

Métavers : au-delà des « casques de réalité virtuelle »

Par Guillaume MOREAU

Directeur délégué à la Recherche et à l'Innovation à l'IMT Atlantique

Depuis le discours de Mark Zuckerberg de 2021 sur le « métavers », force est de constater que la généralisation annoncée n'a pas encore eu lieu. Si les technologies immersives ont indéniablement progressé et atteint des niveaux de prix plus compatibles avec le grand public, certains défis, plus ou moins liés aux compromis effectués sur les interfaces, restent à relever. Non seulement les logiciels doivent s'adapter au compromis masse, énergie, bande passante réseau, capacité de calcul, mais les outils auteurs doivent également progresser. L'ensemble de ces systèmes n'atteint pas les capacités de la perception humaine, engendrant des effets secondaires et potentiellement des conséquences sanitaires. Enfin, il convient d'étudier les enjeux éthiques de ces technologies qui se marient progressivement à l'intelligence artificielle.

Depuis l'annonce de Mark Zuckerberg sur le « métavers » en novembre 2021, la tendance s'est retournée et la généralisation de la réalité virtuelle auprès du grand public n'a pas encore eu lieu. C'est donc l'occasion de faire un point d'étape sur les grands défis qui restent à résoudre en vue d'une adoption plus massive.

Tout d'abord, qu'entendait-il par « métavers » ? Il s'agissait de la combinaison d'une réalité virtuelle (un monde virtuel appelé Horizon Worlds), des moyens d'accès à ce monde virtuel, de possibilités d'interactions avec d'autres humains et quelques éléments autour de la monétisation. Par réalité virtuelle, on entend ici le fait de s'immerger, par exemple avec un casque comme le Meta Quest, dans un environnement artificiel pouvant représenter en partie la réalité et d'interagir en temps réel avec ce monde virtuel (Arnaldi *et al.*, 2018). On peut par exemple interagir avec des représentations d'êtres humains qu'on appelle alors des avatars, correspondant à des personnes réelles ou potentiellement mus par l'intelligence artificielle. La réalité mixte est une forme de généralisation de la réalité virtuelle où l'utilisateur perçoit simultanément des éléments du monde réel et des éléments de synthèse. Les technologies permettant d'accéder à cette réalité virtuelle font l'objet d'un autre article dans ce numéro, nous nous concentrerons sur les autres aspects. Parmi ceux-ci, il y a bien sûr des considérations logicielles liées aux nécessaires compromis matériels, à la représentation des mondes virtuels et en particulier des avatars mais aussi aux capacités du système sensori-moteur et cognitif humain. Enfin, qui dit enjeux humains dit nécessairement enjeux éthiques.

C'est ce que nous nous proposons d'aborder dans la suite de cet article.

ENJEUX LOGICIELS

Même si les matériels progressent, ils relèvent nécessairement d'un compromis entre leurs performances techniques, leur encombrement et leur poids. Tous ces paramètres

influent sur leurs capacités de calcul, de communication et leur autonomie énergétique. Se pose également la question de leurs moyens d'interaction avec l'environnement et de la création des mondes artificiels, y compris pour le cas particulier des avatars.

Du point de vue informatique, le rendu d'une scène en réalité virtuelle ressemble de très près à la synthèse d'images en temps réel telle qu'elle est pratiquée dans les jeux vidéo : à partir d'un modèle 3D de la scène généralement représenté sous forme de formes géométriques simples comme des triangles, d'un modèle d'éclairage, d'un modèle de caméra virtuelle et de lois simplifiées du rendu (modèle de Phong), on calcule le rendu graphique « suffisamment vite » (25 images par seconde sont un minimum pour la rémanence rétinienne, mais 60-90 images par seconde semblent constituer des valeurs limitant la cybercinétose). La différence se situe dans les paramètres de la caméra et le calcul de sa « pose », *i.e.* sa position et son orientation. Les casques modernes sont dotés de capteurs permettant ce calcul moyennant un algorithme de fusion de données. Le problème est bien plus complexe en réalité augmentée, comme on le verra plus loin. Ces calculs peuvent être effectués à différents niveaux : directement sur le casque ce qui limite le besoin en bande passante réseau au détriment de la consommation énergétique, sur un PC avec carte graphique dédiée, ou carrément dans le « nuage ». On retrouve les problématiques habituelles des systèmes embarqués entre technologies avec et sans fil, sécurité, bande passante et autonomie énergétique.

La création de monde virtuels est également plus complexe que la création de monde 3D pour l'image de synthèse ou l'animation (cinéma, jeux vidéo) : elle nécessite la création de moyens d'interaction entre le ou les utilisateurs et les différents éléments du monde afin de sélectionner, manipuler ou modifier les objets du monde virtuel. C'est un domaine de recherche très actif entre extension des paradigmes 2D comme l'utilisation de la souris en 3D à des expériences très diversifiées comme le suivi de la direction du regard pour la sélection en 3D (Paulus, 2021), la détection de gestes, la commande vocale... À l'inverse, comme il est quasi-impossible d'utiliser simultanément un clavier et un casque de RV, l'entrée de texte pour annoter un monde virtuel reste extrêmement peu efficace quelle que soit la métaphore proposée.

La combinaison du monde virtuel et des différents scénarios d'interaction qui était directement programmée par-dessus des *middlewares* se limitant à une couche d'interface matériel-logiciel est aujourd'hui à peu près intégrée dans un outil devenu un standard de fait, comme Unity 3D. Il reste néanmoins à créer et animer des avatars : il faut un modèle géométrique (basique ou visuellement réaliste, ce que les futurs outils de Meta se promettent de générer à partir d'un simple *smartphone*), un modèle d'animation en lien avec la biomécanique humaine et surtout un modèle comportemental. Ce dernier qui a fait l'objet de recherches actives depuis les années 1990, avec un beau succès dans le domaine de la simulation de foules, se construit aujourd'hui avec les techniques d'intelligence artificielle. Tous ces types de modèles se trouvent aujourd'hui en ligne, mais ils doivent être adaptés : par exemple, on ne peut pas simplement transformer un modèle d'humain virtuel de 1,65 m en un modèle d'1,90 m, qui ne marchera pas de la même manière. Il faut respecter certaines contraintes biomécaniques, c'est l'objectif des techniques de "motion retargeting". Certains peuvent aussi souhaiter un avatar qui ne leur ressemble pas.

Le dernier élément régulièrement évoqué est celui du couplage entre avatars réalistes et intelligence artificielle. Du point de vue technique, cela implique d'importants coûts de calcul qui ne peuvent pas être réalisés aujourd'hui sur un dispositif portable et ont pour conséquence des besoins en bande passante. L'IA embarquée est un important défi impliquant des travaux sur les architectures des composants, la compression de réseaux de neurones... La crédibilité d'un avatar intégralement piloté par intelligence artificielle n'est pas acquise aujourd'hui et soulève des enjeux éthiques que nous étudierons plus loin.



Figure 1 : Exemple de simulation de foule virtuelle (Source : Inria).

CAS PARTICULIER : LA RÉALITÉ AUGMENTÉE

En matière de réalité augmentée, la problématique du calcul est bien plus importante parce qu'il faut faire « coïncider » le monde réel et le monde virtuel : si un objet virtuel est posé sur une table réelle, il ne doit pas apparaître ni au-dessus ni en dessous de celle-ci, ni encore donner l'impression de trembler autour de celle-ci. Le calcul précis et fréquent de la position et de l'orientation du casque est donc une condition nécessaire.

En intérieur, les systèmes incorporés comme celui l'HoloLens 2 font appel à des techniques complexes mêlant projection de lumière structurée, caméras infrarouges et optiques, SLAM et fusion de données qui fonctionnent dans un espace de travail de 5 à 10 m de côté. En extérieur, le problème se corse puisque les caméras de profondeur utilisant l'infrarouge sont inefficaces. Les systèmes GNSS ne fournissent qu'une position de précision décimétrique, pas d'orientation et fonctionnent à une fréquence d'environ 1 Hz. Cela ne résout pas en outre le fait que les systèmes de réalité augmentée actuels ne sont pas suffisamment lumineux pour permettre d'afficher une information dans un environnement ensoleillé.

On retrouve cette question d'éclairage dans la scène mixte : il faut respecter les interactions entre les objets réels et les objets virtuels. En premier lieu, il convient de respecter les occultations entre les objets ; techniquement, l'écran d'un équipement de réalité augmentée est placé entre l'œil et le monde réel, les objets virtuels seront donc toujours devant les objets réels. Comme on ne peut pas altérer le monde réel, il faut alors calculer les interactions entre les différents objets réels et virtuels *via* une technique de "phantom rendering" qui suppose l'acquisition d'un modèle 3D de la scène réelle (Kalkofen *et al.*, 2011). Les techniques de modélisation 3D d'une scène réelle existent mais sont difficilement exploitables en temps réel lorsque la scène est dynamique. Une reconstruction purement géométrique de la scène n'est pas suffisante, il faut y ajouter une reconstruction photométrique afin d'assurer la cohérence de l'éclairage des deux parties de la scène. En sortie de cette chaîne, nous arrivons donc à une image virtuelle calculée qu'il faut afficher, ce qui renvoie à un problème de matériel : l'immense majorité des systèmes de réalité augmentée actuels ne savent qu'ajouter de la lumière à un pixel donné superposé au monde réel. Pas pratique pour montrer l'ombre d'un objet virtuel projetée sur un objet réel !

ENJEUX HUMAINS

Perception

La conséquence première de l'imperfection de ces dispositifs matériels et logiciels est l'impact sur le système perceptif humain. Si on compare les performances d'un casque de réalité virtuelle (par exemple un Meta Quest 3) au système visuel humain, le champ de vision horizontal atteint aujourd'hui 100 à 120° (au lieu de 180°), l'acuité visuelle monoscopique est à peu près du même ordre mais il subsiste un problème fondamental : en vision naturelle, lorsque l'humain regarde un objet, l'œil accommode sur l'objet et les deux yeux convergent également sur l'objet d'intérêt. Dans un dispositif stéréoscopique de type casque ou CAVE, les yeux accommodent toujours sur l'écran mais convergent sur l'objet d'intérêt qui peut être dans un plan différent de celui de l'écran. Cette rupture de la relation entre accommodation et convergence n'est pas naturelle (et d'ailleurs pas toujours possible). De plus, en vision naturelle, les objets qui ne sont pas autour de la zone d'intérêt ne sont pas fusionnés (diplopie) et sont flous. En vision stéréoscopique, faute de savoir suffisamment rapidement ce que les yeux regardent, il n'est pas possible de flouter en temps réel les zones non fusionnées, ce qui engendre une fatigue visuelle.

L'ensemble de ces imperfections engendrent des différences de perception entre le monde réel et le monde virtuel. Typiquement, celui-ci est perçu comme étant de taille différente (Peillard, 2020).

Qui plus est, la qualité de reproduction des *stimuli* sur l'ensemble du système sensoriel est différente selon les modalités sensorielles, le contenu de l'application et les dispositifs techniques retenus, engendrant des incohérences sensorielles pouvant être source d'inconfort mais aussi de limitation de la plausibilité ou du sentiment de présence dans le monde virtuel. Cependant, le lien entre complétude de la perception et présence est loin d'être établi : si un minimum de qualité des *stimuli* est nécessaire pour engendrer un sentiment de présence, la théorie de la « vallée de l'étrange » de Mori (2012) tend à l'inverse à dire que la perception se focalisera sur les différences les plus ténues.

Effets sanitaires de la réalité virtuelle

Tout dispositif de réalité virtuelle, artificiel par nature, sollicite le système sensori-moteur et donc cognitif de l'être humain qui l'utilise. Notamment en raison de performances encore éloignées du système perceptif humain, il n'est pas exempt d'effets sur la santé humaine, ce que démontre un rapport de l'ANSES de 2021 en séparant ces effets en fonction des niveaux de preuve (Burkhardt *et al.*, 2021).

Les effets avérés, dont l'intensité dépend des interfaces utilisées, du contenu proposé mais aussi de la sensibilité individuelle, sont :

- la cybercïnétose (30 à 50 % des utilisateurs) qui peut avoir comme symptômes des nausées, des vomissements, des maux de tête ainsi que divers autres effets physiologiques ou vestibulaires : la cause en reste mal connue, les incongruences sensorielles restent les plus citées ;
- l'altération de la coordination sensori-motrice et des capacités perceptives après exposition ;
- les effets liés au fonctionnement des casques en tant que dispositifs émetteurs de rayonnement : lumière bleue qui perturbe le cycle circadien et le sommeil en particulier chez les enfants, modulation temporelle de la lumière qui a un effet sur le déclenchement de crises chez les personnes souffrant d'épilepsie et engendre de la fatigue visuelle.

Le rapport souligne aussi des effets possibles sur lesquels il n'est pas aujourd'hui possible de conclure en raison d'un niveau de preuve encore insuffisant :

- les effets liés aux interfaces elles-mêmes et à leur ergonomie, incluant les accidents (collisions avec l'environnement par distraction ou occultation), des troubles musculosquelettiques dus par exemple au poids des casques et à leur répartition des masses ;
- les effets psychologiques et psycho-sociaux à l'instar des jeux vidéo (risques émotionnels, déréalisation, dépendance, isolement social...). Ces risques sont naturellement plus liés aux contenus qu'aux dispositifs de réalité virtuelle ; ces risques induisent de façon indirecte des risques sur le développement émotionnel et cognitif ;
- les effets sur la représentation de soi dans le monde virtuel sur lesquels nous reviendrons dans le paragraphe consacré à l'éthique.

Enjeux éthiques

Un récent rapport du CNPEN (2024) pointe les risques liés au « métavers » tout en mentionnant la difficulté d'étudier un ensemble de technologies à vocation intégrative, elles-mêmes en cours de développement.

Si certains pointent à raison que le métavers peut être un moyen d'économiser un certain nombre de déplacements, l'équilibre écologique du métavers et des technologies associées est loin d'être une évidence : les moyens opto-électroniques sont conséquents, le cycle de vie avant obsolescence est aujourd'hui très court et si la consommation électrique des casques reste par définition limitée, il n'est pas de même pour l'infrastructure de calcul (rendu, IA), de stockage et de communication.

Le métavers est-il inclusif ? Il y a plusieurs moyens de répondre à cette question : d'un côté, malgré la baisse sensible des coûts d'accès aux dispositifs et à Internet, il est bien évident que tout le monde n'est pas égal devant l'accès à ces technologies, notamment mais pas seulement dans les pays du Sud. À l'opposé, ces inégalités de richesse essentiellement peuvent être partiellement compensées grâce au métavers : un concert ou un événement sportif dont le nombre de places est limité devient accessible à plus de monde, un site historique situé à des heures d'avion ou fermé à cause du surtourisme est à nouveau accessible, même s'il faut rester conscient que l'expérience est dégradée.

Certains points sont d'un abord plus complexe, notamment en ce qui concerne les avatars : ces représentations des êtres humains, de plus en plus précises et « convaincantes » soulèvent quelques questions importantes :

- L'identification des personnes dans le métavers : si ressembler à autre chose que son corps physique peut être souhaitable, on ne peut pas négliger la possibilité d'usurpation d'identité très convaincante (*deepfake*) avec ses conséquences possibles : influence, vol de données, abus de confiance...
- De façon plus subtile encore, la littérature scientifique commence à démontrer que la modification de la représentation de son propre corps est à l'origine de modifications comportementales : être incarné dans un corps de couleur tend par exemple à diminuer les préjugés racistes (Peck *et al.*, 2013), tandis qu'une incarnation dans un corps de super-héros rend l'utilisateur plus susceptible d'aider son prochain. On peut se demander ce qu'une incarnation dans un corps représentant quelque chose de moins positif pourrait donner.
- Dans tous les cas, la notion de responsabilité doit être établie. Si le vol de données est qualifié aujourd'hui, le vol de biens virtuels n'est pas complètement défini et certaines notions restent à compléter : il y a d'ores et déjà quelques ressentis de violences sexistes et sexuelles voire de viols dans le métavers qui vont au déjà de la définition

légale du viol du code pénal. Pour autant, ceci est en contradiction avec l'objectif d'un certain nombre de jeux vidéo, moyen clé d'accès au métavers où tuer d'autres êtres « vivants » est un objectif en soi, qu'ils soient animés par le moteur de jeu ou des représentations d'autres êtres vivants. Deux systèmes moraux sont amenés à cohabiter.

CONCLUSION

Les technologies immersives sont d'ores et déjà utilisées dans de grands nombres de domaines de l'industrie, du monde de la santé et du divertissement. La généralisation est lente faute d'application incontournable (*silver bullet*) même si des défis techniques restent à régler. Pour autant, on voit progressivement les outils de réalité virtuelle passer du praticien au patient dans le domaine médical. La réalité virtuelle a des bénéfices réels et prouvés pour le traitement de certaines pathologies comportementales comme les phobies (Silva Freitas *et al.*, 2021). Elle propose néanmoins aujourd'hui une expérience dégradée par rapport à la réalité et si nous faisons une analogie avec le fameux test de Turing, le mélange ultime IA-RV ne passerait pas ce test et ne le passera probablement jamais complètement pour une personne avertie. La généralisation nécessite que le législateur se penche sur les vides juridiques qui subsistent au-delà de l'AI Act, mais gageons que nous saurons raisonner en termes de bénéfices – risques à l'instar de ce qui se fait en thérapie pour ne pas tuer dans l'œuf une technologie qui représente aussi une immense opportunité pour proposer des outils inclusifs au service de l'ensemble de la société et en particulier auprès des populations privées de capacités physiques ou cognitives, financières.

BIBLIOGRAPHIE

ARNALDI B., GUITTON P. & MOREAU G. (dir.) (2018), *Réalité virtuelle et réalité augmentée : mythes et réalités*, Londres, ISTE.

BURKHARDT J.M., BEHAR-COHEN F. GRYSZPAN O. KLINGER E., LOBJOIS R., MOREAU G. NANNIPIERI O., PALJIC A. PIOLINO P., HUNG T.V., TISSERON S. VIAUD-DELMON I., DAVIDOVICI M., ATTIA D. & BAYEUX T. (2021), « Expositions aux technologies de réalité virtuelle et/ou augmentée », Paris, Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail.

CNPEN (2024), « Avis n°9 Métavers : enjeux d'éthique », Paris, Comité national pilote d'éthique du numérique.

KALKOFEN D., SANDOR. C, WHITE S. & SCHMALSTIEG D. (2011), "Visualization techniques for augmented reality", *In: Handbook of Augmented Reality*, pp. 65-68.

MORI M., MACDORMAN K. & KAGEKI N. (2012), "The uncanny valley" (traduction), *IEEE Robotics & Automation Magazine*, 19(2), pp. 98-100.

PAULUS Y. T. & REMIJN G. B. (2021), "Usability of various dwell times for eye-gaze-based object selection with eye tracking", *Displays*, 67.

PECK T.C., SEINFELD S., AGLIOTI S.M. & SLATER M. (2013), "Putting yourself in the skin of a black avatar reduces implicit racial bias", *Consciousness and Cognition*, 22(3), pp. 779-787.

PEILLARD E. (2020), *Vers une caractérisation des biais perceptifs en réalité mixte : une étude de facteurs altérant la perception des distances*, thèse de doctorat, École Centrale de Nantes.

SILVA FREITAS J.R., SILVA VELOSA V.H., NUNES ABREU L.T., JARDIM R.L., VIEIRA SANTOS J.A., PERES B. & CAMPOS P.F. (2021), "Virtual reality exposure treatment in phobias: a systematic review", *Psychiatric Quarterly*, 92, pp. 1685-1710.