire ce choix pour lui-même et tout son royaume.

Illustration 40: Le joueur aux côtés du Roi Vendrick, assis au premier plan.

« One day, fire will fade, and Dark will become a curse. Men will be free from death, left to wander eternally. Dark will again be ours, and in our true shape… We can bury the false legends of yore…

Only…

Is this our only choice[50] *? »*

La voix de Vendrick se brise sur cette dernière phrase.

Dans ce dialogue, le *grain* de la voix du monarque renseigne autant sur le doute qui s'est emparé de lui que le texte du dialogue. Par ailleurs, l'écriture de l'ensemble des dialogues de Vendrick dans un anglais ancien, rendue sensible par exemple par l'emploi du mot « *yore, »* issu du vieil anglais, permet la rencontre de cette voix et de ce langage.

La dernière spécificité de l'écriture vocale de ***Dark Souls 2 : Scholar of the First Sin*** réside dans l'absence de hiérarchie entre les différents éléments qui constituent son univers sonore. Les voix des personnages sont très souvent faibles, usées ou difficilement intelligibles car le locuteur est, par exemple, engoncé dans une lourde armure. De plus, les voix ne jaillissent que très rarement des autres éléments sonores, à savoir la musique, les ambiances et les bruitages.

De cette absence de hiérarchie préconçue, il résulte que toutes les voix de ***Dark Souls 2 : Scholar of the First Sin*** sont des voix d'ambiance que le joueur pourra éprouver pour déterminer la fiabilité des informations qu'elles contiennent et les hiérarchiser à sa manière.

---

50 [Traduction] : Un jour, la flamme s'affaiblira plus encore, et les Ténèbres deviendront une malédiction. Les hommes seront libérés de la mort, condamnés à errer pour l'éternité. Les Ténèbres nous appartiendront de nouveau, et sous notre véritable forme… Nous pourrons enterrer les fausses légendes des temps lointains. Seulement… Ce choix est-il le seul choix qui s'offre à nous ?

Malgré son hostilité apparente vis-à-vis du joueur, l'univers de ***Dark Souls 2 : Scholar of the First Sin*** est d'autant plus beau qu'il se livre à chaque joueur d'une façon différente. Chaque joueur qui arpente Drangleic en est tantôt l'archéologue, tantôt l'explorateur, tantôt le dépositaire de tous ses savoirs. Ce faisant, il acquiert une connaissance de l'univers du jeu synonyme dans ce cas là d'une maîtrise croissante du *gameplay*.

Les **voix d'ambiance**, habituellement reléguées à un rôle d'immersion, d'enluminure voire de remplissage, sont dans ce jeu mises à l'honneur et composent à elles seules son écriture vocale. Cela participe à faire de la série des ***Dark Souls*** une série de jeux uniques par leur univers ainsi que par le rapport au *gameplay* imposé au joueur, invité à se perdre toujours plus profondément, à ses risques et périls, corps et âme.

«Would you kindly ?» *: **Bioshock** ou l'aboutissement d'une **écriture vocale unidirectionnelle***

Illustration 41: C'est dans cette salle que s'amorcent les rebondissements principaux du scénario de **Bioshock**.

Au cours de ce chapitre, nous avons décliné les types de voix qui composent la grammaire vocale vidéo-ludique. Ils sont au nombre de trois : **voix système**, **voix narratives** et **voix d'ambiance**.

Chacun de ces types de voix correspond à un des concepts autour desquels des développeurs conçoivent l'écriture vocale d'un jeu vidéo : **Mécanique**, **Narration** et **Immersion**.

Bien que codifiée très tardivement, l'utilisation des voix dans le jeu vidéo est aujourd'hui à son apogée, à l'heure où la recherche de photo-réalisme et le rapprochement avec le cinéma sont affirmés et assumés (quoique controversé). Depuis les années 1990, la **grammaire vocale vidéo-ludique** s'est considérablement étoffée.

Il en résulte l'apparition d'une forme de classicisme dans l'élaboration des écritures vocales de nombreux jeux vidéo. Cela se manifeste à la fois par une soumission aux contraintes imposées par le genre de jeu choisi mais aussi par une volonté de développer une écriture vocale conforme aux écritures précédentes ayant pu devenir canoniques.

Un jeu s'est justement imposé en 2007 comme un nouvel incontournable, établissant de nouveaux codes, réactualisant l'utilisation d'outils jusqu'alors délaissés et en créant de nouveaux.

Il s'agit de ***Bioshock,*** développé par Irrational Games.

Aux commandes d'un personnage sobrement nommé Jack, nous y découvrons Rapture, une cité utopique bâtie dans les fonds marins du nord de l'Océan Atlantique sur l'impulsion d'Andrew Ryan. Ce dernier, un entrepreneur dont l'ambition n'a d'égale que la puissance de l'empire financier qu'il dirige d'une main de fer, a décidé au milieu des années 1940 de fuir les querelles politiques et religieuses incessantes pour fonder Rapture et en faire un refuge pour les cerveaux les plus brillants et les hommes les plus idéalistes.

Néanmoins, lorsque Jack arrive à Rapture suite au crash en pleine mer de l'avion dans lequel il voyageait, l'utopie s'est effondrée. La cité est partiellement inondée et les rares personnes qui y vivent encore ont sombré dans une folie sans limite.

Au fond de l'océan, les équipes scientifiques d'Andrew Ryan ont auparavant découvert une nouvelle ressource, l'Adam, rendant possibles les modifications

génétiques de l'humain. La télékinésie, l'invisibilité ou la capacité de faire sortir des torrents de flammes de la pointe de ses doigts sont désormais à la portée de tous. Malheureusement, l'Adam est une substance extrêmement addictive, amenant à des injections régulières de doses toujours plus importantes, causant des dommages irréversibles non seulement au corps de l'humain ainsi modifié, mais aussi à sa psyché, métamorphosée en un tourbillon de peur, de haine et de colère.

Si le joueur arrive bien après la chute de Rapture, celle-ci lui est donnée à entendre par le biais de nombreux journaux audio qu'il pourra récupérer et écouter. En documentant les derniers instants paisibles de la cité, sa déchéance brutale puis sa déliquescence actuelle, le joueur découvre autant de personnages qui resteront très souvent pour lui des voix sans corps, mais dont les histoires personnelles apportent des nuances au scénario et l'éclairent sous un nouveau jour.

Ces journaux audio rassemblent des **voix d'ambiance** et des **voix narratives** ni **linéaires**, ni **non linéaires** car pouvant être écoutées autant de fois que le joueur le désire et à n'importe quel instant, ce qui fait d'eux des éléments vocaux remarquables. ***Bioshock*** n'est cependant pas le premier jeu à présenter ce type de contenu, apparu pour la première fois avec ***System Shock***, précurseur de ***Bioshock*** aussi bien pour son gameplay que pour son ambiance claustrophobique et une partie de son intrigue.

Illustration 42: Un journal audio tels qu'ils apparaissent en jeu.

Ainsi, bien que le jeu de Looking Glass Studios soit à l'origine de l'utilisation de journaux audio au sein d'un jeu, c'est avec ***Bioshock*** que le recours à ces éléments vocaux si particuliers va se démocratiser jusqu'à apparaître dans des jeux comme ***Call of Duty***, ***Gears of War***, ***Dead Space*** ou ***Batman Arkham Asylum***, éloignés de l'esprit de ***Bioshock*** mais partageant avec lui des éléments de *gameplay*.

Plus qu'un nouveau code, c'est une véritable mode initiée par le jeu d'Irrational Games. Car cette proposition rencontre un important succès critique auprès des joueurs. Dans cette cité sous-marine dépourvue d'interlocuteur n'ayant pas sombré dans la folie, le joueur s'entoure des voix des témoins des derniers instants de la splendeur de Rapture et imagine, voire rêve de cette utopie qu'il ne verra jamais.

Si, selon Lev Manovich, ***Myst*** et ***Doom*** rattachaient définitivement le jeu vidéo aux nouveaux médias numériques du fait de leur inscription dans la tendance de ceux-ci à se manifester par la représentation d'un espace navigable, ***Bioshock,*** ainsi que les jeux qui prennent sa suite transforment le jeu vidéo, en mettant à jour une autre forme consubstantielle aux nouveaux médias aux yeux de Lev Manovich : la *base de données*.

Définie au départ comme « un ensemble structuré de données qui sont organisées de manière à permettre une recherche et une récupération rapides au moyen d'un ordinateur[51], » la *base de données* est une forme intimement liée aux supports numériques. Dès lors, toutes les données d'un jeu vidéo, en tant que programme informatique, sont méticuleusement organisées et forment une base de données. Un jeu vidéo correspond donc aux méthodes de recherche et de récupération, évoquées par la définition et appropriées pour interagir avec sa base de données. Pour Lev Manovich, le joueur qui souhaite terminer un jeu recherche en réalité l'algorithme, c'est à dire la suite d'opérations à réaliser pour atteindre les objectifs imposés par le jeu[52].

Un jeu vidéo est donc en réalité une base de données avec laquelle un utilisateur, appelé dans ce cas joueur, va interagir en suivant un algorithme pour atteindre de façon ludique des objectifs fixés par un programme.

Nous venons non seulement de donner une définition alternative de ce qu'est un jeu vidéo mais aussi de montrer que la forme de la *base de données* est absolument intrinsèque au jeu vidéo.

En offrant au joueur la possibilité de rassembler des journaux audio, c'est à dire des **voix d'ambiance** ou des **voix narratives**, correspondant à autant de fichiers audio dans les données du jeu, ***Bioshock*** révèle la forme de la base de données qui se trouve à son fondement, ainsi qu'à celui de tout jeu vidéo.

---

51 Lev Manovich, **Le Langage des Nouveaux Médias***, op. cit., p.394.*

52 Ibidem, p. 399 : « L'algorithme est également une composante essentielle de l'expérience des jeux vidéo, mais à un autre titre cette fois. En avançant dans la partie, le joueur découvre peu à peu les règles en vigueur dans l'univers construit par tel ou tel jeu. Il s'initie à la logique cachée de celui-ci ; bref, à son algorithme. »

Accumuler ces journaux audio permet en général au joueur de reconsidérer sa façon d'envisager l'univers de ***Bioshock*** et d'interagir avec lui. Dans une interview donnée à Shannon Drake pour le site Internet The Escapists, Ken Levine, co-fondateur d'Irrational Games, parle de *player-powered gameplay*[53], défini comme une situation de jeu dans laquelle c'est le joueur, et non le développeur qui conçoit le *gameplay* en jouant. La révélation que la forme de la base de données est indissociable du concept de jeu vidéo par le biais de la collecte de journaux audio constitue un sous-texte auto-réflexif qui place ***Bioshock*** dans la catégorie très restreinte des jeux vidéo qui parlent de jeux vidéo.

Cette volonté de mise en abyme atteint son paroxysme dans l'intrigue même de ***Bioshock***.

Dans la dernière partie du jeu, le chemin de Jack l'amène inéluctablement face à Andrew Ryan. Celui-ci lui révèle tout d'abord qu'il est son père puis que Jack a, dès son plus jeune âge, été modifié grâce à de l'Adam pour systématiquement obéir à tout ordre qui serait suivi de la formule de politesse « Je vous prie » (en anglais *« Would You Kindly »*). Il se servira une dernière fois de cette commande vocale pour ordonner à Jack de le tuer en lui fracassant le crâne à coups de club de golf, au cours d'une démonstration définitive de son pouvoir et de son libre-arbitre.

Or, n'est-ce pas là ce que fait tout joueur de jeu vidéo ? N'obéit-il pas, souvent aveuglément, à des séries d'ordres transmises par le jeu ?

Bioshock révèle dans ce second élan auto-réflexif que le joueur de jeu vidéo n'a

53 *http://www.escapistmagazine.com/articles/view/video-games/editorials/interviews/1227-Inside-The-Looking-Glass-The-Escapist-Talks-With-Ken-Levine*

ni son mot à dire ni d'autres choix que d'écouter des consignes auxquelles il devra obéir pour finir le jeu. La grammaire vocale, telle qu'elle est actuellement codifiée, est à sens unique, donnant lieu à ce que nous appelons des **écriture vocales unidirectionnelles**.

Des jeux ont cependant tenté de donner voix au chapitre au joueur et proposent d'inclure la voix du joueur et tout autre son qu'il pourrait émettre au sein du *gameplay*. Ces jeux proposent des **écritures vocales bidirectionnelles**. Nous allons maintenant étudier ce que cette **bidirectionnalité** signifie et ce qu'elle implique.

CHAPITRE 3 : Pour des écritures vocales bidirectionnelles

Dans ce chapitre, nous allons étudier la possibilité d'inclure le joueur dans le *gameplay* d'un jeu grâce à sa voix.

Au terme d'un aperçu des tentatives, fructueuses ou non, qui ont déjà été menées en la matière, nous exposerons les contraintes qu'imposent ce parti-pris de réalisation ainsi qu'une partie des technologies utilisées pour sa mise en oeuvre.

Enfin, nous rendrons compte des travaux menés dans le cadre des parties expérimentale et pratique de ce mémoire de master en décrivant les différents processus qui ont abouti à la réalisation de ***v0x***, un jeu dont le *gameplay* repose sur l'analyse spectrale de la voix du joueur.

L'intégration de la voix du joueur au gameplay

Bien que la majorité des jeux vidéo propose actuellement des écritures vocales uni-directionnelles, de nombreux jeux témoignent comme autant d'expériences et d'essais un intérêt historique pour l'évolution des interfaces de jeu par l'ajout de périphériques de détection sonore.

En effet, le *Famicom* (pour *Family Computer*), version japonaise du *NES* de Nintendo, était différente des versions occidentales de la console sur de nombreux aspects, dont la présence d'un microphone intégré à la manette de jeu. Cependant, peu de jeux exploitent les possibilités offertes par cet appareil.

Nous pouvons néanmoins remarquer son utilisation au sein de certains niveaux de la version japonaise de ***The Legend of Zelda***. Le joueur pouvait hurler dans le microphone pour immobiliser et vaincre un certain type d'ennemis. Cette fonctionnalité n'était pas cependant exploitée outre-mesure, encore moins pour la résolution d'énigmes essentielles à la progression du joueur.

Il faut attendre l'apparition de la licence ***Pokémon*** pour observer une nouvelle tentative d'intégration et d'utilisation d'un microphone au sein de l'interface de jeu. En 1998, le jeu ***Hey You Pikachu !*** (Nintendo), sur *Nintendo 64*, est un aboutissement logique des jeux centrés sur l'interaction avec un animal de compagnie virtuel.

En utilisant le microphone d'un périphérique externe de la *N64*, la *Voice Recognition Unit*, le joueur peut interagir vocalement avec Pikachu, la mascotte de la licence ***Pokémon***. Pour cela, il peut activer le microphone pour lire à voix haute des mots ou des séries de mots, formant des commandes, interprétables par Pikachu. Cependant, ce dernier dispose d'un vocabulaire restreint d'environ 200 mots et ne peut comprendre que les mots apparaissant en rouge à l'écran. Malgré ces restrictions, Pikachu est capable d'obéir à des commandes complexes comme *« Stay at my house »*

ou de répondre par des gestes ou des expressions faciales à différentes assertions, dont *« You're so cute »* ou *« Thunder ![54] »*.

Par ailleurs, les notices américaines et japonaises du jeu indiquent respectivement que l'Anglais et le Japonais sont les deux seules langues dans lesquelles le joueur peut espérer communiquer avec le jeu.

Enfin, l'interface de jeu intègre différents éléments de *feedback* visuels indiquant que les sons émis par le joueur sont reçus et interprétés dans des conditions satisfaisantes ou au contraire que le niveau sonore auquel ils ont été émis ne permet pas de les interpréter.

Si le jeu rencontre un certain succès, sa pratique par des adultes révèle des dysfonctionnements. Les procédés de détection vocale ont en effet été calibrés pour des voix d'enfants, plus aiguës que celle d'un utilisateur plus âgé.

Par la suite, avec la sixième génération de consoles, et plus particulièrement la *Dreamcast* de Sega, la *Playstation 2* de Sony et la *GameCube* de Nintendo, on distingue alors l'apparition de trois tendances dans la mise en œuvre de technologies de détection vocale :

- proposer au joueur de converser apparemment librement avec des personnages non-joueur.
- transmettre à des personnages non-joueur des commandes correspondant à des suites de mots-clé aisément détectables et reconnaissables par le programme du jeu.

54 [Traduction] : Reste chez moi. Tu es si mignon. Attaque Éclair !

- détecter la voix du joueur et les variations de ses caractéristiques acoustiques.

Seaman (Vivarium, SEGA, 1999), sur Dreamcast, est le principal représentant de la première tendance ci-dessus. Dans la lignée de ***Hey you Pikachu !***, ***Seaman*** permet au joueur d'interagir vocalement avec un animal virtuel, en l'occurrence un poisson arborant un visage humain en lieu et place de sa tête. Néanmoins, contrairement à Pikachu, le Seaman est capable d'étendre son vocabulaire à de nouveaux termes spécifiques au joueur. Il pourra donc, au cours de sa croissance, apprendre et mémoriser différentes informations sur son maître qu'il pourra par la suite restituer ou ré-employer.

Illustration 43: Le Seaman.

Seaman est cependant une expérience artistique plutôt qu'un jeu vidéo, au regard des caractéristiques du jeu et de sa dimension ludique limitée. Les sujets de conversation affectionnés plus particulièrement par le Seaman sont extrêmement divers et concernent des concepts complexes comme la liberté, Internet ou la censure. Il aborde de plus ces sujets-là avec un cynisme et un dédain qui limitent le public cible du jeu à un public adulte curieux et à la recherche d'expériences vidéoludiques

étranges.

Le Seaman pourra par exemple questionner le joueur pour savoir s'il communique avec ses amis par *e-mail* ou par téléphone. Si le joueur répond *« e-mail, »* la créature lui répondra avec humour qu'il le comprend et qu'il serait bien incapable de retranscrire par des mots les émoticônes (en japonais *emoji*) florissant alors dans les courriels et les SMS japonais. Cette réponse, particulièrement décalée, est emblématique de l'étrangeté que cultive ce jeu, qui le place à l'avant-garde technique et artistique des jeux vidéo intégrant un microphone et des technologies de détection sonore.

La deuxième tendance s'illustre essentiellement par son utilisation dans des jeux de tir tactiques ou dans des jeux de stratégie. La série de jeux ***SOCOM : US Navy Seals*** (Zipper Interactive, Sony Computer Entertainment, 2002), sur *Playstation 2* place le joueur à la tête d'une escouade de soldats d'élite auquel il peut donner des ordres via un microphone connecté au port USB de la console. Pour ce faire, le joueur doit prononcer une suite de noms de codes et de verbes pour que le soldat concerné de l'escouade effectue l'action désirée.

Par exemple, pour que les soldats identifiés par le nom de code Bravo utilisent une grenade, le joueur doit prononcer la phrase : *« Bravo Deploy Frag. »*

Ce système sera repris et amélioré avec la sortie en 2008 de ***Tom Clancy's Endwar*** (Ubisoft Shanghai, Ubisoft) sur *Xbox 360*, *Playstation 3* et ordinateur personnel. Dans ce jeu de stratégie en temps réel, ce n'est plus une escouade que le

joueur peut contrôler par sa voix, mais une armée entière, de l'unité d'infanterie la plus basique au véhicule d'artillerie le plus sophistiqué.

En cours de partie, la prononciation par le joueur du nom de code d'une unité déclenche l'apparition à l'écran d'un menu en arborescence permettant à la fois d'indiquer les options disponibles et de confirmer au joueur que sa requête a bien été prise en compte. Une fois la commande complète, elle est exécutée par l'unité sollicitée.

Illustration 44: Capture d'écran de Tom Clancy's Endwar. Le menu déroulant de commande vocale apparaît à l'écran.

Le système est très performant et une proportion infime de commandes vocales est rejetée, nécessitant une seconde prononciation. Cela est rendu possible par la restriction du vocabulaire compris par le jeu aux différents termes qui composent les commandes ainsi que par la performance des algorithmes de détection et de reconnaissance du langage utilisés.

Enfin, la troisième tendance est associée aux jeux de karaoké dont ***Singstar*** (Sony Computer Entertainment London, Sony Computer Entertainment, 2004) est un représentant notoire.

La septième génération de consoles, constituée de la *Wii* de Nintendo, de la *Xbox 360* de Microsoft et de la *Playstation 3* de Sony voit l'avènement puis l'abandon du *motion gaming*. De plus en plus de jeux pouvant être contrôlés par des mouvements du joueur voient le jour.

Si la caméra *EyeToy* de Sony, périphérique externe dédié à la *Playstation 2* et commercialisé dès 2003 constituait une première tentative de l'industrie en la matière, Nintendo choisit avec la *Wii* d'occuper une position radicale. Dès sa sortie en 2006, la console de la firme japonaise se distingue très nettement de ses deux concurrentes, la *Xbox 360* et la *PS3,* commercialisées respectivement à partir de 2005 et 2006. En effet, la *Wii* n'utilise pas de manette traditionnelle, mais deux contrôleurs appairés : le *Nunchuk* et la *Wiimote*. Cette dernière, dont le nom dérive du terme *remote controller,* désignant en anglais les télécommandes de téléviseur, fonctionne en interaction avec une barre de diodes électroluminescentes émettant des lumières infrarouges que la *Wiimote* va détecter pour déterminer la position et l'orientation du contrôleur par rapport à l'axe que constitue la barre de diodes. Enfin, des accéléromètres intégrés à la *Wiimote* permettent de détecter des variations d'inclinaison de l'appareil.

Illustration 45: La Wii accomagnée du Nunchuk (au centre) et de la Wiimote (à droite)

Illustration 46: La Sensor Bar de la Wii, contenant une série de diodes électroluminescentes

Combinés, ces différents outils permettent de mettre en œuvre des jeux dont le *gameplay* repose sur des interactions gestuelles permises par cette réinvention de l'**ergonomie** vidéoludique. Le jeu ***Wii Sports***, commercialisé avec la console, permet au joueur de s'improviser boxeur ou joueur de tennis et d'affronter un ami ou une intelligence artificielle en mimant les gestes d'un sportif.

La console de Nintendo rencontre un immense succès auprès d'un public étonnamment large. La grande accessibilité et l'intuitivité de leurs **ergonomies** lui permettent d'atteindre des publics jusqu'alors peu attirés par le jeu vidéo, dont des personnes âgées.

En 2010, face à la réussite commerciale et critique de leur concurrent, Microsoft et Sony commercialisent des périphériques de détection de mouvement pour leurs consoles. Si le *PS Move* est esthétiquement et fonctionnellement très similaire aux

contrôleurs de la *Wii*, Microsoft se démarque à son tour avec le capteur *Kinect*.

Illustration 47: La deuxième version de la **XBox 360** *accompagnée du capteur* **Kinect**

Tout d'abord, *Kinect* propose un *motion gaming* sans contrôleur en main. Le périphérique comporte en effet un regroupement de capteurs vidéo et infrarouges permettant le suivi précis des différentes parties du corps du joueur.

De plus, quatre microphones intégrés délivrent un signal audio PCM 24-bits traité au sein de l'appareil par des algorithmes de suppression d'écho et de réduction du bruit ambiant. Ce signal microphonique peut être utilisé par une application ou un jeu pour mettre en place des interactions vocales impliquant des sons émis par le joueur.

Par exemple, le jeu d'horreur ***Rise of Nightmares*** (Sega, 2011) réutilise une idée développée par les développeurs de ***Manhunt*** (Rockstar North, Take2 Interactive), jeu sorti sur *PS2* en 2003, dans lequel le joueur contrôle un tueur en série pouvant attirer l'attention de ses victimes en émettant des sons captés par le microphone USB de la *PS2*. Dans ***Rise of Nightmares***, le principe est inversé : le joueur doit rester silencieux pour ne pas attirer à lui les créatures qui le poursuivent.

D'autre part, le kit de développement de *Kinect* comprend des modules logiciels de reconnaissance vocale adaptés à différentes langues[55] et permettant au joueur de contrôler vocalement sa console en prononçant des séries de commandes.

Malgré la qualité de ses fonctionnalités vocales, *Kinect* est peu à peu délaissé par les joueurs et les développeurs qui lui reprochent l'imprécision de son module de détection et de suivi de mouvement.

Une certaine lassitude du public amène peu à peu les constructeurs de consoles à abandonner le *motion gaming*[56], faute de développeurs attirés par ces technologies et donc de jeux convaincants. Néanmoins, les périphériques de détection vocale ne sont pas victimes du même désintérêt. Ainsi, Microsoft développe une deuxième version de *Kinect* commercialisée à partir de 2013 avec sa nouvelle console, la *Xbox One*.

55 *http://www.microsoft.com/en-us/download/details.aspx?id=43662*

56 *Article collectif,* **« Le Motion Gaming, cette étoile filante, »** *JV, n° 3, janvier 2014, p. 49.*

Illustration 48: La **XBox One** *et* **Kinect 2***.*

Kinect 2 propose une détection plus performante de la voix du joueur et augmente les possibilités de contrôle vocal de la console. De plus, dans la mesure où le capteur a été commercialisé avec la *Xbox One* au lancement de celle-ci, le nombre de jeux exploitant, même de façon simple et discrète les possibilités offertes par le capteur va augmenter sensiblement.

Par exemple, les fonctionnalités de contrôle de l'interface visuelle et en particulier, la navigation dans les menus de ***Dead Rising 3*** (Capcom Vancouver, Capcom, 2013) peut être effectuée par la prononciation des commandes appropriées.

Chez la concurrence, les microphones vont disparaître dans un premier temps des consoles de salon au profit des consoles portables. La *Nintendo DS* ainsi que la *Playstation Vita* comportent toutes les deux un microphone que quelques jeux

exploitent savamment.

Nintendo développe la série de jeux ***Nintendogs*** (Nintendo EAD, Nintendo, 2005) qui propose au joueur de s'occuper d'un chien virtuel. Dans la lignée de ***Hey You Pikachu!*** ou de ***Seaman***, le joueur peut communiquer avec le chien pour le dresser, le congratuler ou le réprimander.

Par ailleurs, Nintendo utilise également le microphone de sa console portable pour d'autres types de jeux et notamment l'inclut de nouveau dans le *gameplay* de jeux appartenant à la série ***The Legend of Zelda***.

Cependant, les interactions vocales permises par le microphone dans ces jeux se limitent généralement à détecter si le joueur souffle sur le microphone. Ce procédé, qui correspond à une détection de niveau sonore ou de saturation du microphone est néanmoins utilisé très intelligemment dans ***The Legend of Zelda : Spirit Tracks*** (Nintendo EAD, Nintendo, 2009). Le joueur doit en effet souffler en rythme sur le microphone en suivant différentes instructions visuelles et rythmiques pour utiliser une flûte magique indispensable à la résolution de certaines énigmes.

Sur la *PS Vita* de Sony, le constat est sensiblement le même. Très peu de jeux emploient le microphone intégré à la console. Les interactions mises en œuvre lorsqu'il est utilisé par un jeu n'occupent pas une place centrale dans le gameplay et sont généralement basiques voire accessoires, malgré des idées rafraîchissantes et sophistiquées.

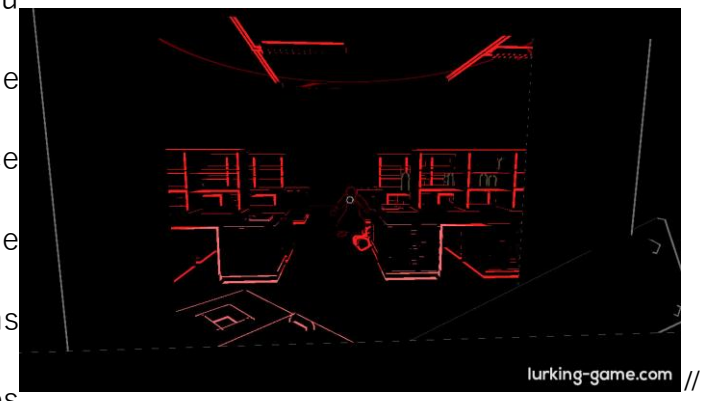
Illustration 49: La citrouille de **TearAway**.

Par exemple, dans ***TearAway*** (Media Molecule, Sony Computer Entertainment, 2013),

conçu par ses développeurs comme une vitrine pour les différentes interactions rendues possibles par la console portable de Sony (utilisation des pavés tactiles avant et arrière, des deux caméras, du microphone), une citrouille demande au joueur de pousser un cri le plus terrifiant possible. Elle reproduira par la suite ce son pour faire s'éloigner des corbeaux bloquant la progression du joueur.

Si ces exemples sont autant de témoignages de l'intérêt limité mais manifeste des constructeurs de consoles, les scènes indépendantes[57] foisonnent de propositions de *gameplay* novatrices, dont certaines reposent sur l'utilisation de la voix du joueur.

A ce titre, ***Lurking***[58] est un jeu d'horreur utilisant les sons émis par le joueur pour cartographier l'espace autour de lui, comme le ferait une chauve-souris. Cependant, ces sons sont susceptibles d'attirer des créatures malintentionnées.



lustration 50: Dans Lurking, l'environnement se révèle peu à peu au joueur au gré de ses respirations.

La voix est dans ce jeu utilisée très intelligemment et placée au cœur du *gameplay*, amenant le joueur à prendre conscience des sons qu'il émet ou qui l'entourent. Ce principe fondateur de l'expérience procurée par ce jeu amène un

57 *Le terme* indépendant *désigne une équipe de développement qui publierait son jeu sans l'assistance d'un éditeur. Cependant, dans la mesure où des éditeurs indépendants, comme* Focus Home Interactive *ou* Devolver Digital *ont récemment vu le jour, la signification de ce terme est aujourd'hui plus floue.*

58 *Ce jeu est en réalité un projet de fin d'études d'étudiant du* Digipen Institute of Technology *de Singapour.*

supplément considérable d'immersion et permet de créer très rapidement des situations de tension dramatique intenses.

Une conséquence du verbo-centrisme : les dialogues interactifs

L'évolution de l'utilisation des technologies de détection et de reconnaissance vocales est contemporaine de l'émergence d'une autre tendance répondant au même désir de donner une voix au joueur. Celle-ci se manifeste par l'apparition de ***dialogues interactifs***, c'est-à dire de dialogues au cours desquels le joueur va décider de ce que le personnage incarné va dire ou faire.

Encore une fois, cette tendance est un héritage des jeux de rôle papier dans lesquels les joueurs incarnent des personnages fictifs et les interprètent au cours de phases de dialogues avec les personnages des autres joueurs ou avec des personnages non-joueur incarnés par le meneur de jeu.

Ce type de jeux est apparu et s'est démocratisé au départ en Amérique du Nord à partir des jeux de stratégie (de plateau ou par correspondance) puis à la suite du succès grandissant de ***Donjons & Dragons***[59]. Les phases de dialogues y ont une importance toute particulière, au point que la capacité d'un personnage à convaincre ou à comprendre le discours d'un personnage non-joueur et d'en déduire des informations y est généralement régie par un système de statistiques définissant en quelque sorte le profil du personnage incarné par le rôliste.

---

59 Raphaël Lucas, **« Avant *Donjons et Dragons*, »** *Level Up, Niveau 1, op. cit., p. 10-13.*

Ces systèmes de jeux, basés sur des statistiques, des lois de probabilité et des évènements générés, au gré du scénario ou selon l'envie du meneur de jeu, se sont très tôt prêtés à des adaptations vidéo-ludiques. Sans meneur de jeu, les évènements, voire même l'environnement dans lequel évoluent les personnages sont générés aléatoirement selon l'**axiomatique** du jeu. Des systèmes de dialogues à progression par arborescence apparaissent mais sont, comme nous l'avons évoqué précédemment, exclusivement textuels.

Avec l'augmentation des puissances de calcul des machines et des capacités de stockage, les jeux vidéo de rôle vont s'ouvrir considérablement aux joueurs et leur offrir de plus en plus de marge de manœuvre au sein du jeu pour personnaliser leur avatar et déterminer les actions effectuées par ce dernier, parmi des choix toujours plus nombreux.

Dans cette dynamique, les systèmes de dialogue évoluent également et relèguent le texte à une fonction de sous-titre pour laisser place à des vocalités véhiculant le contenu des dialogues. Au cours de ces phases de dialogues, le joueur sélectionne la prochaine phrase prononcée par son avatar parmi différents choix indiqués à l'écran sous la forme de mots-clés, d'humeurs ou de phrases complètes.

L'exemple récent le plus notoire et ayant acquis la plus grande reconnaissance du public est la trilogie de jeux ***Mass Effect*** (Bioware, EA Games, 2007). Les trois jeux constituent un *space opera* influencé entre autres par ***Star Wars*** et ***Star Trek*** dans

lequel le joueur incarne le commandant Shepard. En début de partie, il détermine son sexe, son apparence physique ainsi que son histoire personnelle et sa personnalité. Dès la première mission, ces différents éléments sont mis en exergue au cours de dialogues dont le déroulement est conditionné par les différents choix de questions ou de réponses effectués par le joueur parmi différentes options disposées autour d'une roue contextuelle.

Deux ans après la sortie du premier opus, ce système de dialogue fait des émules et Bioware en développe un pendant plus tactique et situé dans un univers médiéval fantastique à ***Mass Effect*** : ***Dragon Age Origins*** (2009). Ces deux séries de jeux vont profondément marquer et renouveler les jeux de rôle par le dynamisme et la qualité de leurs dialogues, mis en scène avec maîtrise et écrits avec une intelligence rare.

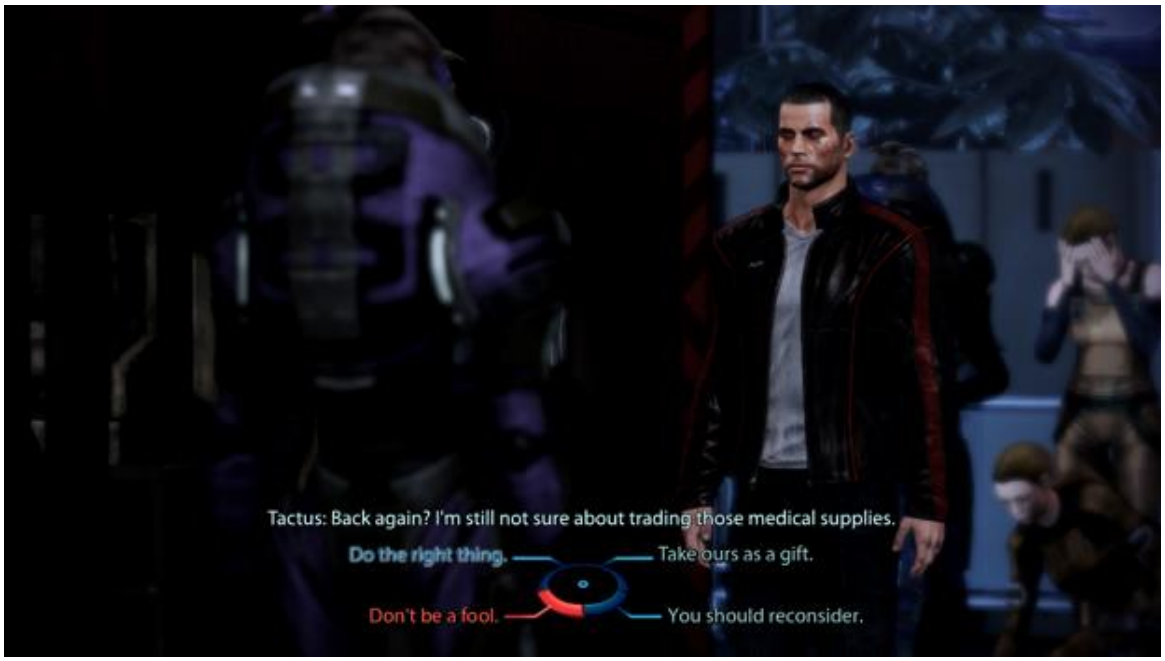


Illustration 51: Sur cette capture d'écran de **Mass Effect 3***, on peut observer une roue de sélection de dialogue. En plus des choix classiques, à droite, des choix liés à l'alignement du joueur (si Donjons et Dragons distingue le Bon du Mauvais, avec des nuances, Mass Effect fait évoluer ses personnages entre la Conciliation et le Pragmatisme). Le bleu désigne les choix permettant de gagner des points de Conciliation tandis que le rouge indique ceux qui*

L'émergence de systèmes de dialogues toujours plus poussés témoigne d'une volonté des développeurs – ou d'un désir des joueurs – d'interagir verbalement, par l'intermédiaire de leur avatar, avec les personnages non-joueur rencontrés.

Pour contourner les difficultés imposées par la mise en place de technologies de reconnaissance du langage sans recourir à des systèmes de dialogues écrits, parfois lourds et néfastes pour la fluidité et le rythme de la narration, des développeurs imaginent des méthodes alternatives pour donner au joueur la sensation tant recherchée de communiquer avec un jeu ou avec les créatures qui l'habitent.

Pour n'en citer que deux, ***Oddworld : l'Odyssée d'Abe*** (Oddworld Inhabitants, GT Interactive, 1997) et ***Baten Kaitos*** (Monolith Software et Tri-Crescendo, Namco, 2003), reposent sur deux propositions de *gameplay* très différentes mais permettant

une implication profonde du joueur.

Dans le jeu d'Oddworld Inhabitants, un système de dialogues permet à Abe, incarné par le joueur, de communiquer avec ses pairs via des combinaisons de touches aboutissant à la prononciation de phrases, d'interjections ou d'éructations par le personnage dans le but de susciter une réponse de son destinataire. Ce système de jeu, nommé *Gamespeak* par les développeurs, indispensable à la résolution d'énigmes

*Illustration 52: Abe entouré de ses congénères, les Mudokons, dans le remake de 2015 de l'***Odyssée d'Abe**, **New n' Tasty**.

complexes se révèle être un élément nécessaire pour atteindre un des objectifs secrets du jeu : sauver tous les congénères d'Abe et les amener en lieu sûr.

Le jeu de rôle de Monolith Soft et Tri-Crescendo est une proposition tout à fait intéressante par le simple fait que le joueur participe activement à la narration en tant qu'ange gardien du personnage principal, Kalas, et non en incarnant ce dernier ou en étant spectateur d'un scénario qui ne lui laisserait aucune place. Cette proposition narrative présente deux intérêts remarquables : donner

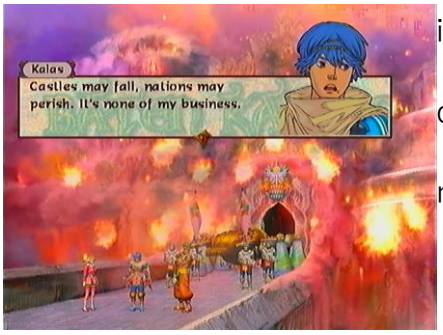


Illustration 53: Un dialogue de Baten Kaitos dans lequel Kalas, son personnage principal, s'éloigne des canons du genre, faisant

au personnage principal une véritable profondeur voire une « existence propre », expression utilisée par Georges Grouard[60] ; et offrir au joueur « son plus beau rôle » avec à la clé l'estime des personnages de ***Baten Kaitos*** et le sentiment d'avoir plus que jamais sauvé le monde.

Que ce soit dans les jeux de rôle à arbres de dialogues ou dans les deux derniers jeux évoqués, les développeurs ont fait le choix d'impliquer le joueur par une simulation d'interactions verbales permises par le *gameplay*. Ce faisant, l'accent est mis sur la dimension sociale du langage et de la parole ainsi que sur les rapports hiérarchiques ou affectifs qu'ils mettent à jour au sein d'un récit.

Il y aurait donc un Graal du jeu de rôle, qui unirait l'implication physique du joueur par sa voix à la richesse et à la profondeur atteintes par les systèmes de dialogues déjà en place à l'heure actuelle. Les nouvelles possibilités d'interaction offertes laissent rêveur.

60 Georges « Jay » Grouard, **« Baten Kaitos : Magna Carta, »** *Level Up, Niveau 1, op. cit., p. 145*

Reconnaissance du langage et jeu vidéo : une association souvent infructueuse

Illustration 54: L'assistant personnel intégré au système d'exploitation des Windows Phone, Cortana, tel qu'il apparaît à l'écran

Pourtant, à l'heure où il est possible, lors d'un trajet en transport en commun, de demander en chuchotant à son smartphone[61] de trouver le restaurant le plus proche de sa destination, comment expliquer que le jeu vidéo n'ait pas d'ores et déjà investi les possibilités offertes par la reconnaissance du langage ?

61 *Nous remarquerons d'ailleurs, non sans cynisme, que l'assistant personnel intégré au système d'exploitation pour smartphones de Microsoft est nommé Cortana, en référence à leur saga phare* **Halo**.

Pour quelles raisons ne conversons-nous pas avec les personnages non-joueur d'un jeu alors qu'il nous est aujourd'hui possible de le faire avec un assistant personnel doté d'une intelligence artificielle ?

Les principes de fonctionnement et la multiplicité des méthodes de mise en œuvre de ces technologies sont à l'origine de la difficulté de leur adaptation au médium vidéo-ludique.

En effet, depuis les premiers travaux scientifiques sur le sujet, de nombreux procédés différents de détection et de reconnaissance automatique de la parole ont été développés. Très peu d'entre eux sont effectivement passés de l'état de prototype à celui de « système assez général et fiable pour être industrialisé.[62] » Les deux adjectifs employés dans cette citation, « général » et « fiable » sont révélateurs.

Ainsi, pour être industrialisé et commercialisé pour le plus grand nombre et donc quitter les laboratoires de recherche pour les ordinateurs personnels, les consoles ou un autre appareil électronique, un système de reconnaissance du langage doit être performant indépendamment de l'utilisateur et de l'environnement dans lequel il est utilisé. De fait, des facteurs de complexité ont été définis pour qualifier et quantifier les contraintes auxquelles un tel système doit répondre et en déterminer le fonctionnement.

Dans leur ouvrage, les rédacteurs de l'équipe de recherche CALLIOPE recensent les facteurs de complexité suivants :

62 *CALLIOPE,* **« La parole et son traitement automatique, »** *Masson, Collection technique et scientifique des télécommunications, 1989, p. 492.*

◆ « Langage » :

La reconnaissance de phrases plus ou moins complexes par le système est profondément dépendante du langage de l'utilisateur. L'intégration de ses règles syntaxiques aux algorithmes de détection de mots séparés va permettre au système de reconnaître la fonction de chacun d'entre eux au sein de la phrase afin d'en déterminer le sens.

De même, le nombre de syllabes qui composent la langue le rendent plus ou moins aisé à analyser.

Par exemple, la structure du japonais, constitué d'une centaine de syllabes principales associées à autant d'idéogrammes a permis le développement précoce de systèmes très performants de conversion de signaux vocaux en documents textuels spécifiques à cette langue.

◆ « Reconnaissance mono-locuteur, ou non » :

Ce facteur décrit les données sur lesquelles se base le système pour analyser la parole de l'utilisateur. Si ces données sont issues d'un locuteur unique, pouvant être l'utilisateur, on parlera de reconnaissance mono-locuteur.

À l'inverse, si pour détecter et reconnaître les mots prononcés par un utilisateur, le système se base sur des données issues de locuteurs multiples, on parlera de reconnaissance multi-locuteur. Ce dernier type de reconnaissance permet la mise en œuvre d'algorithmes d'apprentissage par le système, grâce à une collecte de données vocales auprès de ses utilisateurs et à des opérations de moyennage, de

discrimination et de classification de ces données.

En constituant, pour chacun des mots de son lexique, une base de référence d'empreintes acoustiques issues de locuteurs parfois très différents, un système multi-locuteur présente l'avantage de se développer perpétuellement du simple fait de son utilisation.

◆ « Reconnaissance de mots isolés ou de parole continue » :

Les auteurs précisent que « si le(s) locuteur(s) marque(nt) une pause après chaque mot de l'énoncé, la complexité du problème est réduite[63]. » Les différents mots prononcés sont dès lors plus facilement identifiables et comparables à la base de données de référence du système : son lexique.

Cependant, la reconnaissance de mots séparés requiert l'établissement de règles d'analyse définissant une tolérance du système. Autrement dit, il doit être conçu pour fonctionner même si le mot analysé ne correspond pas exactement à l'empreinte acoustique de référence correspondante. Il faudra ainsi définir des paramètres de distance spectrale, de compression temporelle et de comparaison dynamique pour que le timbre du locuteur et son élocution ne nuisent pas à l'analyse de son discours et à sa comparaison aux données de référence du système.

La détection et la reconnaissance de phrases plus ou moins complexes prononcées avec fluidité sont logiquement plus difficiles à réaliser. La reconnaissance de parole continue suppose une connaissance de la syntaxe du langage analysé par le système ainsi que l'introduction de règles de segmentation des phrases et de

63 *CALLIOPE, op.cit., p. 493.*

détection des frontières des mots, essentiellement par comparaison dynamique.

◆ « Taille du vocabulaire » :

Ce critère détermine le nombre de mots pouvant être reconnus par le système. Plus son lexique est étendu, plus la complexité du système augmente, en particulier si les mots qui le composent sont phonétiquement proches les uns des autres.

◆ « Environnement » :

Les caractéristiques acoustiques du lieu dans lequel le système va être utilisé constituent un ensemble de paramètres pouvant parasiter le signal vocal de l'utilisateur et donc perturber son analyse par le système. Le contenu spectral ainsi que le niveau du bruit environnant sont les deux paramètres principaux à prendre en compte.

Ces facteurs de complexité conditionnent la conception des systèmes de reconnaissance de la parole, que l'on peut dès lors diviser en deux types : ceux qui adoptent une approche globale, permettant une reconnaissance mono-locuteur ou multi-locuteur indépendante de l'utilisateur, et ceux qui, au contraire, reposent sur une identification de l'utilisateur indispensable à son bon fonctionnement.

En mettant en relation l'historique de l'application au jeu vidéo de technologies de reconnaissance vocale avec ces cinq facteurs de complexité ainsi que les deux types de systèmes qui en découlent, nous pouvons désormais comprendre pourquoi

l'industrie vidéo-ludique s'est peu à peu détournée de ces technologies.

Les processus de création et de commercialisation d'un jeu vidéo obéissent à une série de contraintes supplémentaires encore plus spécifiques, limitant les choix disponibles pour les développeurs quant au système de reconnaissance le plus adapté au jeu en cours de conception.

En effet, quand bien même un jeu aurait un *gameplay* très spécifique le destinant à un public de niche, son édition et sa publication à une échelle nationale, voire internationale, imposent un système de reconnaissance dont les performances sont indépendantes du joueur, de son environnement de jeu et de sa langue.

De fait, les systèmes de reconnaissance de la parole qui ont été appliqués au jeu vidéo sont principalement des systèmes multi-locuteurs conçus pour être utilisés dans un environnement sonore calme.

Pour limiter plus encore le parasitage du signal vocal de l'utilisateur par des sons indésirables, le signal audio est traité par des algorithmes de suppression d'écho et de réduction de bruit de fond afin d'en augmenter le rapport signal sur bruit.

Les langages détectés et reconnus par ces systèmes, à l'heure actuelle en nombre très limité, restreignent l'accès à ces technologies à des développeurs exclusivement occidentaux ou japonais.

Par exemple, le capteur *Kinect* de Microsoft fonctionne pour des utilisateurs s'exprimant en français, en allemand, en italien, en japonais ainsi qu'en espagnol et en

anglais. Pour ces deux dernières langues, le *Kinect* est en mesure de différencier des accents régionaux et distingue notamment l'anglais britannique de l'anglais néo-zélandais ou le castillan de l'espagnol couramment parlé au Mexique.

Néanmoins, dans la mesure où les langages de programmation informatique reposent sur l'utilisation de mots issus de la langue anglaise, cette limitation du nombre de langages reconnaissables pénalise assez peu les développeurs. Les utilisateurs ne maîtrisant aucune des langues reconnues par le système ne pourront pas s'en servir et tirer profit de toutes les fonctionnalités de leur jeux.

Par ailleurs, les systèmes de reconnaissance vocale peuvent être d'autant plus difficiles à adapter au jeu vidéo que ceux-ci sont susceptibles de présenter des lexiques spécifiques, composés de néologismes ou de noms de lieux et de personnes fictifs. Un jeu vidéo devra donc inclure dans son programme soit un module de reconnaissance de la parole à part entière, qui sera en général accompagné d'un périphérique de détection articulé autour d'un microphone (c'est le cas pour ***Hey You Pikachu !***), soit une extension pour un module de reconnaissance déjà implanté au sein d'une interface de jeu, permettant d'étendre son lexique aux mots fréquemment employés dans le jeu. Ce dernier principe de fonctionnement est, par exemple, celui qui est utilisé sur les consoles de Microsoft employant un capteur *Kinect*.

Enfin, un autre élément, bien plus difficile à appréhender, car lié à la subjectivité du joueur, est susceptible de conditionner l'adhésion de celui-ci à l'univers d'un jeu partiellement basé sur la reconnaissance de la parole : l'**ergonomie** du jeu.

Un jeu que le joueur contrôle par sa voix en lui donnant des directives vocales est-il agréable ? La reconnaissance de la parole ne complexifie-t-elle pas l'interface de jeu au point d'en faire un obstacle venant s'immiscer entre le joueur et l'univers du jeu ?

La fréquentation de forums de discussions, la lecture d'articles de journaux spécialisés et de blogs traitant de jeux intégrant des principes de reconnaissance de la parole sont révélatrices d'un sentiment de déception et d'insatisfaction de la part des joueurs[64]. La moindre imprécision du système utilisé, pouvant amener le joueur à répéter en vain une même commande vocale, est vécue comme une interruption de la fluidité du jeu, comme un rappel de l'interface. Celle-ci est alors révélée pour ce qu'elle est : un prisme par lequel le joueur fait l'expérience sensorielle du jeu et agissant comme un filtre conditionnant la perception de l'univers du jeu. Qui plus est, la prédominance des systèmes de reconnaissance de mots séparés sur les systèmes de reconnaissance de parole continue au sein des interfaces de jeu est à l'origine d'une scansion du discours du joueur pouvant être vécue comme dissonante par rapport au rythme du jeu. De même, l'élocution et la prosodie imposée au joueur par le système de reconnaissance peut lui donner la sensation de s'adresser à l'interface ou au jeu plutôt qu'aux personnages avec lesquels il est censé converser.

Un système de reconnaissance dont l'ergonomie ne serait pas irréprochable par sa précision et sa réactivité peut donc être un obstacle au bon fonctionnement mécanique du jeu, à la fluidité de sa narration et à l'intensité de l'immersion qu'il procure.

---

64 *http://calmdowntom.com/2010/10/top-ten-reasons-kinect-sucks/*

L'application de technologies de reconnaissance de la parole est donc susceptible de mettre en péril en tout point le pacte ludique d'un jeu. Pour cette raison, elle représente un risque considérable pour les développeurs qui souhaiteraient l'appliquer. Un risque d'autant plus conséquent que ces technologies sont très coûteuses à développer ou à implanter et que leur maîtrise totale requiert une très bonne compréhension des principes informatiques et mathématiques qui régissent leur fonctionnement. A l'heure où les coûts et les durées de développement des jeux dits « AAA » atteignent chaque année de nouveaux sommets, la frilosité des développeurs et des éditeurs vis-à-vis de ces technologies est un secret de polichinelle.

Cependant, l'émergence récente des scènes indépendantes et de leur influence sur les exigences en terme de créativité ainsi que la sophistication et la démocratisation croissantes des technologies de reconnaissance de la parole augurent un renouveau de leur utilisation. Celui-ci passera probablement par un détournement de ces technologies afin d'en pallier les faiblesses.

Par exemple, le jeu ***Keep Talking and Nobody Explodes***, actuellement en cours de développement par Steel Crate Games, est une relecture de ***Bomb Squad***, évoqué dans le premier chapitre de ce mémoire.

En effet, si un joueur, équipé d'un casque de réalité virtuelle, est placé dans la position du démineur confronté à une bombe à désamorcer, le personnage non-joueur qui le guidait dans ***Bomb Squad*** est ici remplacé par un second joueur ayant devant

lui un mode d'emploi de la bombe. Les deux joueurs sont reliés par un système de communication vocale pour que le démineur décrive d'abord la bombe dont le circuit électronique est généré aléatoirement à chaque nouvelle partie. Un dialogue débute ensuite entre le démineur et son conseiller pour que celui-ci consulte le mode d'emploi de la bombe et guide son partenaire en fonction de ses descriptions.

Illustration 55: La bombe telle qu'elle est vue par le démineur de Keep Talking et Nobody Explodes.

Illustration 56: Aux côtés du joueur, ses coéquipiers décortiquent le mode d'emploi de la

Véritable exploration des possibilités offertes par la réalité virtuelle, ***Keep Talking and Nobody Explodes*** est symptomatique d'un intérêt croissant pour la mise en œuvre d'interactions vocales à la fois innovantes, intuitives et complexes.

Par ailleurs, le jeu de Steel Crates Games utilise savamment les principes de communication vocale entre les joueurs, principalement utilisés au cours de parties en ligne pour se coordonner ou se guider mutuellement. Ils récupèrent, ce faisant, les éléments de méta-jeu que peuvent véhiculer ces modes de communication.

En effet, lorsque les joueurs d'un jeu de rôle en ligne se retrouvent sur un chat vocal pour y incarner verbalement leurs personnages et donner ainsi leur voix à un paladin orque engoncé dans sa lourde armure ou à un elfe voleur malicieux, les

interactions vocales qui les lient tendent à quitter ce qu'Axel Stockburger[65] nomme « l'environnement utilisateur » pour rejoindre « l'environnement de jeu. »

Dès lors, la communication vocale entre les joueurs tend à les inclure dans une méta-narration, à défaut d'appartenir au *gameplay* du jeu. Ces démarches d'incarnation des personnages par leurs joueurs est également symptomatique d'un désir, que nous avons déjà évoqué, de dialogue avec le jeu et les personnages qui l'habitent.

L'analyse spectrale de la voix du joueur : une alternative accessible et attrayante

Dans cette même volonté d'explorer de nouvelles possibilités d'intégration de la voix du joueur au *gameplay*, nous avons étudié l'emploi de techniques de détection et d'analyse sonores indépendantes de tout langage. Pour ce faire, nous avons choisi d'étudier un des principaux paramètres acoustiques de la voix d'un joueur : son spectre.

La voix d'un individu résulte de l'excitation d'un milieu par ses organes phonatoires. On distingue deux principales sources des sons vocaux : les cavités supra-glottiques (cavité buccale et fosses nasales) et le larynx.

65 *Axel Stockburger,* **The Game Environment from an Auditive Perspective**, *Level Up, Digital Games Research Conference, Utrecht, 2003.*

Au sein de ce dernier se trouvent les cordes vocales. Ce terme très répandu désigne en réalité un ensemble complexe de tissus et de fluides indispensables à la phonation :

- les muscles thyro-aryténoïdiens régulent la tension des cordes vocales.
- les ligaments thyro-aryténoïdiens inférieurs réunissent les cartilages qui constituent le larynx « entre eux ou avec les organes voisins[66]. »
- une muqueuse qui « recouvre la paroi interne du larynx[67] » et est également mise en vibration lors de la phonation. Son influence sur les caractéristiques de la voix d'un individu est d'autant plus importante qu'elle présente une « certaine indépendance [mécanique] par rapport à la couche musculaire.

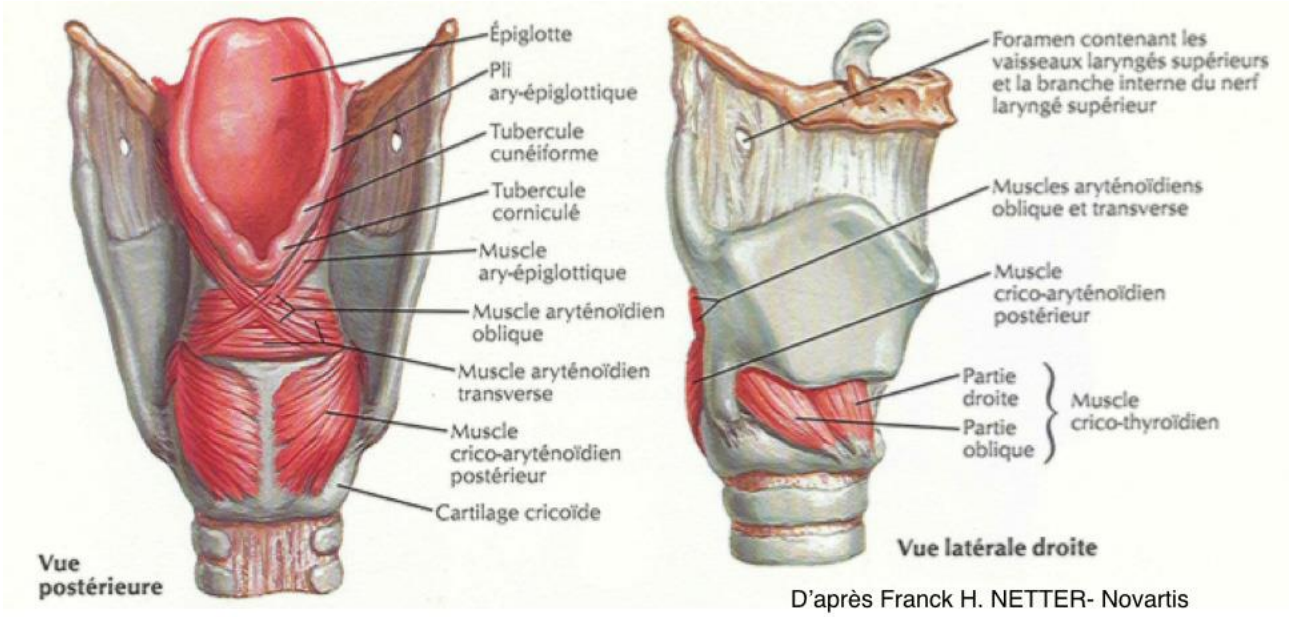


Illustration 57: Un schéma anatomique légendé du larynx.

En s'écoulant dans le larynx, l'air fait vibrer les cordes vocales, séparées en deux parties, et oscillant selon une fréquence dépendant de leur tension par les muscles thyro-aryténoïdiens.

66 *CALLIOPE, op. cit., p. 22.*

67 *Ibidem, p. 22.*

L'écoulement d'air est en retour perturbé par la vibration des cordes vocales. Un son est émis. Il est, au cours de son trajet du larynx aux lèvres du locuteur, altéré en raison de l'interdépendance de différents facteurs dont « les propriétés aérodynamiques de l'air qui excite le larynx, l'ajustement des cordes vocales déterminé par l'activité nerveuse de ses différents muscles [⋯], l'interaction mécanique entre les cordes vocales[68] » et le couplage aérodynamique entre le larynx et les cavités intervenant dans la phonation.

Les rédacteurs de ***La parole et son traitement automatique*** décrivent synthétiquement le phénomène de la phonation de la façon suivante :

> « En d'autres termes, les cordes vocales vibrent sous l'effet du passage de l'air à travers la glotte [⋯]. A la vibration purement mécanique se superpose un contrôle et des ajustements opérés par les muscles du larynx excités par le système nerveux. La répartition de la masse, la tension longitudinale, la compression latérale, la disposition des cordes vocales étant, dans toutes leurs nuances, déterminées par l'excitation des muscles du larynx. Comme pour toute source, le régime de vibration du larynx peut être influencé par le couplage avec les éléments auxquels il est connecté[69]. »

Les différents paramètres qui influent sur le larynx ainsi que sur les cavités résonantes auxquelles il est couplé définissent le timbre de la voix d'un individu ainsi que sa tessiture vocale. Les différences morphologiques et physiologiques entre deux individus, qu'elles soient liées à leur âge ou à leur condition physique, vont donc être

68 Ibidem, p. 27.
69 Ibidem, p. 27.

synonymes de timbres vocaux extrêmement variés. Cela indique que la voix est un son complexe, d'autant plus riche fréquentiellement que son contenu spectral évolue considérablement entre la voix chantée et la voix parlée d'un même individu.

Par ailleurs, la cavité buccale, située au terme du trajet de l'écoulement d'air originaire du larynx, peut amener une diversité de sons émis encore plus conséquente. En effet, par l'occlusion du conduit vocal par les lèvres ou des variations de la position de la langue par rapport à son palais ou à sa dentition, un individu est en mesure de perturber l'écoulement d'air pour façonner le son émis.

Dès lors, différentes méthodes d'analyse spectrale de la voix sont susceptibles de fournir au programme d'un jeu des informations permettant d'interagir avec lui. En particulier, lors du voisement, c'est-à dire de la vibration des cordes vocales, il est possible d'identifier une fréquence qui serait une estimation de la fréquence de vibration laryngée. Par la suite, nous appellerons cette fréquence « fréquence fondamentale » et utiliserons le symbole F_0 pour la désigner.

Il est essentiel d'assimiler le fait que l'analyse de F_0 est discriminante du simple fait qu'elle présuppose une vibration des cordes vocales. De fait, comme l'indique les rédacteurs de CALLIOPE, les sons dont la source est « une explosion produite après occlusion du conduit vocal, » comme les consonnes occlusives (p, t, k, b, d, g), beaucoup plus complexes que les sons voisés, ne peuvent pas être analysés si le voisement est implicitement présupposé lors de l'analyse. Ces sons seront donc considérés comme du bruit par rapport aux sons voisés que l'on souhaite analyser et

seront, si possible, identifiés comme tel par l'appareil ou algorithme d'analyse.

Par la suite, on distingue deux grandes familles d'analyseurs de F_0 opérant dans deux domaines différents : les domaines temporel et fréquentiel.

METHODES D'ANALYSE TEMPORELLE

Ce type d'analyse est désigné généralement par les termes de méthode de Prony ou de méthode des coefficients de prédiction linéaire (en anglais LPC pour Linear Prediction Coefficients) et implique l'adoption d'une modélisation du signal vocal grâce à un modèle source-filtre qui considère le signal vocal comme un bruit traité par un filtre ou une série de filtres excités par un train d'impulsion de fréquence F_0. Dans ce modèle, le train d'impulsion correspond à la contribution des cordes vocales, le bruit blanc à la contribution de phénomènes de friction au sein du larynx et le filtre à la contribution du de l'ensemble du conduit vocal.

Les analyseurs temporels identifient la période fondamentale T_0 (définie telle que $T_0 = \frac{1}{F_0}$) par une détection des passages par zéro de la valeur de l'amplitude du signal vocal complexe, filtré au préalable pour en atténuer les harmoniques et ainsi faire correspondre le nombre de passages par zéro du signal filtré à celui du signal laryngé avant qu'il ne traverse les cavités buccales et nasales. Une déconvolution du signal par la fonction de transfert du filtre plus ou moins élaboré qui a été déterminé

au sein du modèle permet l'identification de cycles temporels afin d'évaluer T_0 puis F_0.

Cependant, certains sons voisés, comme le son voyelle [u] présentent d'importantes composantes harmoniques qui compliquent la conception du filtre à appliquer au signal. Cette méthode est, par ailleurs, susceptible de déphaser certaines des composantes harmoniques du signal, ce qui pourra, selon CALLIOPE, « provoquer des erreurs dans la mesure des périodes successives[70]. »

Pour pallier les défauts potentiels de ces analyseurs, il est possible de traiter le signal filtré pour « renforcer l'amplitude de la fondamentale » ou de lui appliquer une fonction d'autocorrélation. Grâce à une fenêtre d'analyse de durée programmable, cette fonction permettra de comparer les périodes successives détectées et de déduire une valeur plus précise de T_0.

METHODES D'ANALYSE FREQUENTIELLE

L'analyse spectrale d'un signal audionumérique est généralement réalisée par l'application d'un algorithme de transformée de Fourier rapide (en anglais FFT pour Fast Fourier Transform), en particulier lorsque l'analyse doit être réalisée en temps réel.

Ces algorithmes réalisent une série d'opérations mathématiques correspondant aux calculs nécessaires à la réalisation d'une transformée de Fourier discrète (en

70 *CALLIOPE, op. cit., p. 304.*

anglais DFT pour Discrete Fourier Transform), permettant de convertir le signal du domaine temporel vers le domaine fréquentiel en vue de son analyse.

Pour appliquer une FFT à un signal vocal, il faut le numériser au préalable. Dès lors, l'étude de son spectre en temps réel requiert le choix d'une fenêtre d'analyse dont la taille définit la fréquence à laquelle l'algorithme analyse le signal. Sans cette fenêtre, l'analyse porterait sur l'intégralité du signal sonore et permettait d'observer l'évolution de son contenu spectral, seulement au terme de l'ensemble des calculs.

Deux paramètres de réglage de la fenêtre vont s'avérer cruciaux pour que les résultats de la FFT soient d'une précision suffisamment satisfaisante par rapport à l'exploitation qui en sera faite.

Ainsi, le premier paramètre, la taille de la fenêtre, c'est-à-dire le nombre d'échantillons audionumériques que l'algorithme analyse à chaque prélèvement d'information dans le signal, permet de déterminer la résolution fréquentielle de l'analyse. Une résolution élevée apportera une meilleure précision aux résultats au prix de calculs supplémentaires pouvant requérir d'importantes ressources informatiques.

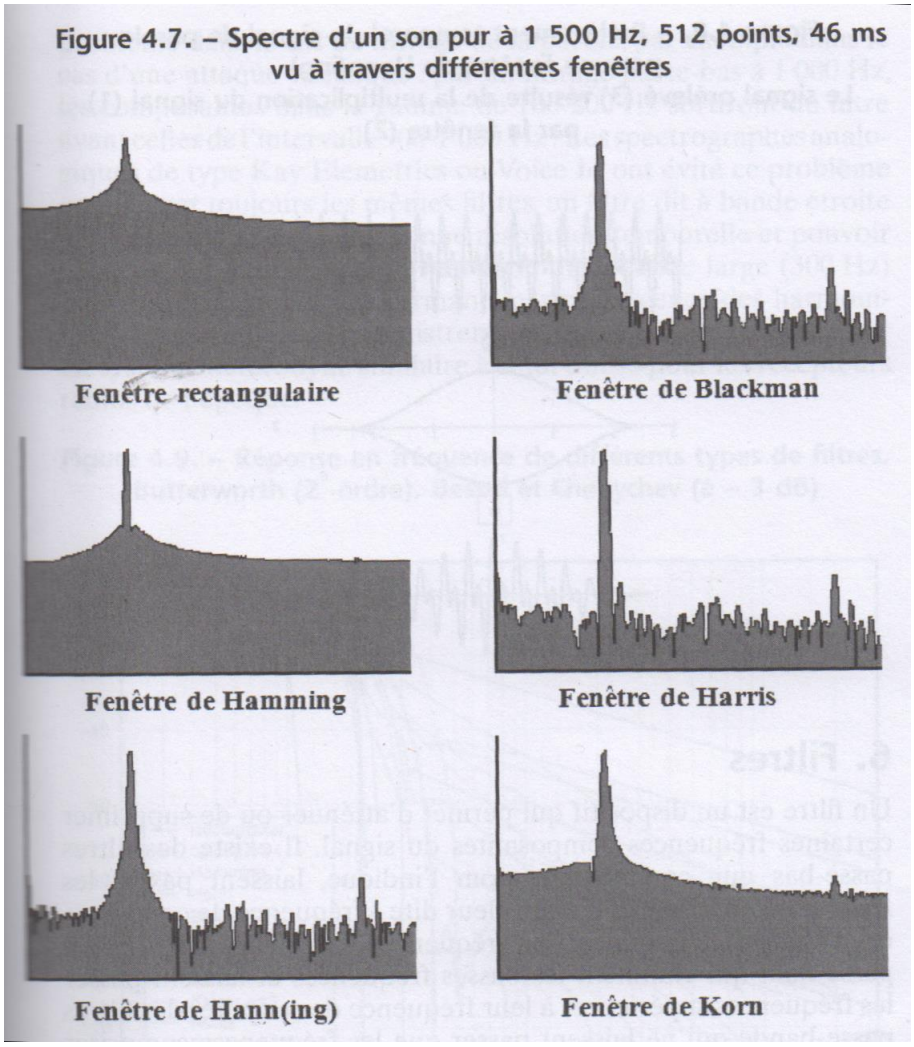


Illustration 58: Sur cette figure, issue de l'ouvrage **Phonétique acoustique : introduction à l'analyse acoustique de la parole***, de Philippe Martin, on peut observer l'influence du type de fenêtre sur le*

Le second paramètre, plus crucial encore, mais aussi plus technique, concerne la type de la fenêtre d'analyse. La fenêtre d'analyse doit être comprise comme un filtre opérant en quelque sorte un second échantillonnage. A ce titre, le fenêtrage peut provoquer des distorsions en réalisant une troncature violente du signal. Pour limiter ces altérations du signal tout à fait nuisibles à la pertinence des résultats de l'analyse, deux solutions ont été élaborées. La fenêtre d'analyse la plus simple à réaliser, la fenêtre rectangulaire, prélève l'intégralité du signal et réalise justement des troncatures brutales du signal à ses extrémités. D'autres fenêtres, correspondant à

autant de fonctions mathématiques précises appliquées au signal ont peu à peu vu le jour. Les plus utilisées sont la fenêtre en cosinus, la fenêtre triangulaire, la fenêtre de Hamming, la fenêtre de Hanning ou encore la fenêtre de Blackman-Harris. Chacune d'entre elles fait subir au signal une série d'opérations mathématiques supplémentaires afin de limiter le plus possible les distorsions causées par le fenêtrage.

Par ailleurs, il est possible de paramétrer une FFT pour que les fenêtres d'analyse se recouvrent au lieu de se succéder simplement. La mise en place d'un recouvrement des fenêtres dans le temps, combinée au choix d'une fenêtre adaptée, limite les erreurs d'analyse.

Ceci étant dit, une FFT ne permet pas d'identifier immédiatement F_0. Elle établit en réalité un spectrogramme qui est une représentation en deux dimensions des contributions de chaque fréquence au spectre du son analysé. Afin de déduire F_0 des résultats d'une FFT, il faut procéder, selon CALLIOPE, à « l'évaluation du plus grand commun diviseur des maxima du spectre d'amplitude71. » Cela permet, même dans un cas de fondamental absent, d'identifier F_0 comme la fréquence ayant le plus d'harmoniques supérieures au sein du spectre. Pour cela, on applique un lissage cepstral qui va réaliser les opérations nécessaires, en accord avec un modèle source-conduit différent des modèles utilisés dans les méthodes d'analyse temporelles, car distinguant les contributions de la glotte de celles des conduits oral et nasal.

71 *CALLIOPE, op. cit., p. 304.*

Quelle que soit la méthode utilisée, la détection de la fréquence fondamentale d'un son permet de tracer une courbe mélodique indiquant l'évolution de F_0 tout au long du signal.

Dès lors, rien ne s'oppose à l'utilisation de la valeur de F_0 comme d'un paramètre de suivi de la fréquence fondamentale d'une voix chantée. Par la suite, la définition de fréquences à atteindre, accompagnées de seuils de tolérance, pourra donner lieu à l'évaluation de la justesse d'une note produite par le joueur dans le cadre d'un jeu de karaoké, par exemple.

Cependant, dans le cadre de la partie pratique de ce mémoire, nous avons appliqué ces principes d'analyse du contenu spectral de la voix du joueur pour les placer au centre d'un jeu vidéo d'un tout autre genre que les jeux musicaux afin de proposer un *gameplay* inhabituel, même pour des joueurs réguliers.

v0x *: un exemple d'emploi de l'analyse spectrale comme mécanique de gameplay*

v0x est un jeu pour ordinateur personnel que l'on pourrait classer parmi les jeux de plateformes et de réflexion en deux dimensions dont ***Braid*** (Number None, 2008) et ***Teslagrad*** (Rain Games, 2013) sont deux représentants récents. Ainsi, si le premier repose sur une manipulation du temps au sein du jeu et si le deuxième est centré sur l'utilisation des propriétés magnétiques d'objets du décor, ***v0x***, jeu expérimental, place la voix du joueur au centre de son *gameplay*.

Dans un premier temps, nous avons déterminé les différents paramètres de la voix du joueur qui allaient être pris en compte. Notre choix s'est porté sur l'intensité sonore et la fréquence fondamentale de la voix du joueur. La première étant extrêmement simple à détecter, nous avons choisi de l'inclure au sein du *gameplay* du jeu en complément de la fréquence fondamentale dont l'intégration représentait un

intérêt plus prononcé.

Dès les premières phases de la conception du jeu, nous avons décidé de mettre en place une trame narrative très simple, servie par une esthétique minimaliste : ***un agent d'une entreprise fictive est numérisé et introduit dans un réseau informatique pour y éliminer un virus***. Cette volonté de simplicité et de sobriété est due à un besoin de concentrer nos efforts sur la programmation du jeu et sur la conception d'interactions vocales fonctionnelles intéressantes.

Les dix tableaux qui composent le jeu décrivent donc la découverte de ce nouvel environnement par l'agent, incarné par le joueur puis sa conversion en programme.

Le choix de la thématique d'un parcours initiatique est propice à la mise en place d'interactions vocales. Le joueur va devoir les découvrir puis les maîtriser pour progresser dans le jeu.

Les tableaux et leur fonctionnement ont tout d'abord été conçus sur le papier, via une série de croquis préparatoires, afin de donner une cohérence globale à l'ensemble et de faciliter la communication avec Nicolas, étudiant en informatique, pour que la réalisation du jeu en binôme soit des plus aisées. Dès lors, les éléments visuels et sonores constituant le jeu ont été intégrés au sein du moteur de jeu vidéo Unity3D 5.

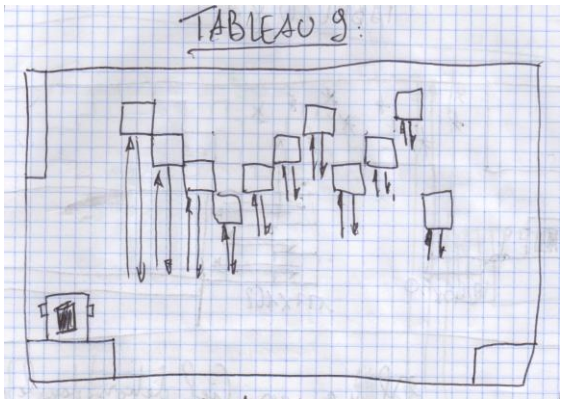


Illustration 59: Le croquis préparatoire du Tableau 9 : La Salle du Temps

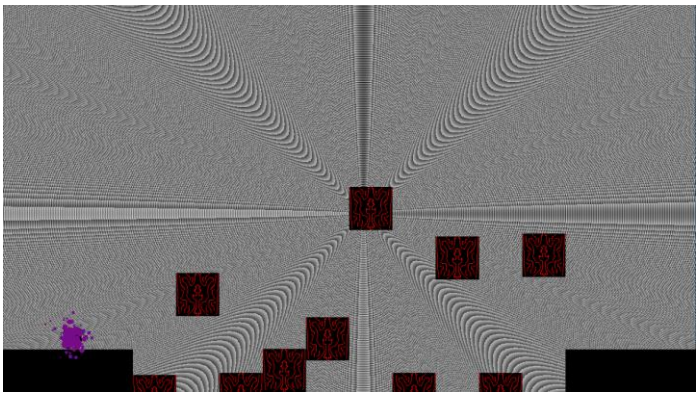


Illustration 60: Le Tableau 9 tel qu'il apparaît en jeu.

Unity3D est un moteur d'édition de jeux vidéo développé par Unity Technologies accompagné d'une interface de programmation (en anglais API pour *Application Programming Interface*), MonoDevelop, indispensable à la programmation de scripts. En effet, à la différence des tout premiers moteurs de jeu vidéo, Unity est à la fois très complet et très accessible pour des utilisateurs dont les compétences de programmation sont peu développées. Il est donc possible de créer des jeux basiques sans avoir besoin d'ouvrir MonoDevelop grâce à un système orienté-objet.

Dans Unity, le développeur crée puis édite des *GameObjects* auquel il peut associer des *Components* qui vont donner au *GameObject* des propriétés précises, comme le fait d'être soumis à la gravité ou de se déplacer lorsque le joueur appuie sur les flèches directionnelles de son clavier.

Par la suite, le développeur pourra disposer les différents *GameObjects* dans un espace virtuel en trois dimensions et définir les éléments visuels et sonores qui seront transmis au joueur par le biais de l'interface de jeu.

Néanmoins, notre volonté de tirer parti des fonctionnalités sonores natives de Unity 5 impliquait nécessairement de programmer ce que Unity nomme des scripts.

Dans Unity 5, les scripts sont des fichiers informatiques, programmés dans notre cas en C#, qui déterminent les modalités d'interactions entre les différents *GameObjects* ainsi que leur gestion par l'interface de jeu. Pour cela, des éléments de programmation tels que des variables ou des fonctions sont associés à chaque *GameObject* ainsi qu'à chaque *Component* afin d'en régir le comportement.

Ainsi, l'API de Unity comprend une classe permettant l'utilisation d'un microphone intégré à l'interface de jeu. En nous inspirant d'un script programmé par le développeur finlandais Teppo Kaupinnen[72], nous avons créé des outils de détection et d'analyse sonore qui correspondait à notre cahier des charges.

Toutes les interactions vocales sont régies par trois scripts :

- Le script MicrophoneInput indique au jeu quel microphone doit être utilisé. Dans notre cas, il utilise un microphone par défaut, identifié par la valeur 0.
- Le script AudioAnalysis récupère et traite les informations fournies par le microphone.
- Le script AudioInputEvent indique les paramètres de la voix du joueur pris en compte pour interagir avec les objets du décor.

Si les deux premiers scripts sont associés au joueur, le troisième est, lui, associé aux microphones visibles à l'écran. Dès lors, il est possible de définir autour de ces indices visuels des zones, dans lesquelles le script AudioInputEvent est actif, afin d'indiquer au joueur à quel endroit de l'écran il doit se placer pour émettre des sons.

72 http://www.kaappine.fi/about-me/

L'étude de ces trois scripts nous permettra de comprendre comment fonctionnent les interactions vocales de v0x.

Voici le script MicrophoneInput.cs dans son intégralité, accompagné de commentaires.

```
using UnityEngine;using System.Collections;
// Cette première ligne indique quelles ressources issues du moteur le script va utiliser.

[RequireComponent(typeof(AudioSource))]
// Cette ligne ajoute un Component AudioSource au GameObject associé à MicrophoneInput.

// La partie fonctionnelle du script débute ici.
public class MicrophoneInput : Singleton<MicrophoneInput>
{
	public int mic ;
	public AudioSource audioSource
	{get{return GetComponent<AudioSource>();
	}
// Cette partie du script permet au développeur de définir le microphone par défaut que va solliciter le jeu grâce à l'entier « mic. »

void Start()
// La fonction Start est une fonction de base de l'API de Unity et permet d'initialiser un script pour déclarer des variables ou exécuter des fonctions au lancement du script.

	{
		if(AudioAnalysis.Instance == true)
		audioSource.clip = Microphone.Start(Microphone.devices[mic], true, 1
		, AudioAnalysis.Instance.samplerate);

		else audioSource.clip = Microphone.Start(Microphone.devices[mic], tr
		ue, 1, 44100);
```

```
            audioSource.loop = true;
            audioSource.mute = true;
            while (!(Microphone.GetPosition(Microphone.devices[mic]) > 0)){}
            audioSource.Play();
      }
}
```

// Nous avons dans cette fonction Start une première occurrence des mots *if* et *else* qui permettent de mettre en place des raisonnements conditionnels. Ici, nous pouvons traduire le code entre crochets par la phrase suivante :

« ***SI*** le script AudioAnalysis est également associé au même GameObject que MicrophoneInput, ***ALORS***, créer chaque seconde un fichier audio à partir du microphone spécifié à la fréquence d'échantillonnage déterminée par le script AudioAnalysis. ***AUTREMENT***, effectuer la même opération à la fréquence d'échantillonnage fixe de 44100 Hertz.»

Les dernières lignes du script définissent les paramètres de sortie audio de l'Audio Source : ce Component lira continuellement les fichiers audio générés à chaque seconde, mais à un volume nul pour que le joueur ne les entende pas.

Ce script définit donc le microphone qui va être utilisé par le script d'analyse sonore que nous allons maintenant décomposer.

```
using UnityEngine;using System.Collections;

[RequireComponent(typeof(AudioSource))]
public class AudioAnalysis : Singleton<AudioAnalysis>
{
      public float sensitivity = 100f;
      public int samplerate = 44100;
      public float loudness = 0f;
      public float trueLoudness = 0f;
      public float frequency = 0f;
      public float trueFrequency = 0f;
      public float loudnessDamping = 1f;
      public float frequencyDamping = 1f;
```

// Dans cette première partie du script, nous déclarons les différentes variables que le script va générer et traiter pour les envoyer vers les autres

scripts interactifs.

Sensitivity permet de contrôler le niveau d'entrée du fichier audio lu par le script.

Loudness et trueLoudness correspondent à l'intensité sonore des sons émis par le joueur.

Frequency et trueFrequency correspondent à une approximation de la fréquence fondamentale de ces sons.

Les deux variables se terminant par le terme Damping sont des coefficients de lissage qui vont permettre de rendre les évolutions des valeurs correspondantes moins brutales.

```
void Update()
```

// La fonction Update est, tout comme Start, une fonction de base de l'API d'Unity. Toutes les opérations qui sont programmées dans la partie Update du script sont effectuées à chaque image. Par exemple, si le jeu est exécuté à la cadence de 30 images par seconde, les opérations incluses dans l'Update seront exécutées 30 fois par seconde. Nous en déduisons qu'en utilisant l'Update, nous nous servons de l'image comme d'unité temporelle minimale sur les calculs d'évolution de variables au cours du temps.

```
{
	loudness = Mathf.Lerp(loudness,GetAveragedVolume() * sensitivity, loudness
	Damping * Time.deltaTime);
	frequency =  Mathf.Lerp(frequency,GetFundamentalFrequency(), frequencyDamp
	ing * Time.deltaTime);
	trueFrequency = GetFundamentalFrequency();
	trueLoudness = GetAveragedVolume() * sensitivity;
}
```

// Ici, les opérations de lissage des valeurs sont définies. On remarque donc que les coefficients Damping vont déterminer l'intensité du lissage, effectué par la fonction Mathf.Lerp.

```
	float GetAveragedVolume()
	{
		float[] data = new float[256];
		float a = 0;
		GetComponent<AudioSource>().GetOutputData(data,0);
		foreach(float s in data)
		a += Mathf.Abs(s);
```

```
            return a/256;
      }
```

// Le flottant GetAveragedVolume est calculé à partir des fichiers audio générés par le microphone grâce à la fonction GetOutputData qui va en déterminer le volume de sortie.

```
      float GetFundamentalFrequency()
      {
            float[] spectrum = new float[8192];
            GetComponent<AudioSource>().GetSpectrumData(spectrum,0,
            FFTWindow.BlackmanHarris);
            float s = 0.0f;
            int i = 0;
            for (int j = 1; j < 8192; j++)
            if ( s < spectrum[j] )
            {
                  s = spectrum[j];
                  i = j;
            }
      return i * samplerate / 8192;
    }
}
```

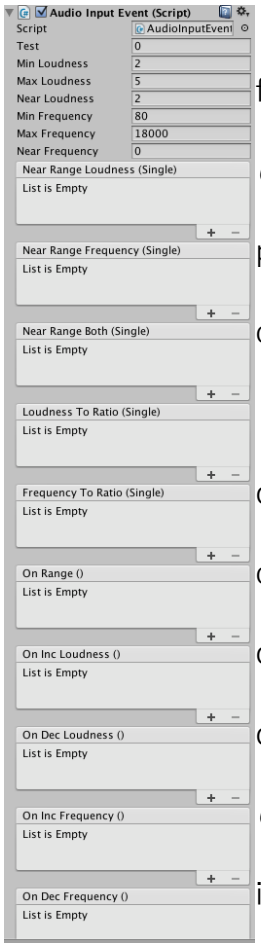
//C'est avec la définition du flottant GetFundamentalFrequency qu'est calculée l'approximation de F_0 à chaque image.

En effet, la dernière partie du script exécute une FFT qui va être appliquée à la sortie de l'Audio Source. Pour obtenir une résolution et une précision optimales, la taille de la fenêtre d'analyse est de 8192 échantillons et la fenêtre utilisée est la fenêtre de Blackman-Harris. Ce choix est motivé par des séries de tests comparatifs ayant montré que cette fenêtre était la plus adaptée.

D'autre part, nous avons jusqu'ici parlé d'approximation de F_0 du fait de la méthode opératoire du script. La variable Frequency ne correspond pas au sens strict à la fréquence fondamentale détectée, mais correspond en fait à la fréquence ayant l'amplitude la plus élevée au sein du spectre. Cependant, il est possible que cette fréquence ne soit pas F_0. Pour améliorer le script, il faudrait donc implémenter un calcul du plus petit commun diviseur des fréquences dont les amplitudes sont les plus élevées au sein du spectre.

Nous avons détaillé le fonctionnement des deux scripts d'analyse principaux.

Enfin, étudions le script AudioInputEvent, qui est au cœur de la quasi-totalité des interactions vocales au sein du jeu. Ce script permet, comme nous allons le voir, de récupérer les flottants *loudness*, *trueLoudness*, *frequency* et *trueFrequency* pour contrôler d'autres *GameObjects* ou d'autres *Components.*



Plutôt que décomposer le script, nous allons observer son fonctionnement au sein de l'éditeur. Une fois associé à un *GameObject*, AudioInputEvent prend la forme d'un *Component* présentant plusieurs paramètres réglables, mais aussi plusieurs champs à éditer, comme l'indique l'illustration ci-contre.

Les six paramètres réglables permettent de définir des seuils d'intensité ou des bandes passantes fréquentielles afin de pouvoir complexifier les interactions vocales. Le script permet en effet d'utiliser ce que le développeur a configuré par un système d'évènements. Parmi eux, par exemple, *OnIncLoudness* et *OnIncFrequency* permettent respectivement de comparer à chaque image la valeur de l'intensité sonore et de la fréquence des sons émis par le joueur par rapport aux valeurs de ces variables à l'image précédente. Si les valeurs à l'image *i* sont supérieures aux valeurs correspondantes à l'image $i - 1$, une fonction peut être appelée et exécutée.

À partir de ces outils logiciels, nous avons concrétisé au sein du jeu les interactions conçues sur le papier. Dès lors, nous avons pu passer à la dernière phase de la réalisation de ***v0x*** : son optimisation et la suppression d'aberrations de

fonctionnement qui risqueraient de rendre le jeu injouable.

Ainsi, nous avons testé le jeu de nombreuses fois pour nous assurer de son bon déroulement ainsi que de la précision des outils d'analyse. Le compte-rendu des parties expérimentale et pratique de ce mémoire, que le lecteur pourra consulter en annexe de ce mémoire, reviendront plus en profondeur sur cette phase.

Les démarches entreprises dans les derniers stades de développement du jeu nous ont essentiellement permis de déterminer les paramètres susceptibles de rendre le jeu instable et de modifier leur programmation, soit pour limiter leur influence, soit pour les remplacer par d'autres paramètres plus pertinents.

Par exemple, des tests comparatifs entre un microphone de mesure, le PRM1 de Presonus, et le microphone intégré de l'ordinateur portable73 sur lequel le jeu allait, au terme de sa conception, être soumis à l'appréciation de joueurs, ont été menés pour déterminer si le microphone intégré était suffisamment précis ou si le bon fonctionnement du jeu allait requérir son remplacement par un microphone externe. Nous avons donc enregistré différents stimuli vocaux auprès de locuteurs de trois tranches d'âge : des élèves de CM2, des élèves de Terminale et des étudiants de l'École Nationale Supérieure Louis Lumière.

L'étude se basait sur l'analyse de deux séries de phrases très spécifiques comportant respectivement des assonances74 et des allitérations75. Les locuteurs ont

---

73 Un MacBook Pro 13'' de fin 2011, équipé d'un processeur Intel Core i5 (2,4 GHz) et employant le système d'exploitation Mac OSX 10.8.5.

74 Première série de phrases : « Tout m'afflige et me nuit et conspire à me nuire. Au firmament qui dort, un soleil vient de naître, comme un papillon d'or. »

été enregistrés et leur voix ont été analysées grâce un logiciel d'analyse phonétique, Praat, fournissant d'une part des spectrogrammes, mais aussi des courbes mélodiques. Ces séries de mesures et leur étude ont permis de conclure que le microphone intégré était suffisamment performant, voire limitait l'apparition d'aberrations lors de l'enregistrement de respirations ou de consonnes dentales fricatives et occlusives.

75 *Seconde série de phrases : « Pour qui sont ces serpents qui sifflent sur vos têtes ? Il dort dans le soleil, la main sur sa poitrine. Tranquille, il a deux trous rouges au côté droit. »*

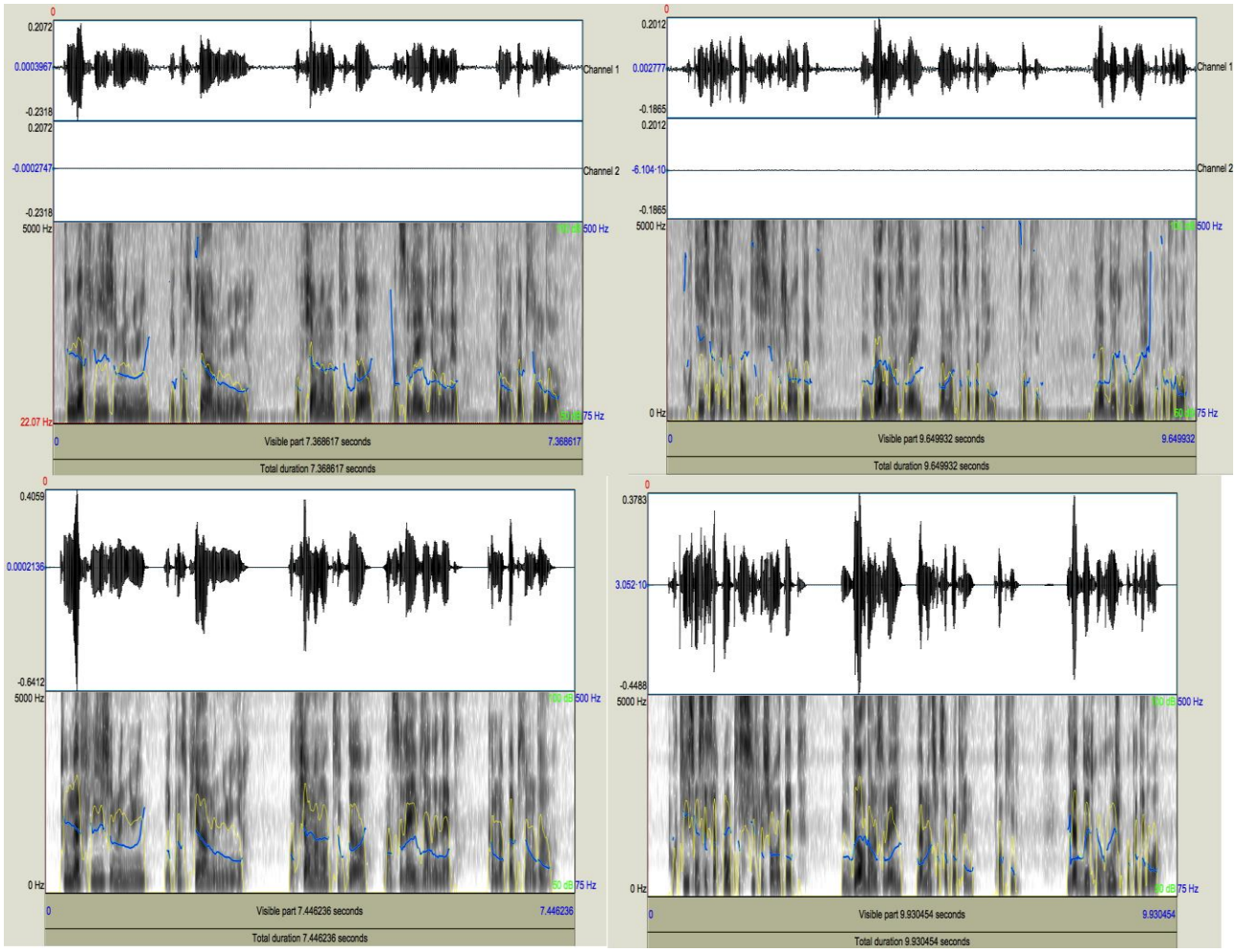


Illustration 61: Le logiciel d'analyse phonétique Praat permet d'éditer ce type de graphiques. Dans chacune des quatre fenêtres : en haut, une représentation temporelle du signal enregistré ; en bas, un spectrogramme sur lequel sont superposées deux lignes de suivi de la fréquence fondamentale (en bleu) et de l'amplitude du signal (en jaune). Les quatre fichiers analysés ont été enregistré auprès d'un même sujet, dans ce cas un étudiant de l'ENSLL.

En haut à gauche : La première série de phrases enregistrée avec le microphone Presonus PRM1.

En bas à gauche : La première série de phrases enregistrée avec le microphone intégré.

En haut à droite :La seconde série de phrases enregistrée avec le microphone Presonus PRM1.

Bien que d'autres tests comparatifs mériteraient d'être menés avec les microphones intégrés d'autres ordinateurs portables, nous avons conclu que ***v0x*** n'exigerait pas du joueur qu'il possède un microphone particulier pour faire fonctionner correctement le jeu.

Ces séries de tests nous ont également permis de déterminer d'autres critères susceptibles de rendre inexploitables les valeurs issues de l'analyse spectrale de la voix d'un joueur.

Nous avions émis l'hypothèse que l'âge du joueur, son sexe, le type de son émis, le type et la taille de fenêtre d'analyse ainsi que le type de microphone utilisé étaient des critères à risque.

Au terme des tests et au regard des principes de fonctionnement des scripts du jeu, nous avons réussi à valider ou infirmer nos hypothèses pour chaque critère.

Ainsi, de la même façon que nous avons conclu que la qualité du microphone avait une influence négligeable sur la qualité de l'analyse, nous avons observé qu'il en allait de même pour l'âge du joueur, son sexe ainsi que la taille de la fenêtre d'analyse.

A l'inverse, le type de fenêtre d'analyse influe sur la précision des résultats. Des imprécisions trop importantes, provoquées par exemple par le choix d'une fenêtre rectangulaire peuvent rendre erratiques les déplacements d'un ascenseur en fonction de la fréquence de la voix du joueur.

Nous avons identifié le type de son émis comme le critère le plus susceptible d'apporter de l'instabilité, du fait de la méthode d'analyse choisie. Comme nous l'avons expliqué, le script d'analyse audio ne détermine pas exactement la fréquence fondamentale d'un signal mais évalue la fréquence ayant l'amplitude la plus élevée au sein du spectre du signal. Lorsque le son émis est particulièrement riche en harmoniques, la fréquence déterminée par le script AudioAnalysis ne correspond pas à F_0.

Nous avons, au cours des tests, observé que certains sons, en particulier les

respirations ou les consonnes dentales occlusives et fricatives, pouvaient causer des évolutions brutales de la valeur de F_0, quand bien même l'algorithme de détection utilisé serait plus perfectionné que le nôtre (c'est le cas de l'algorithme de détection de fréquence fondamentale de Praat).

Afin de limiter l'influence de ce type de son sur la stabilité du jeu malgré les défauts de notre script d'analyse, nous avons configuré des seuils déterminant la borne supérieure de la bande de fréquences analysée. Ainsi, la plupart des scripts ne détectent pas de fréquence supérieure à 10 kHz environ, afin d'atténuer les imprécisions de notre script tout en laissant la possibilité au joueur de siffler pour résoudre les différentes énigmes du jeu.

Par ailleurs, nous craignions que les éléments sonores du jeu perturbent la détection et l'analyse de la voix du joueur. Ceci aurait motivé l'ajout d'une touche permettant, une fois enfoncée, d'activer les scripts correspondants et d'atténuer l'ensemble des sons diffusé par le jeu. Par un travail de mixage des différents objets sonores et de réglages des seuils de détection de l'intensité sonore, l'ajout de cette touche a été abandonné au profit d'une détection ininterrompue permettant également de donner des indices visuels au joueur quant à la détection de sa voix. Ainsi, la bouche du petit personnage qu'incarne le joueur grandit et change de couleur en fonction respectivement de l'intensité sonore et de la fréquence des sons émis par le joueur.

En effet, si les interactions vocales furent aisées à mettre en œuvre au sein du

jeu, la volonté de donner au joueur un minimum d'instructions, pour le laisser tâtonner et découvrir le jeu et ses possibilités par lui-même, nous a amené à en questionner l'ergonomie. Comment inviter le joueur à se servir de sa voix, sans lui donner les solutions des différentes énigmes qui composent le jeu, par des indices visuels ou sonores trop explicites ?

C'est lors de la présentation du jeu à un groupe de testeurs que les principaux ajustements ergonomiques ont été implémentés. En effet, si des éléments de *feedback* visuels – dont l'évolution de la couleur et de la taille de la bouche du personnage- et sonores – dont des sons indiquant le déclenchement d'un mécanisme par le joueur – avaient été intégrés au jeu avant sa présentation au groupe de testeurs, leurs différents retours ont permis de perfectionner l'ergonomie du jeu pour qu'elle soit plus intuitive, plus agréable et plus fonctionnelle.

Le jeu a été présenté à 26 participants au cours de tests répartis sur une durée de trois jours, dans une des salles de cours de l'ENSLL aménagée pour l'occasion. Les participants, principalement des étudiants de l'école, étaient âgés de 21 à 42 ans et présentaient des profils de joueur très différents. Les genre de jeux qu'ils affectionnaient le plus, la fréquence à laquelle chaque participant jouait à un jeu variait considérablement d'un participant à l'autre.

Chaque participant devait, au cours du test, terminer d'une traite le jeu. Seul face à l'ordinateur de jeu, le participant était encouragé à tâtonner et expérimenter. En cas de blocage ou d'incompréhension des objectifs d'un tableau, l'expérimentateur

présent dans la pièce pouvait donner des indices suffisamment vagues pour guider le joueur sans lui offrir la solution du tableau.

Au terme de sa partie, chaque participant devait répondre à un questionnaire destiné à recueillir des informations sur son profil ainsi que des retours sur son expérience de jeu. Ce questionnaire est disponible dans la section Annexes de ce mémoire. Les conditions matérielles de jeu furent strictement les mêmes pour chaque joueur : la luminosité de l'écran et le volume de sortie des haut-parleurs de l'ordinateur portable sur lequel était installé le jeu étaient tous deux réglés à leur valeur maximale.

Parmi les 26 participants, âgés de 21 à 42 ans, un seul n'a pas apprécié son expérience du jeu. Néanmoins, malgré leur apparent plaisir de jeu, de nombreux joueurs ont exprimé des réserves, en particulier par rapport à l'ergonomie du jeu.

Ainsi, 38,5% des participants se sont sentis plutôt mal à l'aise en utilisant leur voix pour contrôler le jeu. Des discussions avec ces participants ainsi que leurs réponses aux questions ouvertes du questionnaire ont indiqué que cette façon de jouer était inhabituelle pour eux et assez étonnante. Quatre participants ont manifesté leur sensation de malaise en cours de test en avouant se sentir démunis, voire idiots, à l'idée de parler, de chanter, de pousser des cris ou de frapper dans leurs mains face à un ordinateur portable.

Ces observations sont révélatrices. Le fait qu'un nombre non-négligeable de joueurs considèrent les interactions vocales mises en œuvre par ***v0x*** comme une source de malaise pourrait remettre en question le bien-fondé d'une intégration aussi profonde de telles interactions au sein du *gameplay* d'un jeu.

Cependant, 25 des 26 participants ont manifesté un intérêt pour les jeux présentant une interface inhabituelle incluant potentiellement d'autres périphériques qu'une manette de jeu classique76. De plus, une grande majorité des joueurs ayant testé le jeu (84,6% pour être exact), a déclaré pertinents les indices visuels affichés à l'écran l'informant que sa voix était prise en compte par le jeu.

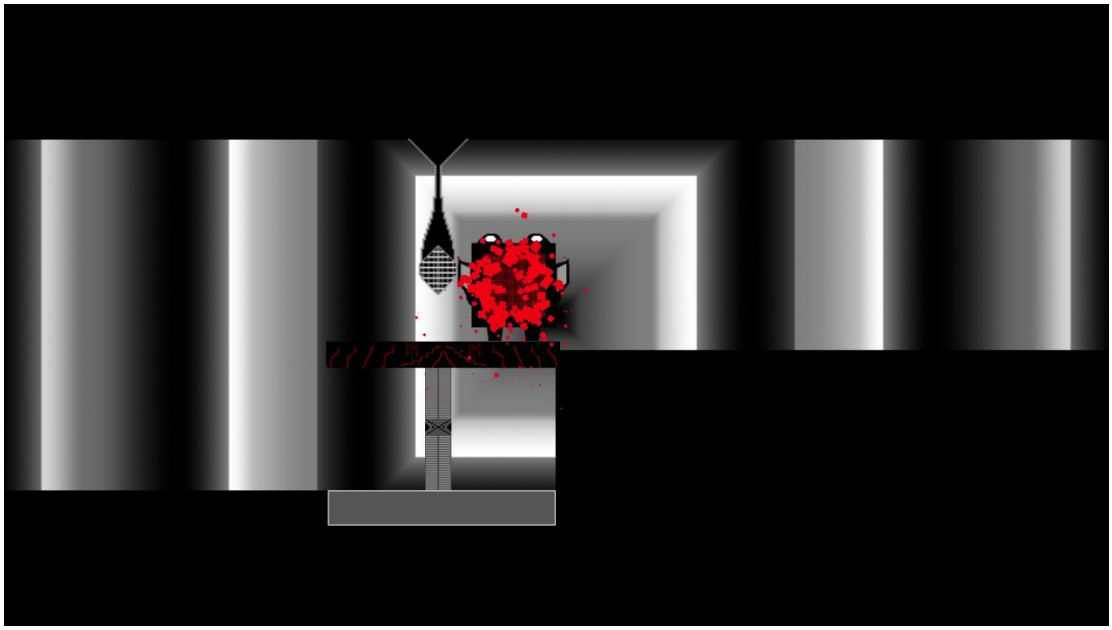


Illustration 62: Dans le tableau 4 de v0x, tous les éléments de feedback visuels sont mis en jeu : la bouche du personnage, le microphone au plafond et la plateforme mobile dont les motifs luisent d'un rouge vif

Ces indices visuels sont constitués de trois éléments :

- la bouche du personnage principal, évoluant en réaction aux différentes variations vocales du joueur ;
- des objets symbolisant des microphones et ayant été identifiés comme tels par 96,2% des participants ;
- des variations de luminosité, d'opacité et de couleur sur des motifs décoratifs

76 Le seul participant ayant indiqué ne pas être attiré par ce type de jeux a expliqué qu'il était friand de la bulle calme et propice à la concentration qui se formait autour d'un joueur seul. Il craignait que des gestes devant une caméra ou l'émission de sons à la destination d'un micro ne perturbe cette posture de jeu.

situés sur différents objets interactifs permettant au joueur de les localiser.

D'autre part, malgré la facilité de prise en main du jeu et le plaisir du tâtonnement et de l'exploration des possibilités du jeu, de nombreux participants ont, principalement au cours de discussions à l'issue du test, manifesté une forme de frustration. S'ils avaient bel et bien pris du plaisir en jouant, en particulier en essayant différents types de sons émis, allant de la simple voyelle chantée au sifflement en passant par des coups frappés en rythme sur la table sur laquelle était disposé l'ordinateur de jeu, ils n'ont pas toujours compris quel élément sonore avait permis la résolution de certaines énigmes.

Plus particulièrement, le dernier tableau, correspondant dans le scénario à la rencontre du personnage avec le Conseil, trois programmes sensés compléter sa conversion numérique, a été majoritairement considéré comme un pic considérable de difficulté.

Ce tableau a évolué au cours des tests. En effet, nous avons décidé de collecter les différentes suggestions des participants de chaque jour pour implémenter des modifications du jour pour le lendemain, afin de proposer une version plus perfectionnée du jeu aux participants du jour suivant. Cette démarche, contestable puisque les participants de jours différents n'ont pas joué au même jeu au sens strict du terme, a cependant permis de perfectionner le jeu et de comprendre plus précisément les sources de problèmes ou d'incompréhension au sein des différents tableaux qui composent le jeu.

Ainsi, si le principe du tableau 10 – émettre un son continu situé dans trois bandes successives de fréquences pour que les trois entités accordent au joueur un droit de passage et ainsi terminer le jeu – est resté inchangé, sa physionomie a beaucoup évolué.

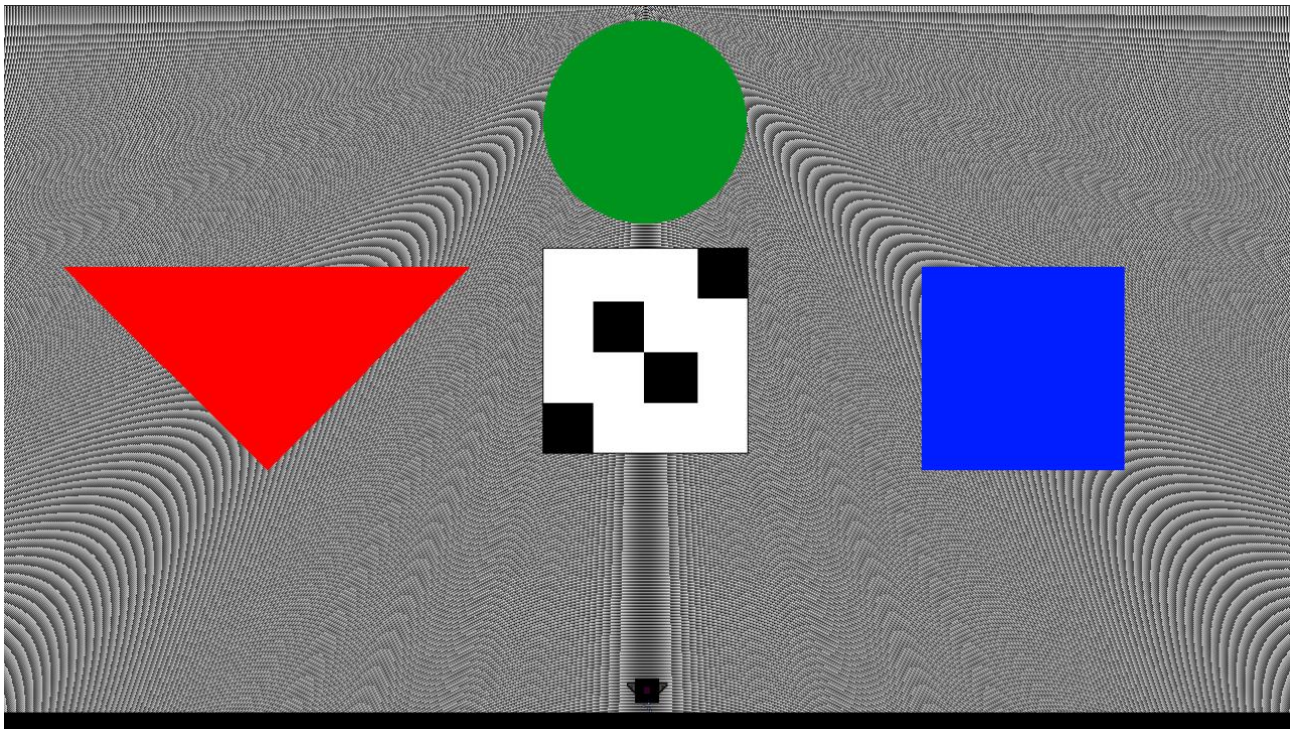


Illustration 63: Le Tableau 10, une fois résolu par le joueur.

Le premier jour, les participants entrant dans la dernière salle du jeu étaient confrontés à trois formes noires (un triangle, un disque et un carré). En atteignant le milieu de l'écran, un script ôtait la capacité du joueur à déplacer son personnage tandis qu'une des trois entités lui indiquait qu'il était écouté. Dès lors, à chaque forme géométrique était associé une bande passante dans lequel la voix du joueur devait se situer pour faire évoluer la couleur du noir vers leur couleur d'origine (rouge pour le triangle, vert pour le disque et bleu pour le carré). Une fois le carré devenu totalement bleu, le personnage du joueur se métamorphosait très bruyamment, perdant ses jambes et ses yeux, tandis qu'un son volontairement sous-échantillonné et sous-

quantifié était lu, pour symboliser sa conversion en un programme numérique. Dans cette configuration, l'objectif à atteindre pour terminer ce tableau était très difficile à comprendre et à deviner bien que l'énigme soit finalement assez facile à résoudre, les bandes de fréquence étant volontairement très larges et progressives : la bande de fréquences de l'entité rouge correspondait aux graves, celle de l'entité verte aux médiums et celle de l'entité bleue aux aigus.

Dès le deuxième jour, la modification suivante fut effectuée pour donner plus de clarté au tableau : une fois proche de sa couleur d'origine, chaque forme se mettait à clignoter pour indiquer que le joueur avait atteint l'objectif correspondant. Cependant, les plages de fréquences n'avaient pas été modifiées. Cela amena plusieurs joueurs à faire une même suggestion : faire correspondre les bandes de fréquence de chaque entité de couleur aux fréquences de la voix du joueur qui feraient correspondre la couleur de la bouche du personnage à celle de l'entité correspondante.

Cette modification fut implémentée le lendemain. En entrant dans la salle, le joueur était confronté aux trois membres du Conseil, dans leur couleur d'origine et non plus en noir. Puis, lors de la désactivation du script permettant au joueur de déplacer le personnage, les trois entités basculaient rapidement vers le noir. Ceci permettait au joueur d'identifier et de mémoriser les trois différentes couleurs pour essayer de les faire réapparaître au niveau de la bouche du personnage.

Ceci impliqua une modification des plages de fréquences associées à chaque entité également. Ainsi à l'entité rouge fut associée une bande de fréquences aiguës

tandis que les graves furent associés à l'entité bleue. La bande de fréquences associée à l'entité verte resta inchangée.

Par ailleurs, pour accentuer les retours visuels et sonores symbolisant la métamorphose du joueur (et donc la complétion du tableau), un objet auparavant intégré au décor fut modifié pour émaner du joueur au cours de sa transformation.

Dans cette configuration, le succès des différents joueurs fut beaucoup plus rapide et intuitif.

Au terme des tests, le jeu pourrait être amélioré sur bien des points, notamment pour perfectionner le script d'analyse sonore afin que l'approximation de F_0 soit beaucoup plus fine, ce qui permettrait de déterminer des objectifs plus difficiles à atteindre. Par ailleurs, la trame narrative pourrait être plus développée au cours de tableaux supplémentaires pour ainsi permettre de mettre au point de nouvelles interactions plus évoluées impliquant, par exemple d'autres personnages non-joueur ou un modèle physique permettant au joueur de se déplacer verticalement sans ascenseur.

Plusieurs participants au test, probablement plus habitués aux jeux vidéo que leurs pairs, ont émis l'idée de pouvoirs que le personnage pourrait gagner afin d'atteindre de nouveaux lieux ou rendre d'autres interactions vocales possibles. Si nous avions envisagé cette possibilité, elle requerrait, pour être réellement intéressante, de créer des lieux dont la conception architecturale serait méticuleusement conçue pour intégrer les nouvelles possibilités offertes par un pouvoir. Ces types d'organisation des niveaux et de contrôle de la progression du

joueur ne sont pas sans rappeler ceux des jeux de la série ***The Legend of Zelda***.

Et dans la mesure où 24 des 26 participants ont indiqué qu'ils auraient aimé que le jeu se poursuive à l'issue du dixième tableau, il pourrait être intéressant de donner une seconde vie à ***v0x*** et de pousser son concept dans ses limites pour créer un véritable jeu, plus long, plus sophistiqué et moins austère.

La réalisation de ***v0x*** ainsi que la collecte des retours des différents participants aux tests nous ont permis de conclure d'une part qu'il était possible de créer une **écriture vocale bi-directionnelle,** sans pour autant avoir besoin de technologies de détection de la parole, et d'autre part que le *gameplay* ainsi élaboré était d'autant plus agréable qu'il était curieux et inhabituel.

En effet, un des retours les plus positifs sur le jeu provint de deux amis proches de l'expérimentateur qui ne purent participer aux tests mais jouèrent à ***v0x*** sur leurs ordinateurs respectifs en commentant leur expérience en direct via un logiciel de visioconférence. Ces deux joueurs assidus et habitués à jouer à des jeux aux *gameplays* très variés firent part de leur étonnement : ils n'avaient jusqu'alors jamais joué à un jeu de la sorte.

Le plaisir de jeu procuré par ***v0x***, intimement lié à son *gameplay* et plus particulièrement à son **écriture vocale**, nous a convaincu de la pertinence de telles expérimentations créatives.

CONCLUSION

« Parole, cris, soupirs ou chuchotements, la voix hiérarchise tout autour d'elle et, de la même façon que la mère s'éveille quand les pleurs lointains de son enfant

dérangent le bruit – souvent plus intense - de la nuit, c'est, dans le torrent des sons, d'abord vers cet autre nous-même qu'est la voix d'un autre que se dirige notre attention. »

Michel Chion – La Voix au Cinéma, 1993, p. 19.

En concluant ce mémoire par cette citation, nous espérons rappeler au lecteur la trajectoire de la voix au cours de l'histoire des jeux vidéo.

D'un élément sonore superflu, balbutiant et maladroit, elle est devenue incontournable. Qu'elle accompagne le joueur dans sa partie, structure une narration dont elle devient la manifestation sonore ou donne toujours plus de profondeur à des univers virtuels dont la raison d'être n'est autre qu'égaler notre monde en terme de cohérence, de crédibilité et surtout de beauté, la voix a permis au jeu vidéo de se distancier de la littérature, de se rapprocher du cinéma, mais aussi d'inventer des formes et des motifs qui lui sont propres.

Comme l'indique Michel Chion, la voix, sans pour autant être le support d'un langage, nous ramène irrémédiablement à notre humanité. Quand bien même elle serait l'émanation désincarnée d'un robot à l'agonie, elle octroierait à cet être de métal et de circuits une étincelle de vie. Les joueurs de ***Halo 4*** qui soutiendraient le contraire seraient d'une mauvaise foi certaine.

Enfin, si le jeu vidéo peine encore aujourd'hui à tendre l'oreille vers le joueur, les tentatives récentes d'inclusion de la voix de ce dernier au cœur même du ***gameplay*** - dont ***v0x***, réalisé dans le cadre de ce mémoire, fait partie – laissent présager un futur

animé par des vocalités aussi nombreuses que variées.

Les travaux menés au cours de la rédaction de ce mémoire ainsi que la réalisation de ses parties pratique et expérimentale constituent un support de réflexion fertile, aussi bien pour ceux que le jeu vidéo passionne, que pour ceux dont la conception ou l'étude de mondes virtuels sont l'occupation principale, voire pour les profanes les plus curieux.

Plus particulièrement, les résultats des différentes séries de tests menées au cours de ce mémoire montrent que les technologies d'analyse spectrale de la voix du joueur sont plus faciles à appréhender et à mettre en œuvre que les technologie de reconnaissance de la parole. Ces dernières, plus complexes et coûteuses sont aussi moins universelles, du fait de leur dépendance au langage. ***v0x*** met à jour la possibilité de créer des *gameplays* aussi ludiques qu'interactifs. L'implication du joueur par sa voix, sans prérequis linguistiques, ouvre la porte à de nombreuses nouvelles expérimentations qui, nous l'espérons, aboutiront à la création de jeux au *gameplay* riche, innovant et agréable pour le plus grand nombre.

Bien entendu, de nombreux questionnements restent sans réponse. La démocratisation prochaine de la réalité virtuelle va-t-elle bouleverser une fois de plus la manière de penser et de concevoir les vocalités vidéo-ludiques ? Le perfectionnement perpétuel des technologies de reconnaissance vocale va-t-il permettre leur application sans contrepartie au jeu vidéo ? Le développement d'intelligences artificielles toujours plus évoluées va-t-il aboutir à l'établissement de dialogues entre le joueur et son jeu ?

De plus, les travaux de formalisation théorique de la grammaire vocale vidéo-

ludique de ce mémoire pourraient être approfondis. À l'heure où de plus en plus de personnes se mettent à jouer à des jeux d'une diversité effarante, à l'heure où des voix s'élèvent contre la linéarité des modes narratifs vidéo-ludiques dominants, nous comptons bien poursuivre notre observation de l'évolution du jeu vidéo, du mutisme au dialogue.

Plus particulièrement, nous souhaiterions nous focaliser sur les jeux de rôle pour étudier l'évolution de la **grammaire vocale** de ce genre aux narrations tentaculaires. Nous fonderions notre étude sur l'analyse du passage d'une narration textuelle à une narration vocale entre le neuvième et le dixième épisode d'une des séries de référence du jeu de rôle japonais, ***Final Fantasy.*** Nous pourrions également étudier l'influence de l'apparition de jeux en monde ouvert (en anglais *open world*) et de jeux en ligne présentant des univers persistants sur l'**écriture vocale** de ces jeux.

Bibliographie indicative

Sur le jeu vidéo :

Laurent Trémel, ***Jeux de rôles, jeux vidéo, multimédia : les faiseurs de monde***, Éditions PUF, Paris, 2001.

Karmen Franinovic et Stefania Serafn, ***Sonic Interaction Design***, MIT Press, Cambridge, 2013.

Karen Collins, ***Game Sound***, MIT Press, Cambridge, 2008.

Tony Fortin, Philippe Mora et Laurent Trémel, ***Les jeux vidéo : pratiques, contenus et enjeux sociaux***, Éditions l'Harmattan, collection Champs Visuels, Paris,2005.

Sylvie Craipeau, Sébastien Genvo et Brigitte Simonnot, ***Les jeux vidéo au croisement du social, de l'art et de la culture***, Question de communication, série actes, n° 8, Nancy, 2010.

Matthieu Triclot, ***Philosophie des jeux vidéo,*** Éditions La Découverte, Collection Zones, Paris, 2011.

Richard Stevens et Dave Raybould , ***The Game Audio Tutorial***, Focal Press, 2011.

Auteurs multiples, ***Level Up***, Niveau 1, Third Editions, Paris, 2015.

Auteurs multiples, ***JV***, périodique mensuel, Wildfire Media, Paris, 2015.

Sur la voix :

Bruno Bossis, ***La voix et la machine : la vocalité artifcielle dans la musique contemporaine***, Éditions PUF, Paris, 2005.

Michel Chion, ***La voix au cinéma***, Cahiers du Cinéma, Collection Essais, Paris, 1982 (réédition de 1993).

Bruno Bossis, Marie-Noëlle Masson et Jean-Paul Olive, ***Le modèle vocal : la musique, la voix la langue***, Presses Universitaires de Rennes, Rennes, 2007.

CALLIOPE, ***La parole et son traitement automatique,*** Masson, Collection technique et scientifque des télécommunications, Paris, 1989.

Philippe Martin, ***Phonétique acoustique : Introduction à l'analyse acoustique de la parole***, Armand Colin, collection Cursus, Paris, 2008.

Gérard Pelé, ***Études sur la perception auditive***, Éditions L'Harmattan, collection Arts et Sciences de l'Art, Paris, 2012.

Roland Barthes, ***Le grain de la Voix****, dans* **Image, Music, Text,** *F*ontana Press, traduction de Stephen Heath, Londres,1977.

Sous la direction de Norie Neumark, Ross Gibson et Theo Van Leeuwen, ***VØ1CE : Vocal Aesthetics in Digital Arts and Media***, MIT Press, Cambridge, 2010.

Sur des sujets proches :

Lev Manovich, ***Le langage des nouveaux médias***, Les presses du réel, Collection Perceptions, Saint Étienne, 2010.

Jacques Rancière, ***Le spectateur émancipé***, Éditions La Fabrique, Paris, 2008.

Filmographie indicative

Spike Jonze, ***Her***, 2014, États-Unis, couleur.

Stanley Kubrick, ***2001 : A Space Odyssey***, 1968, États-Unis, couleur.

Seth Gordon, ***King of Kong***, 2007, États-Unis, couleur.

Lisanne Pajot, ***Indie Game : The Movie***, 2011, Canada, couleur.

Chris Marker, ***Level 5***, 1996, France, couleur.

Jean-Luc Godard, ***Histoire(s) du cinéma***, de 1988 à 1998, Suisse, couleur.

Kenji Mizoguchi, ***L'intendant Sanchô***, 1954, Japon, noir et blanc.

Marguerite Duras, ***L'Homme Atlantique***, 1981, France, couleur.

Ludothèque indicative :

Halo – Série de jeux développée par Bungie et 343 Industries et éditée par Microsoft Games – 2001.

The Elder Scrolls – Série de jeux développée et éditée par Bethesda – 1994.

Tom Clancy's Endwar – Developpé et édité par Ubisoft – 2008.

Fable – Série de jeux développée par Lionhead Studios et éditée par Microsoft Games – 2004.

Final Fantasy – Série de jeux développée et éditée par Square Enix – 1987.

FEZ – Développé et publié par Polytron – 2013.

Journey – Développé par Thatgamecompany et édité par Sony – 2012.

Portal 1 et Portal 2 – Édités et développés par Valve – 2007 et 2011.

Amnesia : the Dark Descent – Édité et développé par Frictional Games – 2010.

Mass Efect – Trilogie développée par Bioware et éditée par Electronic Arts – 2008.

Dead Space – Trilogie développée par Visceral Games et éditée par Electronic Arts – 2008.

Diablo 3 – Édité et développé par Blizzard – 2012.

The Legend of Zelda – Série de jeux éditée et développée par Nintendo – 1986.

Bioshock – Trilogie développée par Irrational Games et éditée par 2K Games – 2007.

Mortal Kombat – Développé par NetherRealm Studios et Midway et édité par Midway – 1992.

The Stanley Parable - Développé par Galactic Cafe et édité par Valve – 2013.

Dark Souls - Série de jeux développée par From Software et éditée par Namco Bandai - 2011.

Thomas Was Alone - Développé et édité par Mike Bithell - 2012.

Kentucky Route Zero - Développé par Cardboard Computer et édité par Valve - 2013.

Borderlands - Série de jeux développée par Gearbox Software et éditée par 2K Games à partir de 2009.

Destiny - Développé par Bungie et édité par Activision en 2014.

Call of Duty : Advanced Warfare - Développé par Sledgehammer Games et édité par Activision en 2014.

Dishonored - Développé par Arkane Studios et édité par Bethesda – 2012.

Metal Gear - Série de jeux développée et éditée par Konami – 1987.

Papers, Please - Développé par Lucas Pope et édité par Valve – 2013.

Sound Shapes - Développé par Queasy Games et édité par Sony Computer Entertainment – 2012.

Hohokum - Développé par Honeyslug et édité par Sony Computer Entertainment – 2014.

Patapon - Série de jeux développée et éditée par Sony Computer Entertainment – 2007.

Batman Arkham – Série de jeux développée par Rocksteady Studios et éditée par Warner Games – 2009.

Bloodborne – Jeu développé par From Software et édité par Sony Computer Entertainment – 2015.

Oddworld : Abe's Odyssey – Jeu développé par Oddworld Inhabitants et édité par GT Interactive – 1997.

Myst – Série de jeux développée par Cyan et éditée par Brøderbund – 1993.

Doom – Série de jeux éditée et développée par id Software – 1993.

Mémoires rédigés au sein de l'École Nationale Supérieure Louis Lumière :

L'interaction sonore dans le jeu vidéo – Nicolas Fournier, Spécialité Son 2011.

Le rôle du son dans l'apparition de la peur dans le jeu vidéo – Baptiste Palacin, Spécialité Son 2013.

Annexes

Table des annexes

COMPTE-RENDU DE LA PARTIE PRATIQUE

Cette partie des annexes a pour objectif de documenter la réalisation de v0x. Pour commencer, voici une description des différents tableaux qui composent le jeu.

Écran-Titre :

Lorsque le joueur appuie sur la touche « X » du clavier de l'ordinateur de jeu, l'écran titre s'efface pour laisser place à une vidéo d'introduction.

Intertitre :

Cet écran annonce que la vidéo à venir est à interpréter comme une communication d'un autre personnage. Le son du codec de ***Metal Gear Solid*** complète l'écran en renvoyant au système de communication de la série de jeux de Konami.

Vidéo d'introduction :

Cette vidéo, réalisée grâce à la librairie GEM de Pure Data, montre un homme au visage masqué et à la voix métallique. Le joueur devine, grâce aux instructions de l'homme, que celui-ci est son supérieur

hiérarchique. L'homme à la capuche indique également l'objectif principal du jeu.

Les éléments sonores, et notamment la piste vocale, a été éditée, comme tous les éléments sonores du jeu, grâce au logiciel Pro Tools.

Tableau 1 :

Ce tableau présente au joueur son personnage. Il n'y a pas d'interaction vocale dans ce tableau. Cependant, le joueur peut d'ores et déjà y comprendre, en observant la bouche du personnage, que sa voix est prise en compte par le jeu, grâce aux particules qui s'en échappent.

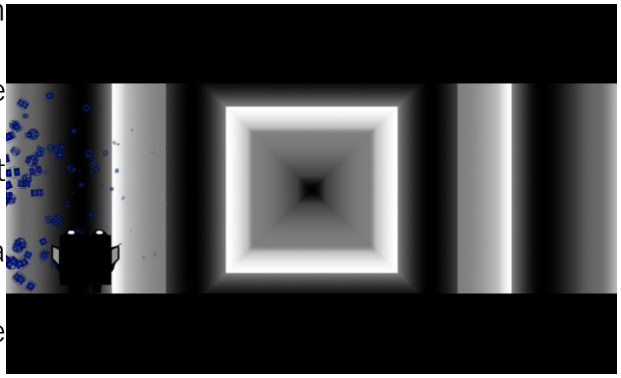


Tableau 2 :

La première interaction vocale se trouve dans ce tableau. En son centre, un microphone indique où le joueur doit se placer. Par la suite, la porte monumentale, au niveau du bord droit de l'écran, s'ouvre si le joueur émet un son dont l'intensité est suffisamment élevée. Ici, c'est la fonction NearRangeLoudness du script AudioInputEvent qui est utilisée pour contrôler l'ouverture de la porte. Par ailleurs, on remarque que le motif graphique sur la porte scintille d'un bleu très clair lorsqu'elle se met en mouvement. Ce code couleur sera associé à tous les objets contrôlés par l'intensité de la voix du joueur.

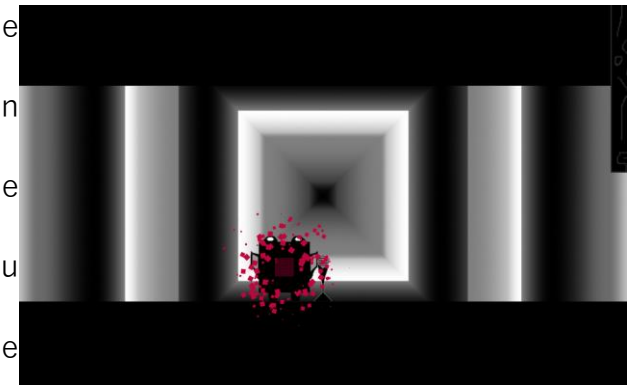
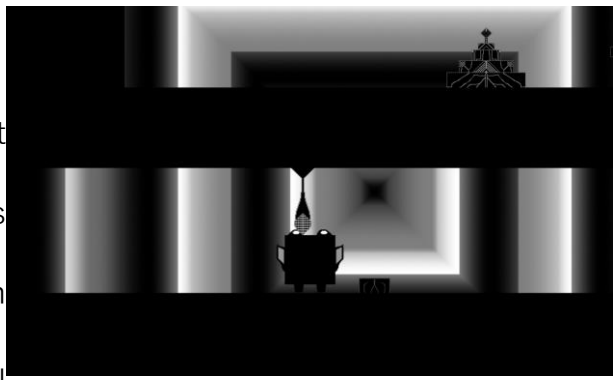
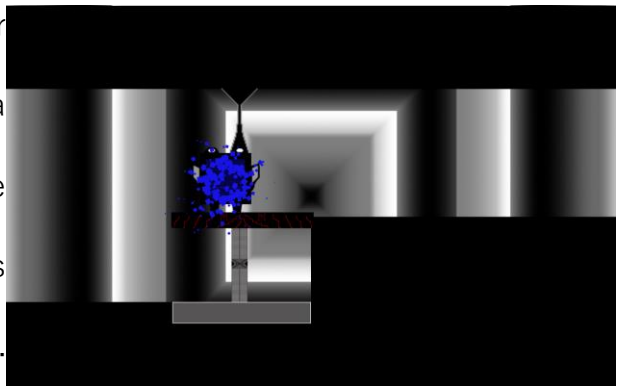


Tableau 3 :



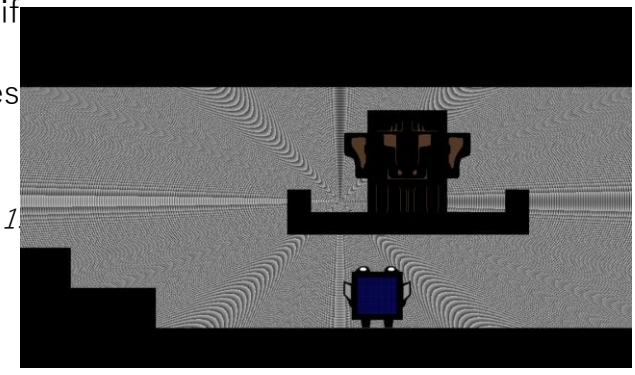
Ici, les fonctions OnIncLoudness et OnDecLoudness contrôlent les mouvements ascendants et descendants de la porte en fonction de l'intensité sonore de la voix du joueur. Pour faire s'élever la porte, le joueur doit émettre un son de plus en plus fort, image après image. La porte présente le même scintillement bleu clair que la porte précédente, pour guider le joueur.

Tableau 4 :



Ici, le joueur est confronté au premier ascenseur du jeu. Cet ascenseur détecte si la valeur approchée de la fréquence fondamentale de la voix du joueur est plus élevée à l'instant *t* qu'à l'instant *t – 1 image*. Si tel est le cas, un script attaché à l'ascenseur incrémente une unité de déplacement vers le haut. Les fonctions OnIncFrequency et OnDecFrequency sont celles qui permettent le contrôle du mouvement de l'ascenseur.

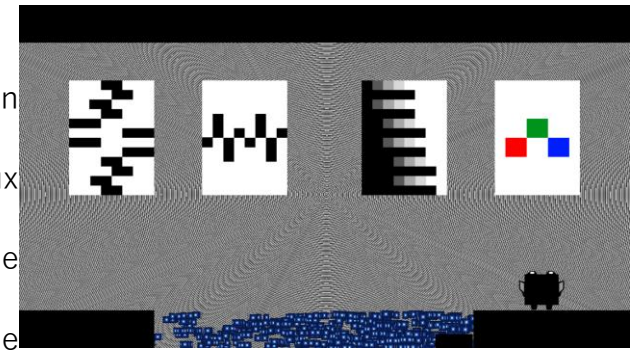
Tableau 5 :



Ce tableau est essentiellement narratif et permet au joueur de comprendre les

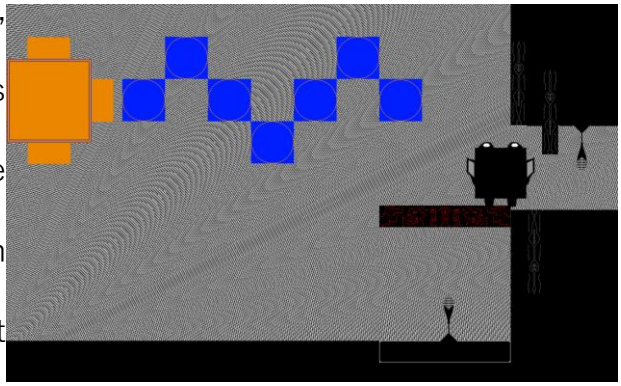
objectifs suivants. Ceux-ci lui sont donnés par l'immense tête qui l'accueille et limite ses déplacements pendant son monologue. Avant d'autoriser le joueur à poursuivre, il lui donne un objet qui réapparaîtra au tableau 10.

Tableau 6 :



Ce tableau est essentiellement un tableau d'ambiance. Il présente les tableaux à venir et est un point de non retour que le joueur doit atteindre d'une traite, sans se retourner. En effet, les blocs qui composent le pont tombent dans le vide une fois que le joueur entre en contact avec eux. Si le joueur tombe également dans le vide, il est renvoyé au début du tableau.

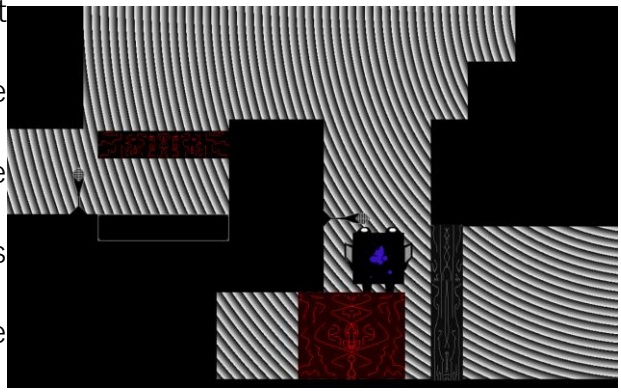
Tableau 7 :



Pour la première fois dans ce tableau, le joueur est confronté aussi bien à des objets contrôlés par l'intensité de sa voix que par la fréquence fondamentale de celle-ci. En effet, l'ascenseur à droite de l'écran est contrôlé par la fonction FrequencyToRatio du script AudioInputEvent, qui fait directement correspondre la hauteur de l'ascenseur avec la valeur approchée de la fréquence fondamentale de la voix du joueur. Plus cette fréquence est aigüe, plus l'ascenseur approchera du haut de l'écran, et vice versa. Les trois blocs de pierre, eux,

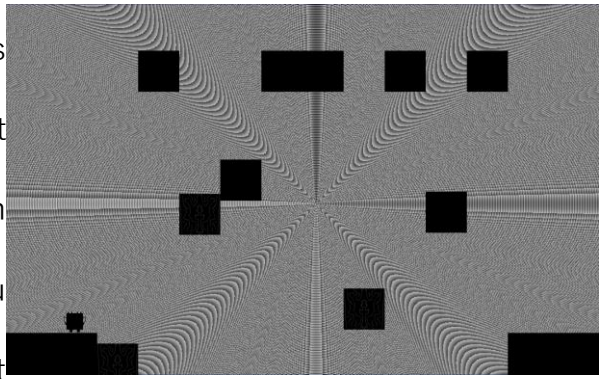
s'ouvrent grâce à la fonction OnIncLoudness, tout comme la porte du tableau 3.

Tableau 8 :



Ici, trois éléments interactifs doivent être contrôlés par la voix du joueur. Le premier ascenseur est contrôlé par la même fonction que l'ascenseur du tableau 4, mais les seuils de détection et la valeur de l'incrément de mouvement vertical ont été modifiées pour que cet ascenseur soit plus dur à mettre en mouvement. Une fois tombé dans le goulet qui suit l'ascenseur, le joueur doit encastrer le bloc massif dans le mur à sa gauche puis faire s'enfoncer un bloc dans le sol. Comme l'indique les différents codes couleurs, l'ascenseur et le bloc massif sont contrôlés par la fréquence fondamentale de la voix du joueur tandis que le dernier bloc s'abaisse lorsque le joueur émet des sons de plus en plus puissants.

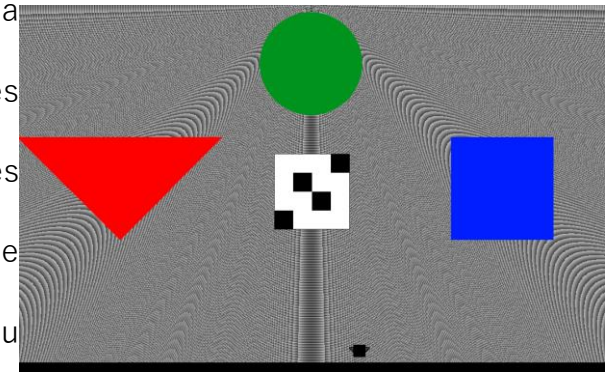
Tableau 9 :



Dans ce tableau, chacun des dix blocs qui constituent le pont que le joueur doit traverser est contrôlé par la fonction FrequencyToRatio, à l'instar de l'ascenseur du tableau 7. Cependant, à chaque bloc est associée une plage de fréquences différente, si bien que le joueur doit tâtonner pour

aligner les blocs dans le bas de l'écran (cela se produit si un son très aigu est émis) ou se jeter d'un bloc à l'autre jusqu'à atteindre le bord droit de l'écran.

Tableau 10 :



Le joueur doit émettre des sons dont la fréquence fondamentale est comprise entre les deux bornes de trois bandes de fréquences différentes. À chaque image, le programme génère un jeton si la fréquence de la voix du joueur est comprise dans la bande de fréquences adaptée. Une fois qu'un nombre suffisant de jetons à été généré, le programme passe à la bande de fréquences suivante.

QUESTIONNAIRE POST-TEST

Première partie : Profil de l'utilisateur

Nom :

Âge :

Sexe :

Activité actuelle :

Deuxième partie : Familiarité de l'utilisateur avec les jeux vidéo

À quelle fréquence jouez-vous à un jeu vidéo ? (cocher le cercle correspondant)

O une à plusieurs fois par jour O une à plusieurs fois par semaine

O une à plusieurs fois par mois O une à plusieurs fois par an

O moins d'une fois par an O jamais

A quels types de jeux jouez-vous ? (cocher le(s) cercle(s) correspondant(s))

O Plateforme (*Mario*, *Sonic*) O Sports (*FIFA*, *NBA*)

O Tir (*Call of Duty*, *Battlefield*) O Combat (*Tekken*, *Street Fighter*)

O Rôles (*Final Fantasy*, *Skyrim, Diablo*) O Stratégie (*Starcraft*, *Age of Empires*)

O Sandbox (*Minecraft, GTA*) O Mobile (*Angry Birds*, *Candy Crush*)

O Infiltration (*Metal Gear Solid*) O Musique (*Guitar Hero*, *Singstar*)

O *Motion Gaming* (*Wii Sports*) O Point & Click (*Myst*, *The Walking Dead*)

O Horreur (*Resident Evil*, *Dead Space*) O MMO & MOBA (WoW, LoL)

O Beat'em all (Devil May Cry, God of War) O Shoot'em up (Ikaruga, R-Type)

O Autres - préciser :

Sur quel(s) support(s) jouez-vous à des jeux vidéo ?

O Console de salon O Ordinateur personnel O Tablette

O Console portable O Smartphone

Que recherchez-vous en jouant à un jeu vidéo ?

(Par exemple : une histoire longue et riche, des sensations de jeu particulières, de beaux graphismes...)

Seriez-vous susceptible d'être attiré(e) par un jeu vidéo présentant une interface de jeu inhabituelle ?

(Par exemple : détection de mouvement, manette spéciale, reconnaissance vocale...)

O Oui O Non

Pourquoi ?

Dans quel type d'environnement jouez vous à des jeux vidéo ?

Comment jouez-vous aux jeux vidéo ?

O Sans le son O Sur enceintes

O Avec des écouteurs ou un casque

Troisième Partie : Au sujet de *v0x*

A.Expérience globale

Avez-vous apprécié ce jeu ?

O Oui O Non

Pourquoi ?

B.Ergonomie du jeu

La prise en main du jeu :

O Pas du tout intuitive

O Plutôt pas intuitive

O Ni intuitive, ni contre-intuitive

O Plutôt intuitive

O Tout à fait intuitive

Vous êtes vous senti(e) à l'aise en jouant grâce à votre voix ?

O Très mal à l'aise

O Plutôt mal à l'aise

O Indifférent

O Plutôt à l'aise

O Tout à fait à l'aise

Les éléments visuels vous ont-ils aidé à comprendre que votre voix était prise en compte par le jeu ?

O Oui O Non

Quels éléments de votre voix étaient pris en compte par le jeu selon vous ?

O Hauteur (fréquence) O Volume sonore O Mot prononcé O Type de son émis

O Autre (préciser) :

Comment améliorerez-vous l'ergonomie du jeu pour qu'elle soit plus intuitive ou plus agréable ?

B.Expérience tableau par tableau

Introduction.

Les paroles étaient-elles suffisamment intelligibles selon vous ?

O Oui O Non

Tableau 1 : Découverte du personnage.

Avez-vous compris immédiatement que le carré noir sur pattes était votre personnage ?

O Oui O Non

Tableau 2 : Premier microphone.

Avez-vous identifié l'objet au centre de l'écran comme étant un microphone ?

O Oui O Non

Tableau 3 : Un bloc s'enfonce dans le sol.

Avez-vous compris qu'une fois à proximité du microphone, vous étiez en mesure d'interagir avec les objets du décor grâce à votre voix ?

O Oui O Non

Tableau 4 : Un ascenseur.

Parmi les paramètres de votre voix que vous avez cochés précédemment, lequel d'entre eux permettait, selon vous, de faire monter et descendre l'ascenseur ?

Tableau 5 : Rencontre avec le Passeur.

Les paroles étaient-elles suffisamment intelligibles selon vous ?

O Oui O Non

Tableau 6 : La voie du Conseil.

Tableau 7 : Les mâchoires de pierre.

Notez la difficulté de ce tableau (0 pour un tableau très simple, 10 pour un tableau impossible) : /10

Tableau 8 : Sous un ciel de pixels.

Notez la difficulté de ce tableau : /10

Tableau 9 : La salle du temps.

Notez la difficulté de ce tableau : /10

Tableau 10 : Le Conseil.

Notez la difficulté de ce tableau : /10

Les paroles étaient-elles suffisamment intelligibles selon vous ?

O Oui O Non

Les indices visuels et sonores vous ont-ils aidé à comprendre la solution de ce tableau ?

O Oui O Non

Auriez-vous aimé que le jeu continue ?

O Oui O Non

Quatrième partie

Avez-vous des remarques ou des suggestions au sujet de ce jeu ?

Les images tournées au cours du test ne seront pas diffusées hors de l'école.

Accepteriez-vous que des images de votre passage soient diffusées au cours de la soutenance du mémoire de master dont ***v0x*** constitue la partie pratique ?

O Oui O Non

Merci beaucoup pour votre participation !

BILAN DES TESTS DE ***V0X***

Participant	Sexe	Age	Jeu apprécié	Intuitivité	Aisance de jeu	Indices visuels pertinents	Microphone identifié comme tel	Jour de passage	Indices du tableau 10 pertinents	Désir de suite
Florine	F	21	1	3	3	1	1	07/05/15	0	1
Celsian	M	21	1	3	1	1	1	06/05/15	1	1
Baptiste	M	21	0	2	1	1	1	06/05/15	0	0
Adrien L	M	22	1	1	4	0	0	06/05/15	1	1
Léo	M	22	1	3	3	1	1	05/05/15	0	1
Eléonore	F	22	1	3	3	1	1	05/05/15	1	1
Léa	F	22	1	3	3	1	1	07/05/15	NSP	1
Clément	M	22	1	2	3	1	1	07/05/15	0	1
Cédric	M	22	1	3	2	1	1	07/05/15	0	1
Adrien S	M	23	1	3	3	1	1	07/05/15	1	1
Chloé	F	23	1	3	1	1	1	07/05/15	1	1
Thomas	M	23	1	3	2	1	1	05/05/15	0	1
Simon	M	23	1	3	3	1	1	06/05/15	0	1
Morgane	F	23	1	3	1	0	1	05/05/15	1	1
Jonas	M	23	1	3	1	0	1	05/05/15	1	1
Paul PDLG	M	24	1	0	1	1	1	06/05/15	0	1
Nicolas	M	24	1	4	1	1	1	06/05/15	1	0
Cyril	M	24	1	1	1	1	1	05/05/15	0	1
Raphaël	M	24	1	4	4	1	1	05/05/15	1	1
Frédéric	M	25	1	4	4	1	1	06/05/15	0	1
Elia	M	25	1	3	1	1	1	06/05/15	0	1
Marie-Angélique	F	25	1	3	3	1	1	05/05/15	1	1
Athalia	F	27	1	1	2	1	1	05/05/15	0	1
Paul Pr	M	28	1	3	1	1	1	05/05/15	1	1
Franck	M	39	1	3	3	1	1	05/05/15	0	1
Roselmy	M	42	1	3	3	0	1	07/05/15	1	1
Moyenne		24,615	96,15%	2,731	2,231	84,62%	96,15%			92,31%
Moyenne pour le 05/05									54,55%	
Moyenne pour le 06/05									37,50%	
Moyenne pour le 07/05									50,00%	

Le tableau ci-dessus reprend les réponses au questionnaire des différents participants.

Par ailleurs, voici les différentes modifications qui furent effectuées :

- *Du 5 mai au 6 mai :*

Différenciation des codes de couleurs : plutôt que de signaler tous les objets interactifs par des motifs rouges, les objets contrôlés par la fréquence de la voix du joueur conservent cette couleur tandis que les objets contrôlés par l'intensité de la voix du joueur sont indiqués par des motifs bleu clair.

Des murs invisibles sont créés ou les murs invisibles existants sont déplacés pour

éviter que le joueur puisse sortir d'un tableau par la gauche ou qu'il se retrouve bloqué entre deux parois.

Quitter l'ascenseur du tableau 7 par la gauche ne provoque plus un redémarrage du tableau mais réinitialise la position de l'ascenseur.

Dans le tableau 10, l'objet donné par le personnage du tableau 5 sort du personnage du joueur lors de la résolution de l'énigme.

Dans le tableau 10, les formes clignotent une fois lorsque le joueur a atteint le nombre de jetons correspondant au passage à la bande de fréquences suivante.

Dans le script correspondant, le type de fenêtre d'analyse est modifié de la fenêtre rectangulaire à la fenêtre de Blackman-Harris. Les différents scripts sont ajustés en conséquence pour que le gain de précision ne bouleverse pas la difficulté du jeu.

- *Du 6 mai au 7 mai :*

Sur le tableau 2, le seuil d'intensité sonore déclenchant, une fois franchi, l'ouverture de la porte est rehaussé pour augmenter la difficulté du tableau.

Sur le tableau 5, la zone de déclenchement du dialogue est déplacée pour éviter que le joueur ne se retrouve bloqué à l'issue du dialogue.

Sur le tableau 7, le microphone est déplacé de la gauche de l'ascenseur pour être placé sur l'ascenseur dans un souci de lisibilité du tableau.

Sur le tableau 10, les plages de fréquences sont modifiées en accord avec les suggestions évoquées dans le troisième chapitre de ce mémoire.

Nous remarquons que malgré leur évolution d'un jour à l'autre, les indices visuels et sonores guidant le joueur dans la résolution de l'énigme ne sont pas jugés plus pertinents par les participants. Cependant, à en juger par les importantes variations du nombre de participants ayant trouvé les indices plus pertinents, il faudrait multiplier le nombre de tests pour obtenir des résultats qui nous permettraient de conclure. En effet, en l'état actuel de l'étude, il semble que la réponse à cette question soit intimement lié au profil de joueur du participant. Pour autant, en comparant les différents profils des participants, il nous est impossible de relier des paramètres tels que le type de jeu le plus pratiqué ou la fréquence de jeu à la réponse du participant en question.

Il nous faudrait donc, en plus d'une multiplication des tests, repenser le questionnaire afin d'obtenir plus d'informations, ou en tout cas des informations plus pertinentes sur les participants.

Il est possible de jouer à ***v0x*** grâce aux fichiers se trouvant sur le CD joint à ce mémoire de master.

Le jeu a été développé sur Mac OSX 10.8.5 et a été réglé avec les microphones intégrés Apple. Même si le jeu a été installé et testé dans un environnement Windows, il est préconisé de l'installer sur un ordinateur dont le système d'exploitation est Mac OSX pour être au plus proche du jeu tel qu'il a été pensé, développé et présenté aux participants.

Pour lancer le jeu, il suffit de copier le fichier .zip correspondant à votre système d'exploitation du CD vers votre disque dur, puis de le décompresser pour obtenir un exécutable.

COMPTE-RENDU DE LA PARTIE EXPÉRIMENTALE

Le principal objectif des travaux menés dans le cadre de cette partie expérimentale était d'identifier les éléments vocaux susceptibles de rendre instable un *gameplay* basé sur l'analyse spectrale, et plus particulièrement la détection de la fréquence fondamentale de la voix du joueur.

Pour ce faire, des séries d'enregistrements ont été effectuées auprès de trois groupes de sujets appartenant à trois tranches d'âges différentes :

- 8 étudiants âgés de 22 à 23 ans.
- 9 élèves de terminale scientifique du lycée public Léonard de Vinci (Commune de Monistrol sur Loire) âgés de 17 à 18 ans ainsi que leurs deux professeurs, âgés respectivement de 40 et 47 ans.
- 11 élèves de la classe de CM2 de madame Valour, à l'école primaire publique Lucie Aubrac (Commune de Monistrol sur Loire) âgés de 9 à 11 ans.

Les enregistrements étaient effectués simultanément sur deux supports : l'ordinateur sur lequel ***v0x*** allait être présenté grâce au logiciel d'analyse phonétique Praat, développé par Paul Boersma et David Weenink, de l'université d'Amsterdam, et au microphone intégré à l'ordinateur ; un enregistreur portable ZOOM H4N associé à un microphone de mesure PRESONUS PRM1.

Les deux supports d'enregistrement étaient disposés de façon à ce que les capsules des microphones soient les plus proches possible et à 50 centimètres du sujet environ.

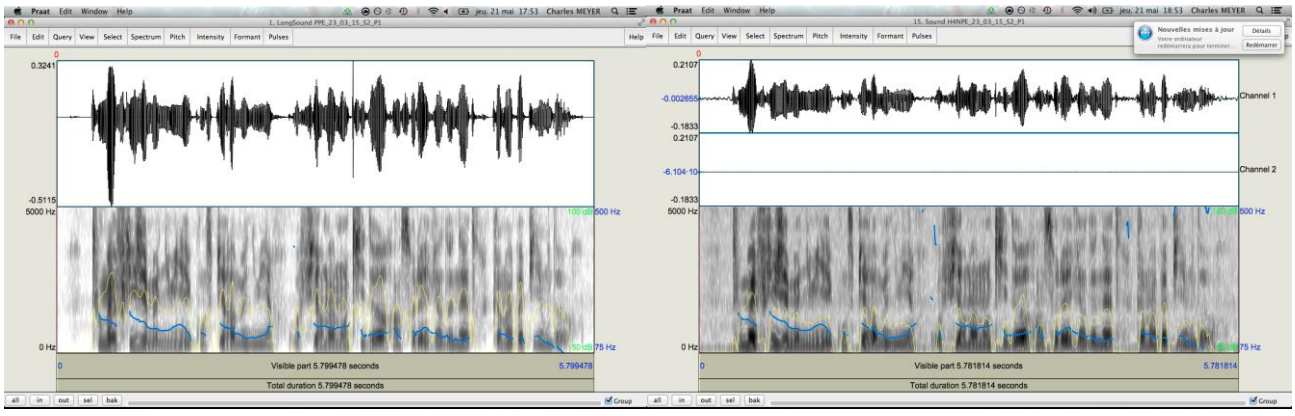
Les sujets du test devaient prononcer deux séries de phrases présentant différents phonèmes de la langue française :

- « Tout m'afflige et me nuit et conspire à me nuire. Au firmament qui dort, un soleil vient de naître, comme un papillon d'or. » : Cette série de phrases, riche en assonances avait pour but d'étudier les sons voyelle de la langue française afin

d'observer leurs manifestations spectrale et dynamique grâce aux outils d'analyse de Praat.

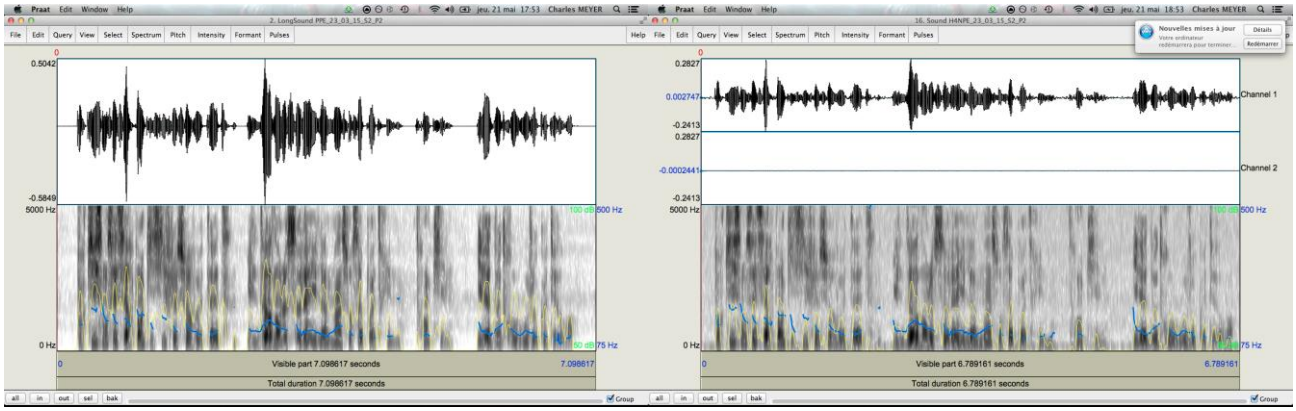
– « Pour qui sont ces serpents qui sifflent sur vos têtes ? Il dort dans le soleil, la main sur sa poitrine. Tranquille. Il a deux trous rouges au côté droit. » : Ce sont ici les sons consonnes de la langue française qui sont observés grâce à une série d'allitérations.

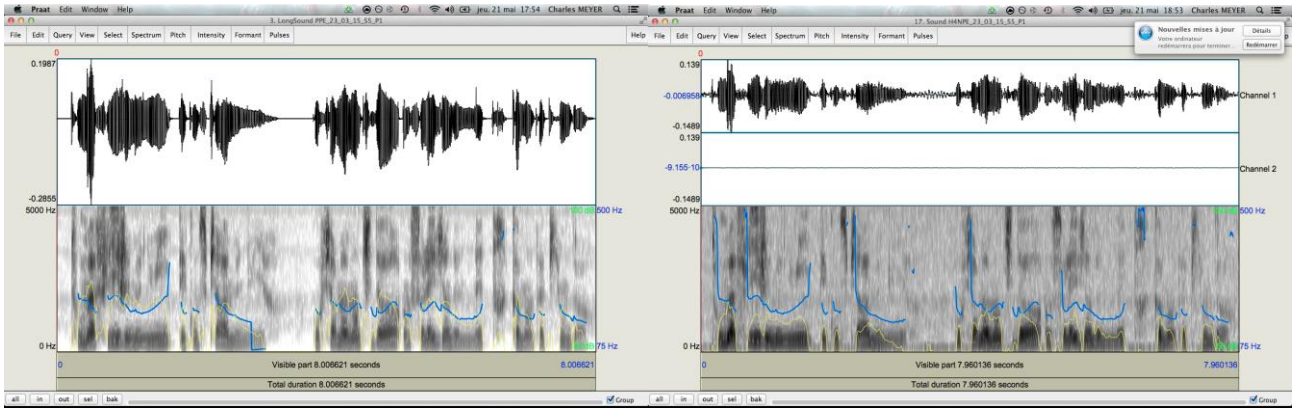
Voici les graphiques que nous a fourni Praat, pour chacune des séries de phrases, pour deux sujets de chaque tranche d'âge ainsi que pour un des professeurs de la classe de terminale. À gauche, les enregistrements effectués avec le microphone intégré. À droite, les enregistrements effectués avec le microphone Presonus PRM1. LA présentation est identique à celle qui a été adoptée dans le mémoire : le graphique bleu décrit l'évolution de F_0, la ligne jaune décrit l'évolution de l'intensité sonore de la

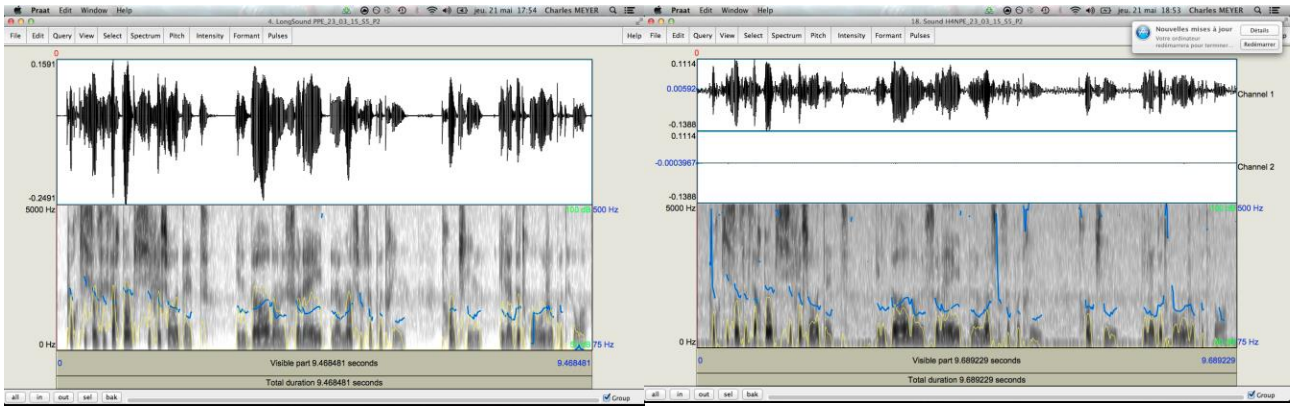


voix du sujet.

1. Première série de phrases, sujet masculin âgé de 25 ans.



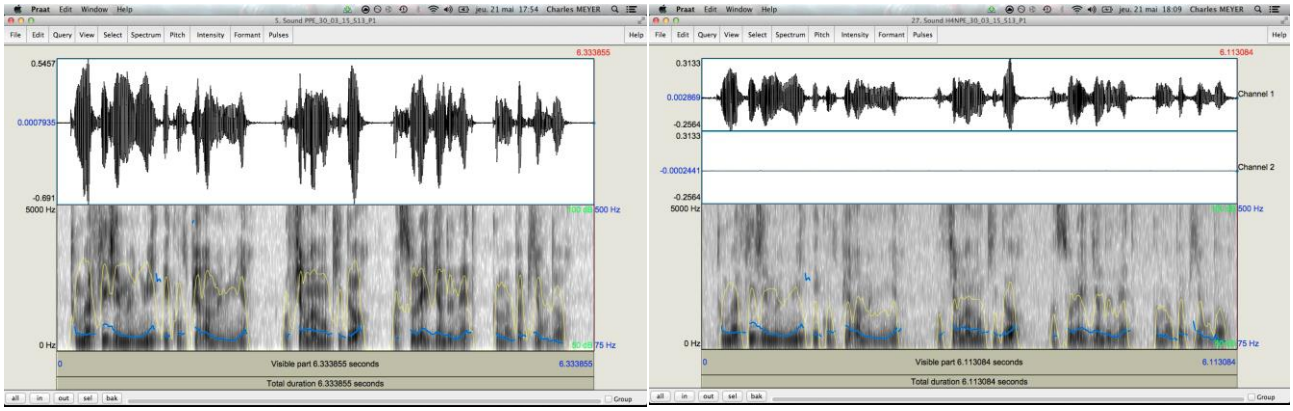




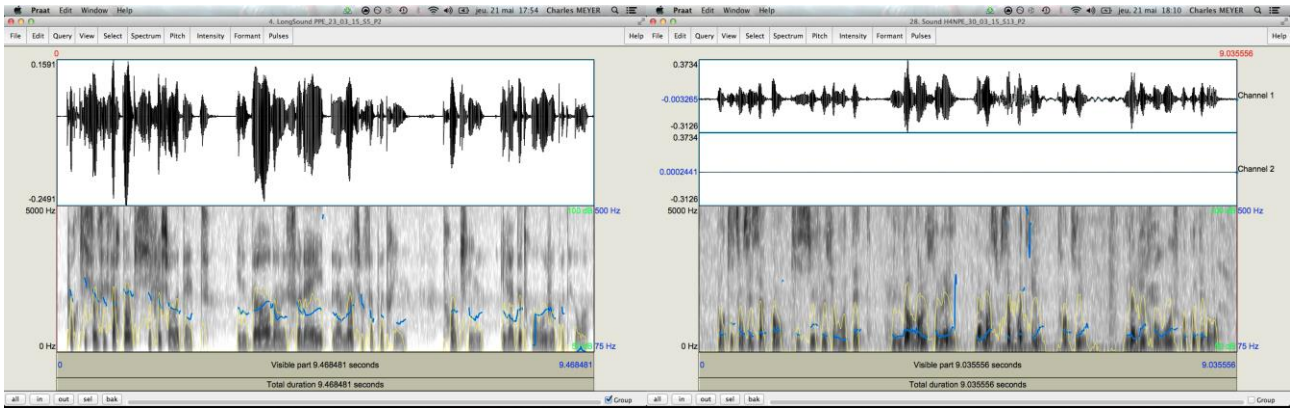
2.Deuxième série de phrases, même sujet.

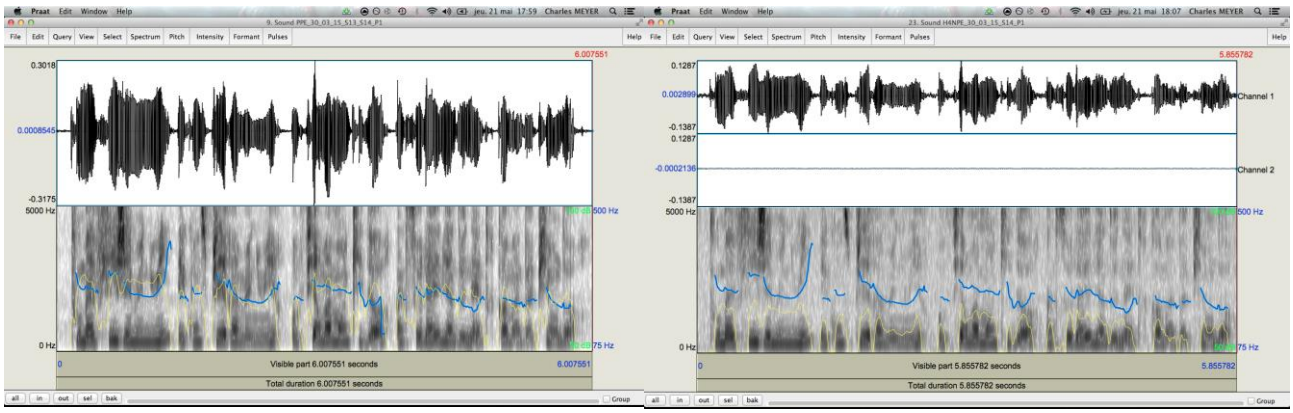
3. Première série de phrases, sujet féminin âgé de 23 ans.

4. Deuxième série de phrases, même sujet.



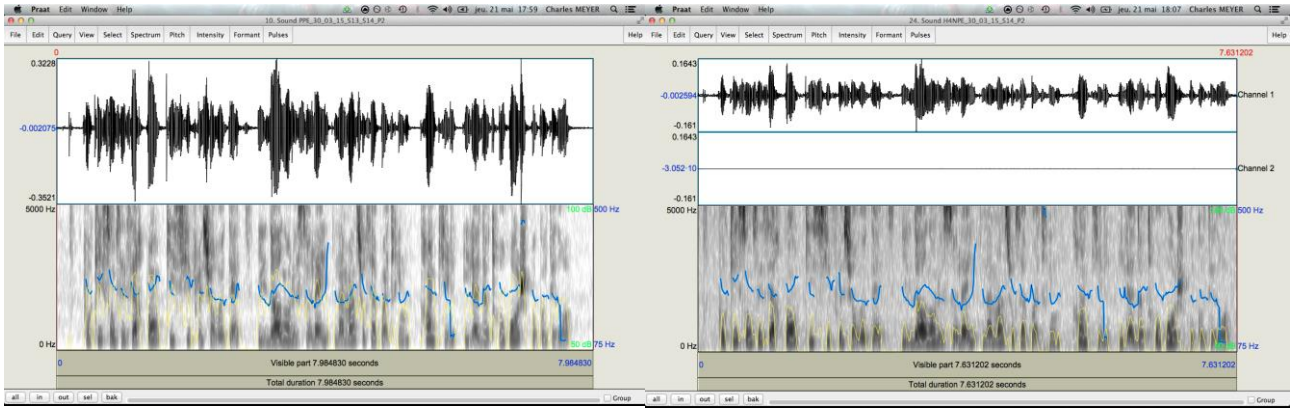
5. Première série de phrases, sujet masculin âgé de 18 ans.



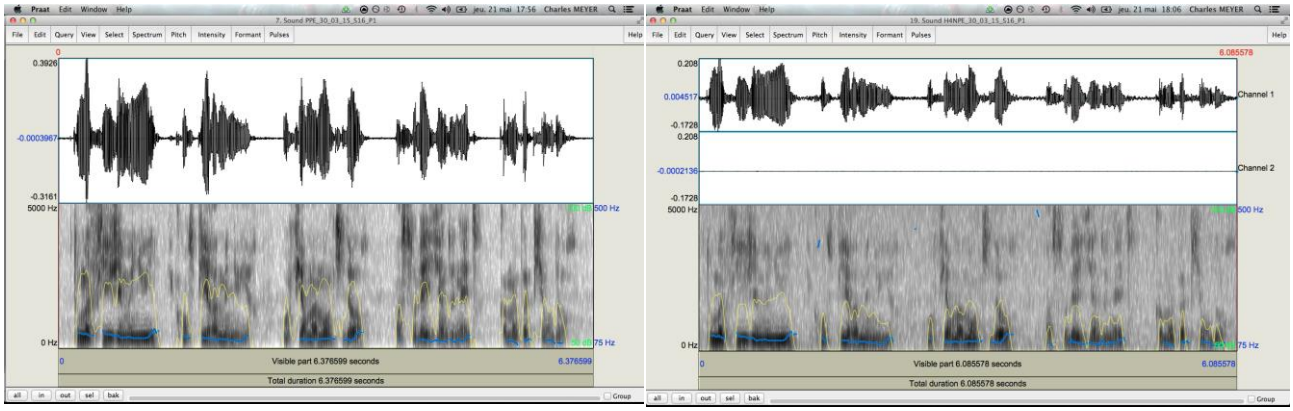


6. Deuxième série de phrases, même sujet.

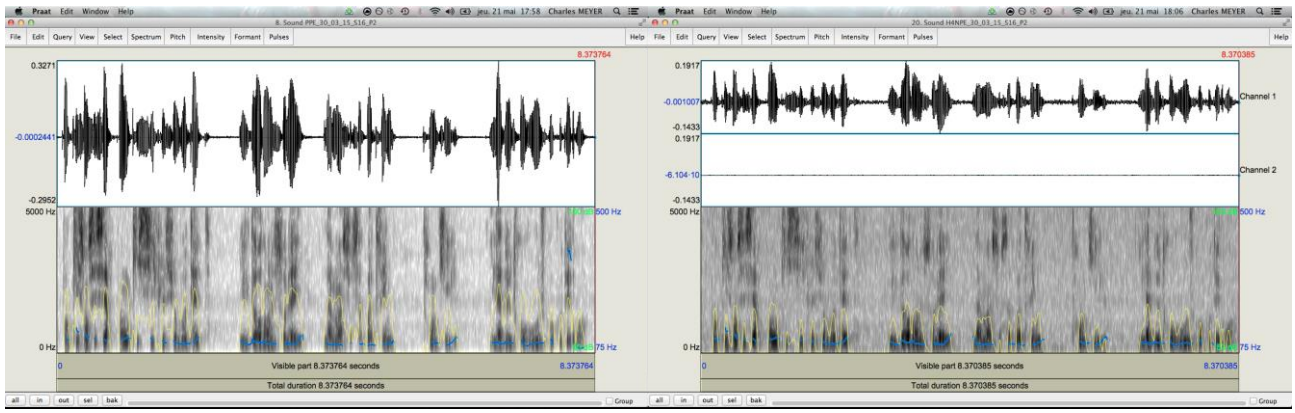
7. Première série de phrases, sujet féminin âgé de 18 ans.



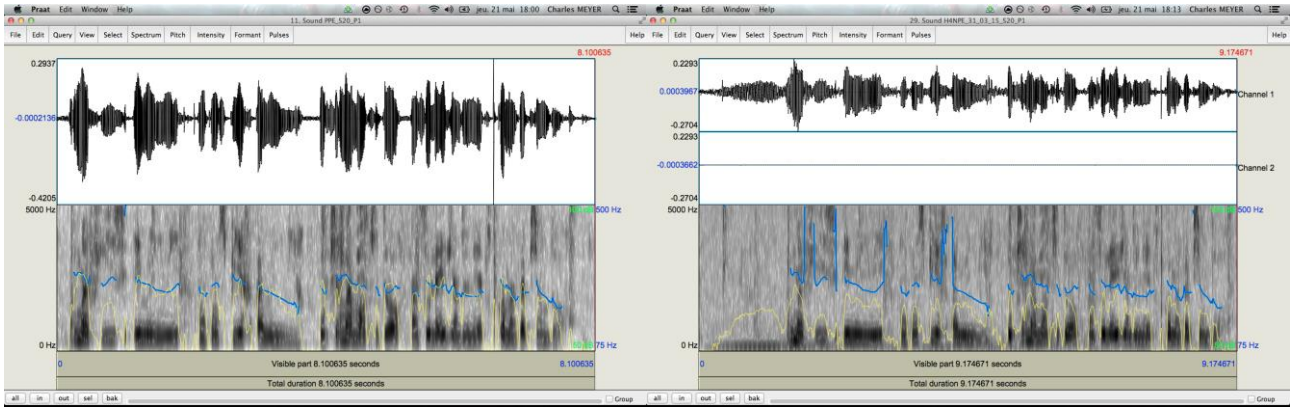
8. Deuxième série de phrases, même sujet.

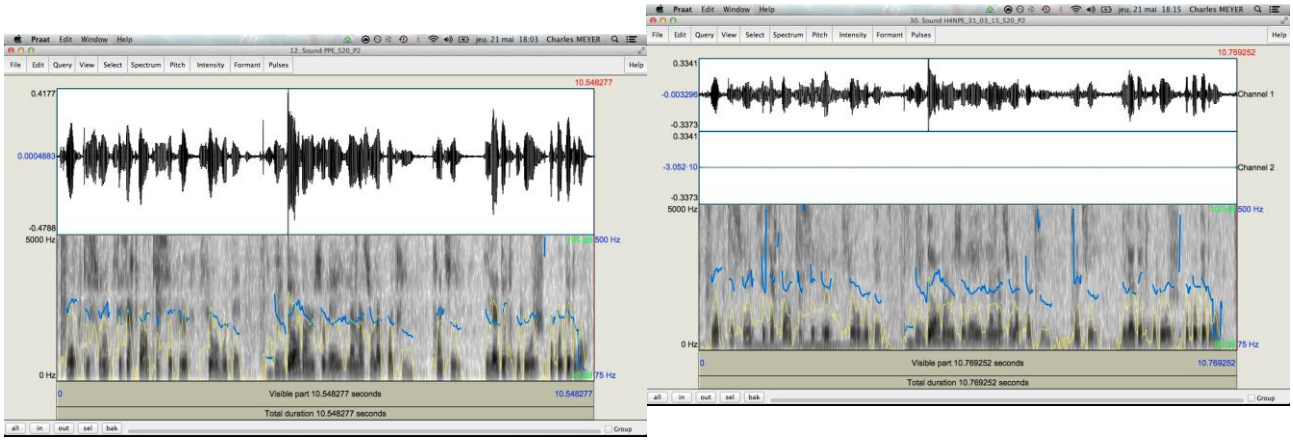


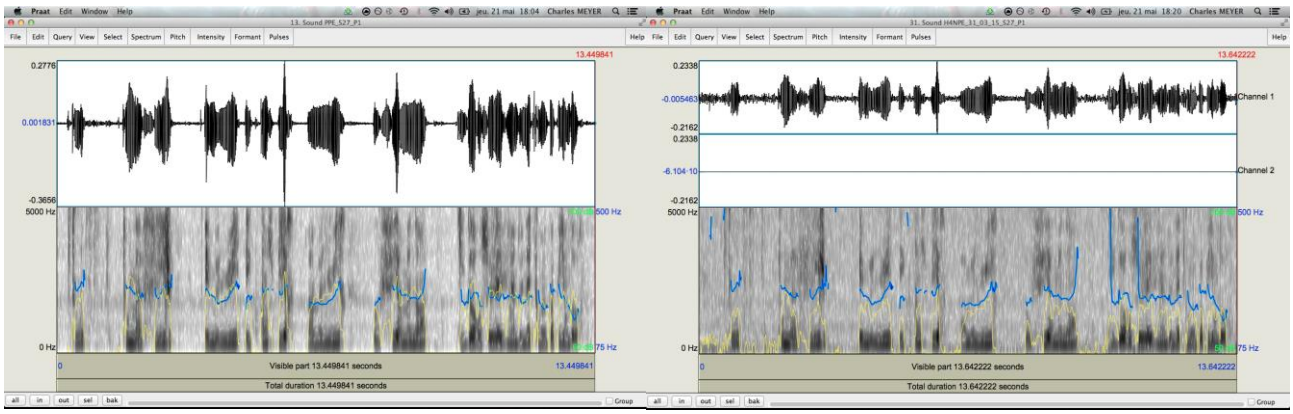
9. Première série de phrases, sujet masculin âgé de 40 ans.



10. Deuxième série de phrases, même sujet.



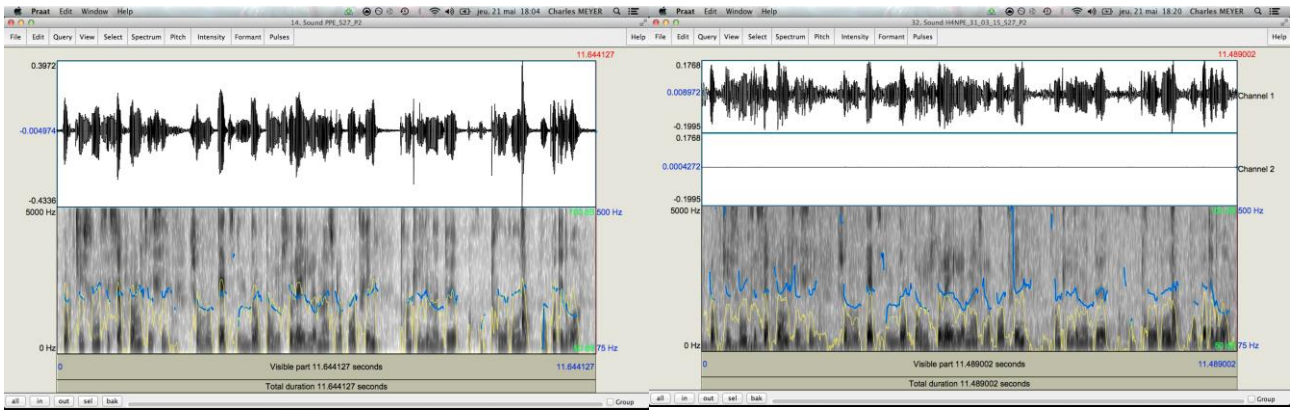




11. Première série de phrases, sujet féminin âgé de 11 ans.

12. Deuxième série de phrases, même sujet.

13. Première série de phrases, sujet masculin âgé de 10 ans.



14. Deuxième série de phrases, même sujet.

Il se dégage de ses mesures différentes conclusions :

– L'âge du sujet semble avoir une influence sur l'écart-type des valeurs de F_0. Plus un sujet est âgé, plus la valeur de F_0 est stable, avec assez peu d'évolutions. À l'inverse, les sujets plus jeunes ont une voix plus agile aux évolutions abruptes.

– Le sexe du sujet influence la valeur de F_0, mais ne semble pas influencer la stabilité des résultats.

– Certains sons, dont les respirations, les claquements de langue, les consonnes fricatives dentales non voisées (le son [s]), les consonnes occlusives dentales non voisées (le son [t]), les consonnes occlusives vélo-palatales non voisées (le son [k]) et les consonnes occlusives labiales non voisées (le son [p]) présentent une dynamique instantanée importante et un contenu spectrale très riche, ce qui explique les rupture de continuité des différentes courbes mélodiques.

– Au cours des enregistrements menés auprès des élèves de la classe de CM2 de madame Valour, des bruits extérieurs ont parasité certains enregistrement. Lors

de leur analyse dans Praat, nous avons pu observer que ces bruits provoquait également une instabilité pouvant être nuisible à l'utilisation de l'analyse spectrale au sein d'un jeu. Ce qui confirme que le bruit de fond de l'environnement dans lequel le joueur se trouve doit être contrôlé pour limiter son influence sur sa partie.

GLOSSAIRE

Jeu vidéo :

Ensemble d'interactions, liant d'une part une interface électronique et audiovisuelle à un joueur, régies d'autre part par un système de règles et dont le résultat est une expérience instrumentée.

Voix :

Ensemble des sons résultant de l'excitation d'un milieu par le larynx et les cavités supra-glottiques d'un individu.

Interface

Frontière conventionnelle entre deux systèmes ou deux unités, permettant des échanges d'informations suivant des règles déterminées.

Dans le cas du jeu vidéo, les deux systèmes, le joueur et le jeu, sont mis en relation par l'intermédiaire d'une machine, qu'elle soit un ordinateur, une console de salon, un téléphone, une tablette, où tout appareil électronique pouvant faire fonctionner un jeu.

Interaction

Action réciproque d'un élément sur un autre.

Gameplay :

Système de règles définissant toutes les interactions possibles entre le joueur et le jeu, qu'elles aient été prévues ou non par ses créateurs. L'ergonomie et l'axiomatique constituent, à elles deux, le *gameplay*.

Ergonomie :

Partie du gameplay qui régit l'interfaçage du joueur avec le jeu.

Axiomatique :

Partie du gameplay qui régit les interactions entre les différents objets informatiques qui constituent le jeu ainsi que leurs conditions d'existence.

Vocalité :

Objet sonore doté d'une qualité vocale, c'est à dire tout élément sonore qui se veut être évocatoire d'un voix.