



HAL
open science

Composer en immersion virtuelle : avatar et représentation

Christine Webster, Sophia Kourkoulakou

► **To cite this version:**

Christine Webster, Sophia Kourkoulakou. Composer en immersion virtuelle : avatar et représentation. Hybrid. Revue des arts et médiations humaines, 2022, 9, 10.4000/hybrid.2382 . hal-04383310

HAL Id: hal-04383310

<https://univ-paris8.hal.science/hal-04383310>

Submitted on 9 Jan 2024

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

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



Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License



Hybrid

Revue des arts et médiations humaines

1

9 | 2022

Arts et sciences de l'avatar technologique

Composer en immersion virtuelle: avatar et représentation

Interroger l'interaction et la coprésence dans un environnement de spatialisation audio 3D pour la VR

Christine Webster et Sophia Kourkoulakou



Édition électronique

URL : <https://journals.openedition.org/hybrid/2382>

DOI : 10.4000/hybrid.2382

ISSN : 2276-3538

Traduction(s) :

Composing in virtual immersion: avatar and representation - URL : <https://journals.openedition.org/hybrid/2968> [en]

Éditeur

Presses universitaires de Vincennes

Référence électronique

Christine Webster et Sophia Kourkoulakou, « Composer en immersion virtuelle: avatar et représentation », *Hybrid* [En ligne], 9 | 2022, mis en ligne le 30 novembre 2022, consulté le 26 septembre 2023. URL : <http://journals.openedition.org/hybrid/2382> ; DOI : <https://doi.org/10.4000/hybrid.2382>

Ce document a été généré automatiquement le 26 septembre 2023.



Creative Commons - Attribution - Partage dans les Mêmes Conditions 4.0 International - CC BY-SA 4.0
<https://creativecommons.org/licenses/by-sa/4.0/>

Composer en immersion virtuelle: avatar et représentation

Interroger l'interaction et la coprésence dans un environnement de spatialisation audio 3D pour la VR

Christine Webster et Sophia Kourkoulakou

Le projet VRAS (*virtual reality auditory space*)

- 1 Depuis 2019, le CICM/MUSIDANSE de l'université de Paris 8 et l'EnsadLab Spatial Media collaborent sur le projet de l'EUR « ArTeC VR Auditory Space »¹. Ce projet, envisagé comme une preuve de concept, met à l'épreuve de la réalité pratique, les jalons d'une interface de spatialisation 3D immersive à destination des compositeurs et compositrices de musique électronique et des artistes numériques². Il procède à une extension compatible pour la réalité virtuelle du système de spatialisation (HOA) Library³, (bibliothèque de spatialisation ambisonique aux ordres élevés). Notre but est de créer un environnement de composition pour la réalité virtuelle en s'appuyant sur les techniques du jeu vidéo, d'une part, et celles de la spatialisation ambisonique 3D et binaurale d'autre part. La première étape du projet a consisté à l'élaboration d'une bibliothèque ambisonique aux ordres élevés (*High Order Ambisonics* ou HOA), pour la VR, qui peut être insérée dans Unity 3D sous la forme d'un insérable ou *component*. La seconde étape, qui est celle de l'élaboration de l'environnement de spatialisation en-jeu (*in-game*), a transposé une partie des fonctionnalités de Unity 3D⁴ nécessaires à la gestion des AudioSources HOA, directement dans l'espace de la scène virtuelle du VRAuditory Space. Se posent alors deux problématiques, celle de l'interaction du sujet en immersion avec les interfaces 3D que nous avons développées et celle de la nécessité d'un retour physique métaphorique ou retour avatarial. *Doit-il y avoir ou non la présence d'un avatar dans l'environnement de spatialisation ou non ?*
- 2 Si l'usage et l'intégration de l'avatar à la troisième personne semblent aujourd'hui optimisés depuis l'apparition des mondes virtuels et les jeux massivement multijoueurs en ligne sur écran, de nombreuses problématiques subsistent dans le cas d'une

immersion VR, depuis l'apparition des casques stéréoscopiques des années 2010, car le corps physique doit alors se synchroniser partiellement ou entièrement avec un avatar. De la conception de l'avatar en passant par les différents systèmes de suivi de mouvements ou *tracking* (cinématique inversée, capture de mouvement, etc.), jusqu'aux problématiques des architectures de *networking*, les questions subsistent pour notre cas d'espèce : quel type d'avatar ? Quelle apparence ? Et avec quelles technologies pouvons-nous l'intégrer au VRAS ?

Compositeur ou compositrice en immersion VR, de nouveaux enjeux multidimensionnels

- 3 En pratique, le compositeur ou la compositrice de musique électroacoustique vit dans un environnement hybride, composé pour une part de machines et d'instruments (instruments acoustiques, synthétiseurs *hardware*, contrôleurs, microphones, table de mixage, effets *hardware* et haut-parleurs) qui font partie du studio d'enregistrement et de mixage classique, mais aussi de représentations informatiques de cet espace, projeté en deux dimensions. Les stations audionumériques telles que ProTools, Ableton Live ou Cubase sont des métaphores numériques du studio électroacoustique initial, car ils synthétisent l'ensemble des processus audionumériques nécessaires à l'élaboration d'une composition en télémanipulant une console, des effets et des instruments virtuels dans les limites de l'espace de l'écran informatique. Ces deux espaces communiquent entre eux par l'intermédiaire des cartes son, du protocole MIDI ou OSC.
- 4 De plus, les techniques de reproduction machiniques du son virtualisent le champ auditif naturel par un processus de transduction de l'énergie acoustique dans le domaine électrique qui suit la chaîne *captation-transformation-diffusion* du signal que l'informatique traduit dans le domaine numérique par une conversion *analogique vers numérique* en entrée et *numérique vers analogique* en sortie. La notion de virtuel et de virtualisation se trouve donc au cœur des techniques de reproduction du son et cela dès le XIX^e siècle (Sterne, 2015), notamment avec la reproduction du principe de fonctionnement du tympan, que l'on retrouve dans la membrane du microphone et du haut-parleur.
- 5 Lorsque nous transposons un dispositif de diffusion sonore par haut-parleurs vers la réalité virtuelle, telle qu'elle a été définie par Fuchs (2000), nous passons d'une forme de virtualisation à une autre. Nous intégrons plusieurs couches de virtualisation existantes et opérant à un certain niveau (machinique, informatique), vers un nouvel ensemble (espace numérique partagé) qui les contient et les virtualise tous.
- 6 Dans son traité, Philippe Fuchs nous dit (2000) :

La réalité virtuelle offre une dimension complémentaire pour le spectateur : elle lui propose un environnement dont il devient acteur, afin de permettre une activité sensori-motrice, dans un monde artificiel, qui est soit imaginaire, soit symbolique, soit une simulation des certains aspects du monde réel. Pour être considérée comme telle, la réalité virtuelle doit répondre à deux conditions qui sont l'immersion du sujet dans l'environnement simulé et l'interaction du sujet avec cet environnement. (p. 380)
- 7 Dans le projet VRAS, l'espace auditif virtuel et l'espace compositionnel, accessibles par l'interface graphique, fusionnent dans l'espace numérique partagé⁵. Dans ce contexte, la place occupée par le compositeur-auditeur ou la compositrice-auditrice, lorsqu'il ou

elle est en immersion VR complète (immersion VR appareillée, avec casque stéréoscopique, casque audio et manettes de contrôle et un suivi de position ou *tracking* en six degrés de liberté), subit un certain nombre de transformations sensorielles et opérationnelles, qui se situent sur plusieurs niveaux.

- *L'espace d'écoute subjective.* Un espace auditif virtuel est synthétisé par le *3D auditory display* du VRAS, c'est-à-dire une bibliothèque HOA à l'ordre 3 que nous avons développée pour le projet, et qui est intégrée à la scène Unity sous la forme d'un insérable. Le champ sonore synthétisé par HOA est décodé en sortie en binaural et parvient à l'auditeur par l'intermédiaire de son casque audio. Le suivi (*tracking*) de position de la tête se fait à partir de la caméra virtuelle Unity qui est couplée au Hmd (casque stéréoscopique), qui prend en charge les écouteurs. Ainsi, le moindre mouvement de tête de l'auditeur-trice est synchronisé avec l'objet caméra, qui balaye le champ auditif synthétisé en-jeu (*in-game*). Ce dispositif permet de positionner dans la scène VR des sources monophoniques, stéréophoniques, multiphoniques et ambisonique 2D ou 3D relativement à la présence effective de l'humain dans l'espace 3D.
 - *L'interface graphique.* Dans le projet VRAS, l'édition des paramètres des sources audio se fait en temps réel (en mode édition/lecture simultané), par l'intermédiaire d'une interface graphique projetée *in-game*. Le compositeur-auditeur ou la compositrice-auditrice en immersion édite les sources de chaque *preset* de spatialisation, mais il ou elle manipule également le positionnement et les trajectoires de sources ponctuelles, qu'il peut disposer librement dans l'espace. L'ensemble de ces interactions se fait avec la manette de contrôle du casque HTC Vive. En fin de processus, chaque session du VRAS peut être enregistrée, partagée et ouverte par un tiers.
 - *La mobilité.* Le VRAS est conçu comme un espace de mobilité pluriel. Nous avons envisagé trois modes de locomotion pour le ou la compositeur-trice et l'auditeur-trice : la déambulation physique naturelle, dans les limites de l'espace de *tracking* ; la téléportation d'un point à un autre de l'espace du VRAS et le mode *human joystick*. Dans le mode *human joystick* (ou « joystick humain »), de simples mouvements du buste avant-arrière, gauche-droite et haut-bas remplacent le contrôle des mouvements classique sur *Gamepad* ou clavier alphanumérique (avec les touches WASD) en lui ajoutant le déplacement dans l'axe de la hauteur. Ce type de mobilité associé à la spatialisation libre permet d'imaginer des structures spatiales complexes et innovantes dans toutes les dimensions de l'espace du VRAS.
 - *La présence.* Le phénomène de présence ou d'*embodiment* (ou incarnation) en immersion VR décrit la sensation psychologique d'*être-là*, présent à l'intérieur de l'environnement virtuel, d'en faire partie intégrante. Cette sensation qui est une des clefs de voûte de l'immersion VR est renforcée par la coprésence environnementale (les objets qui se trouvent dans l'espace), sociale (d'autres personnes/avatars se trouvant dans cet espace) et la perception de soi par l'intermédiaire d'un avatar.
- 8 La version du VRAS que nous avons finalisé en 2021 est mono-utilisateur, en vue subjective et ne propose pas de représentation métaphorique par un avatar représenté. La problématique reste donc complètement ouverte et à l'étude, notamment si nous prévoyons de réaliser une version *multi-user*, possibilité que nous avons commencé à défricher en 2020 et 2021. Il en découle la nécessité de revisiter l'histoire de l'avatar et de ses potentialités, pour approfondir ce questionnement.

Retour sur la représentation avatariale des années 2000 à nos jours

- 9 La représentation avatariale évolue en parallèle avec les sciences et technologies de l'information, dans les domaines de l'informatique et des réseaux, lorsque les applications requièrent une implication partielle ou complète du corps de « l'immersant·e », c'est-à-dire de la personne immergée (simulateurs de vol, jeux vidéo, réalité virtuelle). L'utilisateur ou l'utilisatrice est alors projeté·e dans l'espace virtuel par l'intermédiaire de son « enveloppe virtuelle pilotable » ou avatar (Lucas & Amato, 2013).
- 10 Dans le domaine de l'informatique, l'interaction humain-machine prend réellement son essor à partir de l'année 1983 avec l'interface graphique 2D du système d'exploitation Apple Lisa, son clavier alphanumérique et sa souris. Le principe de cette configuration reste toujours en vigueur dans les systèmes informatiques actuels. Dans ce contexte, le pilotage d'un avatar passe par un écran 2D, les touches du clavier, la souris ou encore une manette de jeu, ou un *joystick*. Ce système de visualisation et de contrôle permet à l'utilisateur·rice de passer alternativement de la vue à la troisième personne (*third person controler*) à la vue subjective (*first person view*). Dans un jeu vidéo massivement multijoueur, on privilégiera plutôt la vue subjective pour un FPS (*first person shooter game*), tandis que l'on préférera utiliser la vue à la troisième personne dans le cadre d'un jeu d'équipe comme *World of Warcraft* où le joueur ou la joueuse doit avoir une vue d'ensemble pour interagir avec son environnement, son interface de jeu et ses coéquipiers et coéquipières. Le principe des vues est également en lien avec la notion de personnage. Au lieu de transposer des concepts cinématographiques tels que « première personne » ou « troisième personne » à la discussion sur les avatars, comme cela a été le cas dans le passé, nous proposons l'idée de « positions modifiées » de l'avatar, qu'il s'agisse de jouer « à l'intérieur » ou « à l'extérieur » du personnage (Gazzard, 2009).
- 11 Par ailleurs :

Daniel et Nadia Thalmann avaient proposé une classification des acteurs virtuels en 4 parties (Thalmann, 1996) : les avatars qui correspondent à des représentations des visiteurs dans les mondes virtuels, les acteurs guidés dirigés par l'utilisateur, mais dont les mouvements ne sont pas exactement les mêmes que son manipulateur, les acteurs autonomes, dotés de capacités d'action et de perception dans le monde virtuel et les acteurs perceptifs conscients des autres acteurs et des humains. (Thalmann cité par Plessiet, 2019)
- 12 La réalité virtuelle utilise un système de représentation audiovisuel et de contrôle différent de celui de l'informatique 2D. L'usage de la stéréoscopie impose une vue subjective constante. Se posent alors la question de la bilocation et de l'incarnation – ou *embodiment* – du sujet. La bilocation désigne le fait d'être présent·e à la fois ici et ailleurs : le corps physique se trouve en situation de base, dans la vraie vie (IRL, *in real life*), tandis que le sujet est projeté dans le monde numérique, *InWorld*, par la vue et l'ouïe. Le résultat de cette projection génère la sensation d'inclusion corporelle, ou *embodiment*, soit la sensation de faire partie de l'espace numérique. Celle-ci dépend en grande partie d'un court-circuitage (ou suspension d'incrédulité), lorsque le corps physique bilocalisé croit, sous le contrôle sensori-moteur du casque stéréoscopique, faire partie du monde numérique. Cette sensation psychologique et cognitive est

perçue globalement, et le sujet se vit entièrement dans cet « ailleurs », avec ou sans représentation avatariale. Cette dernière ne fera que renforcer cette sensation. L'*embodiment* se manifeste également sur le plan auditif. Grâce au suivi de position des mouvements du casque stréréoscopique et du casque audio qui lui est associé, et qui participent tous les deux à la restitution de la scène auditive binaurale, procurant au sujet une sensation d'externalisation du son perçu, (la scène auditive virtuelle est perçue comme le serait une scène auditive naturelle, venant de l'extérieur à soi vers sa localisation dans le monde simulé).

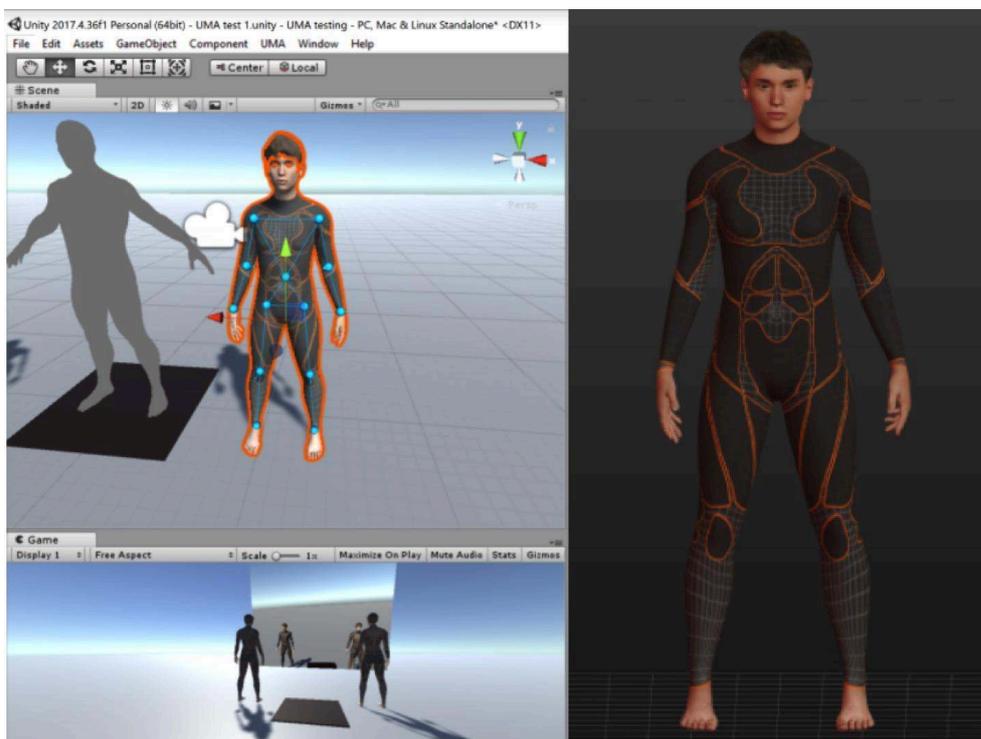
- 13 Depuis les années 2000, l'usage de l'avatar se répand dans les jeux vidéo massivement multijoueurs en ligne, les métavers, les réseaux sociaux et les plateformes collaboratives VR.
- 14 Les Sims, Second Life (SL) ou World of Warcraft (WOW) partagent la même approche, celle d'un avatar à la troisième personne, avec parfois un passage ou *switch* en vue subjective sur SL et sur WOW, configurable, piloté par un clavier alphanumérique et une souris. Les relations interpersonnelles sont enrichies d'un *tchat* vocal intégré qui permet de communiquer par la voix en temps réel avec les autres résidents présents sur la simulation. Les moyens de locomotion proposés comprennent la marche, la course, le *fly* (déplacement en vol) et la téléportation.
- 15 Depuis les années 2010, les Gafa, comme Google et Facebook, se sont intéressés à la réalité virtuelle. En 2012, Facebook teste un prototype de métavers, Cloud Party, accessible par le biais d'un profil créé sur le réseau social. Le dispositif proposé est un clone de *Second Life* et ne sera jamais vraiment exploité. Comme pour beaucoup de réseaux sociaux (Twitter, LinkedIn, Discord, WhatsApp), la représentation avatariale restera donc cantonnée au profil imagé et au texte (faire corps avec ses publications), ainsi qu'à l'usage des émoticônes et des *likes*.
- 16 En 2021, en réponse à la pandémie de la Covid-19 qui a mis en téléprésence forcée des milliards d'individus de par le monde, Facebook lance la version beta de *Horizon Workrooms*, un métavers exclusivement dédié aux détenteur-trices du casque Oculus Quest 2. La représentation avatariale dans *Horizon Workrooms* est assujettie aux performances de *tracking* du casque Oculus Quest 2 qui ne propose par défaut qu'un suivi de position en trois points, tête (Hmd), mains ou manettes de contrôle. De ce fait, les avatars ne représentent que la moitié d'un corps, amputé de ses jambes. Alors que l'année précédente, le salon français Laval Virtual a été entièrement transformé en Laval Virtual Worlds, une plateforme virtuelle 3D via l'application Virbela, fonctionnant sur le même modèle que *Second Life* ou que les jeux en ligne massivement multijoueur. À ce jour, la plateforme sociale de réalité virtuelle la plus perfectionnée en matière de gestion et de représentation des avatars est VRChat. Elle permet aux joueur-ses d'interagir à travers des chambres virtuelles appelées « Mondes ». Dans VRChat, la configuration avatariale est intégrée par défaut à partir d'un panel de modèles publics et/ou la possibilité de modéliser un avatar personnalisé, en passant par Unity 3D. Les avatars, représentés de plain-pied, se pilotent soit à la troisième personne (en 2D sur écran informatique) ou en vue subjective, c'est-à-dire en immersion VR dite *full-body* (corps entier), avec Hmd et manettes de contrôle. Le *suivi* des mouvements du corps est en six points et s'étend jusqu'à la synchronisation labiale et à l'*eye blinking* (battement de paupières). C'est à ce jour le système le plus complet tant au niveau avatarial que sonore, puisqu'il gère également l'audio 3D (hérité de Unity 3D).

Un avatar, pourquoi faire ? Notre retour d'expérience

- 17 Une immersion VR appareillée impose au sujet de faire l'expérience de l'environnement du VRAS en vue subjective. La perspective et la problématique d'une représentation avatariale *full-body* était donc envisagée dès le départ du projet. Nous avons comme source d'inspiration et comme point de repère le projet de réalité mixte OneSight développé par le Jet Propulsion Laboratory de la Nasa pour les lunettes HoloLens de Microsoft. Le projet VRAS étant un projet exclusivement VR, il n'y donc pas de superposition entre la réalité IRL et la scène VR, nous avons préféré travailler avec le casque HTC Vive, qui est polyvalent à ce niveau-là. Nous avons pu observer, dans le cadre de projets immersifs qui ont précédés le VRAS, qu'une absence de représentation avatariale du sujet ne pose pas vraiment de problèmes en mono-immersion, tant que le sujet est en position assise ou assis-debout, qui sont des positions cognitivement sécurisantes puisque le sujet n'a pas obligation de mettre l'ensemble de son corps à contribution pour se déplacer dans l'espace virtuel. Dans le cadre du projet *Empty Room* (Webster & Sèdes, 2018), l'auditeur-trice se déplaçait, tout en étant assis, à l'aide d'un contrôleur de type *joystick* de *Gamepad*. Dans le projet d'écoute comparée entre un acousmonium projeté IRL et sa reproduction en VR (Webster *et al.*, 2020), la comparaison se faisait par un seul point d'écoute de référence (ou *sweet spot*) situé au même endroit en IRL et en VR, debout tête droite, sans plus de sollicitations physiques autre que l'audition et de petits mouvements de la tête.
- 18 À partir du moment où nous sommes passés en position debout et en déambulant à travers nos dispositifs de spatialisation, l'absence de retour visuel avatarial est devenue rapidement problématique, notamment en ce qui concerne le maintien de l'équilibre général : en station debout, tant que le regard est porté dans l'axe frontal, latéral et légèrement en élévation, l'équilibre du corps reste globalement stable. La visualisation et l'utilisation des manettes qui sont associées aux mains par le cerveau aident également le corps à se synchroniser aux mouvements de ces dernières. De ce fait, l'interaction avec les objets se trouvant dans l'espace se fait assez naturellement. À partir du moment où nous nous déplaçons et que nous portons le regard en deçà ou au-delà des manettes et surtout vers le bas à l'endroit où nous devrions rencontrer notre corps, l'absence de l'avatar créé un *trou perceptif* qui peut troubler la déambulation et rendre la sensation de présence inconfortable, voire désagréable. De brefs vertiges peuvent apparaître ainsi qu'une forme de sidération passagère. Le sujet va alors se figer dans un premier temps, puis se déplacer à nouveau, mais de façon peu assurée. Ces dissensions perceptives et cognitives ont pour effet de sortir le sujet de l'expérience immersive.
- 19 Au-delà de cette problématique subjective se pose également, dans le contexte d'une immersion multiprésentielle, une problématique intersubjective qui réintroduit l'intérêt de l'avatar : il faut voir (autrui) et être vu(e) (par autrui), pour interagir avec l'environnement du VRAS de manière collaborative, en écoutant ensemble la scène audio 3D et tout en éditant et déplaçant une source sonore à tour de rôle dans l'environnement du VRAS.
- 20 Dans le cadre de notre projet, un avatar *full-body* correctement synchronisé permettrait de :
- privilégier la position debout en mode *roomscale* avec une représentation avatariale complète à l'échelle des opérateurs, synchronisé à leurs mouvements ;

- consolider la présence physique subjective, en évitant les trous cognitifs et les sorties d'immersion ;
 - travailler en mode collaboratif intersubjectif *in-game* ;
 - d'envisager l'avatar comme une extension de l'interface graphique *in-game* en reportant sur ce dernier une partie des fonctions d'éditations.
- 21 Pour atteindre en partie ces objectifs, nous sommes d'abord passés par une phase de modélisation de l'avatar, suivie par la phase de *tracking* du modèle, puis de l'intégration du modèle *trackable* dans une configuration multijoueur.
 - 22 Pour la modélisation de l'avatar VRAS nous avons utilisé le programme Make Human. Pour la synchronisation humain-avatar nous avons utilisé le plugin Vrik Root Motion. Et pour le *tracking*, nous avons utilisé un dispositif en six points avec le HTC Vive : tête, main gauche, main droite, pelvis, pied gauche, pied droit.
 - 23 Pendant la phase des tests de synchronisation, nous avons rencontré des problèmes d'incohérence au niveau du retour de position envoyé par Vrik, la position de l'avatar ne correspondant pas à la position du sujet humain. Nous avons également constaté des problèmes au niveau de l'homothétie entre la taille du sujet humain et sa représentation avatariale. Une parfaite correspondance entre la taille de l'opérateur et celle de l'avatar est essentielle. Si la hauteur est facilement reproductible, nous avons eu des problèmes de longueur de bras, ce qui par extension induit des problèmes de mains au niveau des contrôleurs. En vue subjective très rapprochée, nous avons constaté des déformations de perspective au niveau des textures et des distorsions physiques au niveau des mains et des bras de l'avatar.

Figure 1



VRAS. Capture d'écran montrant notre avatar VRAS dans Unity 3D. À gauche, visualisation des points d'articulations mis en place avec le programme VRIK sur notre modèle. À droite, le modèle masculin VRAS, réalisé dans Make Human.

Crédits : Christine Webster & Sophia Kourkoulakou.

- 24 Pour la phase des tests d'immersion en mode multijoueur, nous avons testé deux solutions de travail en réseau sur Unity compatibles avec le projet VRAS, le Network Manager intégré à Unity qui permet de créer un système de connexion client/e/serveur local et la solution client/e/serveur *cloud* Photon qui est une plateforme de réseau multijoueur indépendante, compatible avec Unity.
- 25 Après avoir réalisé une première série de tests en réseau local (en mode LAN, *local area network*) entre deux PC, nous sommes passées sur la phase de tests avec Photon PUN et le Photon Unity Network : nous avons réalisé ce test d'immersion VR à deux opérateurs avec la version PUN *free* qui permet d'instancier un maximum de 20 joueurs/ses au moyen du *cloud* Photon. Les tests se sont déroulés à Paris, entre la place des Fêtes et Belleville pendant le confinement total de la Covid-19 de mars à mai 2020.
- 26 Pour atteindre notre objectif d'instanciation par le *cloud*, nous sommes passées par un certain nombre d'étapes, d'abord en immersion VR avec des avatars pilotés en vue à la troisième personne et au final en immersion VR en vue subjective avec nos avatars VRAS.
- 27 En vue à la troisième personne, le test d'immersion à distance avec deux avatars VRAS a immédiatement modifié l'expérience de l'immersion, la simple présence des avatars dans l'environnement de travail VRAS change le rapport que nous avons avec ce dernier. Nous avons fait l'expérience de la perception d'autrui dans cet espace, de son échelle numérique, ce qui transforme également l'échelle de l'espace de travail. Nous avons pu nous projeter dans cette relation. La pertinence d'une immersion VR *full-body* s'est alors pleinement imposée.
- 28 En 2020, nous avons réalisé une première immersion VR *full-body* en vue subjective avec deux avatars VRAS et en 2021 nous avons réussi une triple immersion. Mais il reste encore de nombreux problèmes techniques à régler. La problématique du *tracking* reste ouverte aux études et aux expérimentations les plus diverses. Pour l'heure, l'industrie de la VR privilégie le *tracking* en trois points, l'usage de la main (virtuelle) et des manettes et une représentation avatariale partielle (haut du corps). Et même si VRChat ou Unreal Engine proposent aujourd'hui des facilités d'intégration avatariales sur leur plateforme, la modélisation et la mise en place de ces derniers restent un travail complexe pour des productions artistiques et de recherches indépendantes.

Figure 2



VRAS. Capture d'écran montrant une immersion à distance entre trois avatars réalisé en Mai 2021. Le point de vue de la caméra est l'avatar 1 qui observe les avatars 2 et 3 qui tentent de se passer un objet à distance.

Crédits : Christine Webster & Sophia Kourkoulakou.

Rétrospective de l'avatar, entre corps perdus ou retrouvés et posthumanisme

- 29 Les problématiques de représentation avatariale dans le contexte du projet VRAS sont le résultat d'un long processus de transformation amorcé au XIX^e siècle avec l'essor des technologies de reproduction de l'image et du son par des moyens mécaniques, électroniques et informatiques. Ces techniques reposent, d'une part, sur la virtualisation d'une partie de l'oreille interne (le tympan), par l'intermédiaire des machines tympaniques (microphones, magnétophones, haut-parleurs) et, d'autre part, sur la virtualisation du fonctionnement interne de l'œil (la pupille), avec les techniques photographiques et cinématographiques.
- 30 Les techniques de reproduction visent tous les corps qui se trouvent dans un champ d'expérience donné : le champ acoustique visé par le microphone et le champ visuel visé par la caméra. Ces processus techniques arrachent au réel une partie de sa *substance* comme le fait très bien remarquer R. Murray Shafer (1979) :
- Nous avons dissocié le son de sa source, nous l'avons arraché à son orbite naturelle, nous lui avons donné une existence amplifiée et indépendante (p. 134).
- 31 Dans la chaîne audionumérique, le son devient acousmatique, « ce que l'on entend sans voir la cause dont il provient », nous dit Pierre Schaeffer (1966), un espace auditif virtuel se projette d'où l'interprète et l'instrument sont absents physiquement, privé de son *hic*

et nunc et de son *aura* originelle, dira Walter Benjamin (1953), se réactualisant à l'identique et à l'infini du fait de sa reproductibilité. *In fine*, à travers ces processus, la réalité n'est plus, seule la représentation est réelle : « Le réel ne peut être représenté, la représentation seule peut être représentée. Pour que le réel puisse être représenté, il doit être porté à notre connaissance, et la connaissance est déjà une forme de représentation⁶ » (Altman, 1992).

- 32 Dans l'espace informationnel, l'arrachement au réel par la numérisation se traduit en même temps par un mouvement d'intégration : l'objet virtuel (texte, son, image) est intégré dans l'espace de la *programmation orienté objet* et projeté sur l'écran informatique. *De facto*, la représentation avatariale est toujours à mettre en relation avec une certaine qualité d'espace projeté dont il fait partie. De la même façon que le son est en relation directe avec l'espace acoustique ou que le sujet photographié l'est avec son environnement, l'avatar ne peut être dissocié de l'espace de l'écran, l'espace de la page web, l'espace du métavers, ou l'espace de la scène VR. En fonction de la qualité de l'espace concerné, plusieurs stratégies d'*embodiment* sont possibles. Faire corps et acte de présence avec le texte caractérisait le début du Web, prolonger cette présence dans une fusion multimédia (texte, icône, photo, vidéo, musique) caractérise nos échanges dans les réseaux sociaux actuels. La réalité virtuelle ajoute un niveau de représentation supplémentaire en réintroduisant l'ensemble du corps humain dans la boucle d'interaction homme-machine, à la fois comme un téléopérateur mais aussi comme un téléreprésenté. C'est le retour du ici et maintenant et de l'unique expérimenté en immersion VR.
- 33 Paradoxe donc, mais c'est aussi là tout l'intérêt de la réalité virtuelle que de remettre la question du corps humain au centre de la relation humain-machine et ainsi bousculer les postures radicales ayant émergé dans les technosciences depuis l'ère de la cybernétique. Pour les cybernéticien·nes, l'information est un être immatériel qui fluctue au gré des principes de régulation par rétroaction (ou *feedback*) théorisé par Norbert Wiener⁷. D'autre part, en 1950, Wiener émet l'idée qu'il serait possible d'extraire le modèle informationnel d'un homme et par là de le connecter à d'autres machines informationnelles. En 1954, après la conférence de Macy à Paris, l'informatique se désolidarise du courant cybernétique et entre dans une nouvelle ère, portée par un imaginaire qui se concentre autour de l'analogie cerveau-ordinateur, principe cautionné par Turing et Von Neumann⁸. Cette vision impose la logique mathématique et statistique comme valeur universelle, seule capable de comprendre et de transformer le monde. À terme, l'informatique doit remplacer l'humain dans ses fonctions de prise de décisions et de production de connaissances. L'humain ne serait plus la mesure de toute chose⁹. L'ordinateur devient une machine pensante, capable d'imiter l'humain (voir test de Turing¹⁰). Dans la vision ordinatrice de Turing et Von Neumann, le corps humain a complètement disparu, l'idée n'est plus de savoir *qui* pense mais *qu'est-ce qui* pense.
- 34 La question du corps, et par extension celle du corps numérique et de sa représentation, sera introduite dans les années 1990, dans la continuité du courant cyberpunk, avec le livre de Neal Stephenson, *Snowcrash (Le samouraï virtuel, 1992)*, qui décrit et préfigure de ce que sera la réalité de notre condition avatariale sur les réseaux informationnels. Depuis, l'avatar est devenu une réalité qui doit être prise en considération dans le continuum des technologies de l'information.

- 35 Katherine Hayles, dans *Comment nous sommes devenus post-humains* (How We Became Post-Human, 1999), l'affirme : l'information a perdu son « corps » (*lost its body*) non pas pour être dissimulé, mais pour être incarné autrement. Dans le chapitre « *Toward Embodied Virtuality* », Hayles expose en quatre arguments les manières dont l'humain est déjà en train de devenir posthumain : d'abord, en privilégiant la *texture* informatique au lieu de la/sa matérialité, l'incarnation corporelle est biologiquement vue comme un « accident de la nature » au lieu de l'inévitable conséquence de la vie. De plus, le concept de conscience posthumaine est considéré comme un épiphénomène : un phénomène accessoire qui accompagne un phénomène essentiel, sans être pour rien dans son apparition. Le troisième argument stipule que la théorie de posthumain considère le corps comme une prothèse originale que l'on apprend à manipuler, et dont l'extension ou son remplacement avec une autre prothèse synthétique devient la continuation d'une procédure d'apprentissage profond qui a commencé avant la naissance. Le quatrième argument, et le plus important : la théorie posthumaine configure l'humain de façon qu'il puisse être articulé, autant qu'intégré directement, avec des machines *intelligentes*.
- 36 Dans la vision posthumaniste, il n'y aurait pas de différences essentielles ou de démarcations absolues entre l'existence corporelle et sa simulation numérique, entre le mécanisme cybernétique et l'organisme biologique, entre la technologie des robots et les objectifs humains, mais plutôt celle d'un devenir hybride :
- Le posthumanisme propose de ne plus considérer l'interface avec la technologie sous la forme d'une ergonomie de l'instrument qui prolonge le corps en restant extérieur, mais sous la forme d'une compénétration qui remet en discussion sa séparation et sa centralité en posant la question de son identité et son degré de liberté. (Antonioli, 2011, p. 173-180)
- 37 Cette vision se désolidarise complètement du courant techno-déterministe et *a fortiori* transhumaniste, qui embrasse la technologie de manière totalitaire, sans en critiquer véritablement les enjeux. Elle prend plutôt racine dans l'approche du cyborg tel qu'il a été défini par Donna Haraway (1984) dans une pensée postféministe.
- Une approche « cyborg » implique également un nouveau regard sur la technique, pour lequel la machine ne peut plus être considérée simplement comme « une chose » ou « un outil ». La machine est un aspect de notre corporéité et de notre sensibilité, un élément essentiel de nos productions de subjectivité. Les machines ne se réduisent pas à une menace de domination, puisque nous sommes appelés à en être responsables, comme nous sommes responsables des frontières et des limites que nous instituons. (Antonioli, 2011, p. 173-180)
- 38 À partir de l'approche *cyborg* de Haraway, nous constatons que ce qui caractérise fondamentalement la pensée posthumaniste, c'est d'abord le fait de remettre l'être humain au centre du dispositif humain/machine ; comme c'est justement le cas avec la réalité virtuelle. Ici, l'humain redevient la mesure de toute chose, dans une volonté de symbiose intelligente et raisonnée, entre l'écosystème du vivant et du non-vivant. En cela le posthumanisme se différencie du transhumanisme car il privilégie, dans la continuité de sa critique de l'humanisme des lumières, l'hybridation ; le faire avec ; le créer avec ; qui s'opposent catégoriquement à la perspective d'une dissolution complète de l'homme dans la matrice informationnelle, telle qu'elle est envisagée par les transhumanistes. Haraway propose de réfléchir et de créer, *think we must*, en partant du concept de *symptièse* qu'elle définit comme « une organisation en système complexe, dynamiques, réactifs, situés et historiques » dans une boucle relationnelle humain/machine/vivant, qui casse l'opposition binaire humains/non-humains. Par cette

approche, Haraway dit se situer au-delà du posthumanisme et des enjeux de l'anthropocène/capitalocène, pour aller vers le « Chtuhlucène », en référence aux étranges divinités imaginées par l'auteur fantastique H. P. Lovecraft¹¹.

Conclusion : des virtualités au-delà des formatages actuels ?

- 39 Si la réalité virtuelle se situe à l'une des extrémités du continuum réalité-virtualité proposé par Milgram (1994), et en supposant qu'il n'y aurait rien d'autre qui puisse être concevable au-delà de la théorie de l'information telle qu'elle existe, elle pourrait être considérée comme le lieu d'un aboutissement ultime, où absolument tous les champs et les degrés des virtualisations possibles se synthétisent (multimédia, hypermédia, métamédia, etc.). La question n'est plus de savoir si le virtuel et la virtualisation sont souhaitables, puisqu'ils imbibent notre façon d'être au monde et que nous le percevons depuis le XIX^e siècle à travers des yeux et des oreilles machiniques, l'informatisation n'ayant fait que fluidifier le processus. Pour autant, il se pourrait que nous arrivions à une impasse technologique et structurelle, car les technologies actuelles n'intègrent pas la mise en situation d'une immersion totale en VR, alors que les dispositifs annoncés promettent un monde fait pour le Web 3.0.
- 40 Le concept de posthumain réinitialise les frontières de la vie humaine en incluant êtres vivants et machiniques dans son ensemble. Cependant, les technologies de l'information appartiennent depuis les années 1950 à une élite majoritairement constituée d'hommes blancs, subordonnés au pouvoir du marché, et qui déjouent leur rôle dans ce que Guattari appelle les « micro-fascismes » (Genosko, 2017).
- 41 Depuis les années 1980, on essaye de représenter l'homme (blanc) dans un environnement de réalité virtuelle, fonctionnel. Mais ces fonctions ne se sont développées que partiellement (industrie automobile, simulateurs de vols) et les promesses d'un passage vers une réalité virtuelle dont l'accès serait démocratique et fluide sont irrémédiablement repoussées dans le temps. En l'espace de dix ans, sous l'effet de la captation du marché de la VR par les grands groupes comme Facebook, Apple et récemment ByteDance, la réalité virtuelle est passée du stade expérimental à un stade essentiellement *corporate*, au sens des marques et sociétés privées, et ceci induit un certain nombre de problèmes pour les artistes, comme pour les chercheurs.
- 42 Pour que nous puissions explorer la réalité virtuelle en profondeur et ne pas nous contenter de rester au stade de la *gamification des usages*, nous devrions avoir à nouveau le champ libre, c'est-à-dire pouvoir pratiquer notre recherche en dehors des contraintes et du contrôle des plateformes VR et d'une vision exclusivement entrepreneuriale de la réalité virtuelle. Nous sommes aux portes d'un monde pluriel dont les potentiels ne se sont qu'à peine dévoilés, que nous n'avons fait qu'effleurer et qui va bien au-delà du modèle imposé de la dystopie *futuriste* capitaliste.

BIBLIOGRAPHIE

- Altman, R. (1992). *Sound theory, sound practice*. AFI, Routledge.
- Benjamin, W. (2016). *L'œuvre d'art à l'époque de sa reproductibilité technique* [1935]. Allia.
- Breton, P. (1987). *Une histoire de l'informatique*. La Découverte.
- Fuchs, P. Moreau, G., & Berthoz, A. (2000). *Le traité de la réalité virtuelle volume 1 : L'Homme et l'environnement virtuel*. Presses des Mines.
- Gazzard, A. (2009). The avatar and the player: Understanding the relationship beyond the screen [Conférence]. Conference in Games and Virtual Worlds for Serious Applications, University of Hertfordshire College Lane, Hatfield, Royaume-Uni.
- Genosko, G. (2017). Les trous noirs de la politique : résonances du microfascisme. *La Deleuziana, revue en ligne de philosophie*, 5.
- Haraway, D. (1984). A cyborg manifesto. *Socialist Review*, 80, 65-108.
- Haraway, D. (2020). *Vivre avec le trouble*. Les éditions des mondes à faire.
- Hayles, N. K. (1999). *How We Became Post-Human. Virtual Bodies in Cybernetics, Literature and Informatics*. The University of Chicago Press.
- Lucas, J. F. & Amato, É. A. (2013). Mondes, points de vue, personnages : l'avatar comme enveloppe pilotable. Dans É. A. Amato & É. Perény (dir.), *Les avatars jouables des mondes numériques. Théories, terrains et témoignages de pratiques interactives* (p. 109-133). Hermès-Lavoisier.
- Murray Shafer, R. (1979). *Le paysage sonore*. Jean-Claude Lattès.
- Platon. (1998). *Protagoras*. Flammarion
- Plessiet, C. (2019). *Quand la marionnette coupe ses fils. Recherches sur l'acteur virtuel*. Art et histoire de l'art. Habilitation de diriger des recherches, (INRÉV) du laboratoire Arts des images et art contemporain (AIAC EA4010) Université Paris 8 Vincennes Saint-Denis.
- Shaeffer, P. (1966). *Le traité des objets musicaux*. Seuil.
- Stephenson, N. (1992). *Le Samouraï virtuel (Snow Crash)*. Le livre de poche.
- Sterne, J. (2015). *Une histoire de la modernité sonore*. La Découverte/Philharmonie de Paris-Cité de la musique.
- Thalmann, D (1996). A new generation of synthetic actors: The real-time and interactive perceptive actors (p. 200-219). Computer Graphics Lab. <https://pdfs.semanticscholar.org/9043/36c4809b4f191132642f786927a3d00eeeed.pdf>.
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind*.
- Webster, C. & Sèdes, A. (2018). Empty room, exploring the plasticity of electroacoustic music spatialization in VR with Ambisonic 3D and Binaural techniques [Conférence]. Conference on Sound Ecology and Media Culture, octobre 2018, Darmstadt-Dieburg, Allemagne. hal-02276901
- Webster, C., Raboisson, N., Lamarche, O., Couprie, P., & Genevois, H. (2020). Vers un acousmonium en immersion VR en ambisonie 3D et binaural [Conférence]. Journées d'informatique musicale, octobre 2020, Strasbourg, France. hal-02977660

NOTES

1. Projet de l'EUR ArTeC nommé « VR Auditory Space ».
 2. Démo de VR Auditory Space.
 3. Développée au Labex Arts-H2H entre 2012 et 2015, la bibliothèque HOA est un ensemble de classes C++ et FAUST et de mises en œuvre logicielles sous forme d'objets Max, PureData et VST destinés à l'ambisonie d'ordre supérieur.
 4. Unity 3D est un progiciel multiplateforme unifié, à destination des jeux vidéo et de la réalité virtuelle.
 5. VRAS démo walkthrough.
 6. *The real can never be represented, representation alone can be represented. For in order to be represented, the real must be known, and knowledge is already a form of representation.*
 7. Norbert Wiener (1894-1964), mathématicien américain, théoricien et chercheur, père fondateur de la cybernétique.
 8. La position de Turing et Von Neumann, à propos du concept de l'ordinateur envisagé comme l'équivalent du cerveau a été décrite par Breton (1987).
 9. Sentence issue du Protagoras de Platon, avec l'informatisation et l'intelligence artificielle, la raison humaine serait remplacée à terme par la raison et la conscience de la machine.
 10. Désigne la faculté d'une machine à imiter la conversation humaine. Ce test est décrit par Alan Turing en 1950 dans sa publication « Computer machinery and intelligence ».
 11. Le Chthulucène selon Haraway n'est pas une référence directe au cauchemar raciste et misogynne de H. P. Lovecraft, Haraway orthographe Chthulu à sa manière, mais une métaphore pour décrire les forces tentaculaires qui sont à l'œuvre dans le vivant : « Chthulucène s'enchevêtre à une myriade de temporalités et de spatialités, à une myriade d'entités-au-sein-d'assemblages intra-actives aussi – qui peuvent relever du plus-qu'humain, de l'autre-qu'humain, de l'inhumain et de l'humain-comme-humus » (Haraway, 2017, p. 223).
-

RÉSUMÉS

L'article se présente comme une étude de cas spécifique centré sur la problématique du ou de la compositeur·trice en immersion, telle que nous l'avons expérimenté dans le cadre du projet de l'EUR ArTeC « VR Auditory Space, spatialisation du son en immersion VR ». Comment la création d'un environnement de spatialisation conçu pour la réalité virtuelle détermine-t-il pour le·la compositeur·trice de musique électronique les modalités d'interaction avec cet espace ? La programmation et l'édition en temps réel en jeu ou *in-game* demande de réfléchir à l'intégration d'une représentation avatariale de type *full-body* ou représentation complète. Notre article propose d'aborder l'ensemble des problématiques que nous avons rencontrées d'un point de vue technique, de même que philosophique et critique.

INDEX

Mots-clés : réalité virtuelle, incarnation, embodiment, coprésence, avatar, spatialisation du son, HOA, ambisonie aux ordres élevés, binaural

AUTEURS

CHRISTINE WEBSTER

Christine Webster est compositrice et ingénieure du son. Elle expérimente depuis 2008 la spatialisation de la musique électroacoustique dans des dispositifs de réalité virtuelle. Le projet VRAS a été réalisé dans le cadre de son projet de thèse, qui explore et étudie l'émergence de nouvelles modalités de composition et de spatialisation dans les espaces numériques partagés. Cette thèse est en fin de développement à l'université Paris 8, école doctorale Esthétique, sciences et technologies des arts, et le CICM, sous la direction d'Anne Sèdes et en codirection de François Garnier, responsable du groupe EnsadLab Spatial Media.

SOPHIA KOURKOULAKOU

Sophia Kourkoulakou développe une thèse (*Design pour*) *l'art interactif: une techno-anthropologie numérique* à l'université Paris 8, école doctorale Esthétique, sciences et technologies des arts, labo AIAC (art des images et art contemporain) avec l'équipe Image numérique et réalité virtuelle (INRÉV), sous la direction de Chu Yin Chen et en codirection de François Garnier, et elle est membre du groupe de recherche Spatial Media, EnsadLab, à l'École des arts décoratifs de Paris/PSL.