

Le système EVE : aperçu des recherches en cours en immersion collaborative et en téléopération

Pierre Cazaux, Jean-Marc Vézien, Pierre Martin, Damien Touraine, Nicolas Férey, Patrick Bourdot.

Groupe de RV&A "VENISE" - CNRS/LIMSI - Université Paris-Sud - Orsay.

<http://www.limsi.fr/venise/>

ABSTRACT

Le système EVE, grand dispositif immersif, multi sensorimoteur, collaboratif et reconfigurable du groupe VENISE du CNRS/LIMSI, est entré en service début 2010. Nous décrivons ses caractéristiques technologiques les plus originales, puis présentons deux des principaux axes de recherche que l'équipe développe sur cet équipement, à savoir : l'Immersion Multimodale et Collaborative, et la Téléopération à base de Virtualité Augmentée.

KEYWORDS: Réalité Virtuelle et Augmentée, Haptique, Multimodalité, Stéréoscopie Multi-utilisateurs, Interaction Collaborative, Téléopération, Simulation.

1 INTRODUCTION

Deux des problématiques de recherche les plus passionnantes dans l'usage des technologies de la Réalité Virtuelle et Augmentée, sont probablement celles du travail collaboratif et de la télé-présence. C'est typiquement ce genre de thématiques que permet d'étudier le **système EVE** (acronyme de **E**volutive **V**irtual **E**nvironment) qui a été spécifié par le groupe VENISE du CNRS/LIMSI, puis réalisé fin 2009 par les compagnies Barco [1], Haption [2] & ART.GmbH [3]. Principalement financé par le CNRS et le RTRA Digiteo [4], le système EVE intègre de nombreuses caractéristiques originales dont la combinaison est actuellement unique au monde.

Après une description de ses caractéristiques, nous présentons deux aperçus sur les recherches que ce système nous a déjà permis de lancer. Le premier concerne l'extension de nos modèles de gestion des interactions multimodales en situation immersive à des activités collaboratives co-localisées. Le second éclairage porte sur nos travaux en Virtualité Augmentée visant des télé-opérations multi-sensorielles. Chacun de ces aperçus s'achève par la description du démonstrateur associé, démonstrateurs qui furent présentés lors du meeting européen EuroVR-EVE 2010 [5] et durant ces 5èmes journées de l'AFRV [6]. Nous concluons en pointant les travaux d'évaluation en cours en matière de travail collaboratif co-localisé et de télé-présence.

2 LE SYSTEME EVE

EVE est un grand équipement immersif, multi sensorimoteur, collaboratif et reconfigurable dont les principales caractéristiques sont :

- Dalle rétro projetée de 13 m² entourée d'écrans de 4,70 mètres de haut (dont 1 mobile à terme).
- Technologie de visualisation temps réel, reconfigurable en haute définition (pixel de 1,2 mm) ou en double

stéréoscopie (relief exact pour 2 groupes d'utilisateurs n'ayant pas le même point de vue).

- Rendus audio 3D, avec solutions ambisoniques ou binaurales, associées à des contraintes acoustiques drastiques sur les projecteurs et la salle accueillant le dispositif.
- Divers périphériques haptiques permettant des sensations tactiles (gants) et/ou des retours d'effort (bras 3 DoF et 6 DoF).



Figure 1. Le système EVE dans sa configuration « immersion complète ». L'immersion est assurée à la fois par la dalle rétro projetée, et à la hauteur des écrans verticaux.

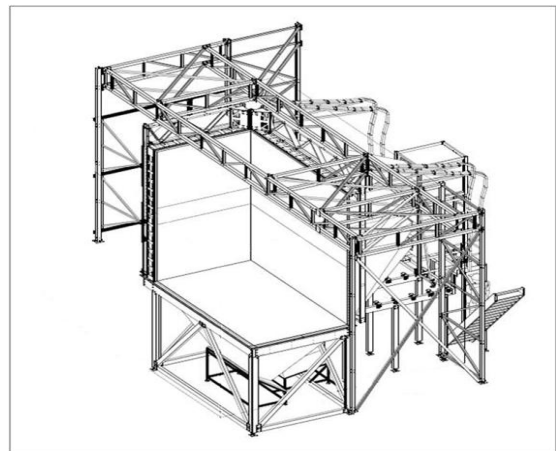


Figure 2. La structure autoportée du système EVE

Il s'agit donc d'un équipement scientifique qui a été conçu pour permettre d'expérimenter le futur de la Réalité Virtuelle et Augmentée. Pour l'heure, les recherches se concentrent sur la mise au point de solutions originales pour développer une nouvelle génération d'applications 3D : des applications immersives, multi-utilisateurs (interaction collaborative « co-localisée » ou « distante »), aux interfaces intuitives voire « naturelles » et dotées d'intelligence (cf. interaction multimodale), via six principaux canaux : parole, capture de mouvements, analyse de gestes, feedbacks tactilo-kinesthésiques, audio 3D, relief visuel.

3 INTERACTION MULTIMODALE ET COLLABORATIVE EN IMMERSION CO-LOCALISEE

La configuration double-stéréoscopie du système EVE permet en particulier de mener des études comparatives entre plusieurs types de collaborations immersives : celles dites distantes par rapport à celles qui sont co-localisées, mais aussi au sein de ces dernières les apports et limites de la multi-stéréoscopie. Ces études se basent sur les précédents travaux de l'équipe sur la multimodalité en RV.

3.1 Etat de l'art

Depuis Bolt [7] et son paradigme du « Put that there », la multimodalité a été un domaine de recherche grandissant en Informatique. Ainsi, depuis plusieurs années, l'équipe VENISE mène des recherches sur la multimodalité en Environnement Virtuel. Le challenge des interfaces multimodales est d'analyser et d'interpréter en temps-réel des informations provenant de plusieurs canaux sensorimoteurs. Martin [8] a établi une classification des interactions multimodales : *transfert* (les informations produites par une modalité sont utilisées par une autre modalité), *équivalence* (les informations peuvent être produites par plusieurs modalités), *spécialisation* (une information spécifique est toujours produite par la même modalité), *redondance* (une même information est produite par plusieurs modalités) et *complémentarité* (différentes informations sont produites par différentes modalités mais doivent être combinées pour être complètes). Cependant, les systèmes de RV intégrant des signaux continus (tracking, gestes, etc.), ils produisent des événements asynchrones qui requièrent une *co-référence passive* (cf. Bellik [9]). Une des caractéristiques fondamentales du système multimodal conçu par le groupe VENISE est de pouvoir s'adapter à des plateformes logicielles synchrones, type Virtools, bien que le processus multimodal soit asynchrone (propriété nécessaire pour la gestion d'événements continus telles les reconnaissances vocale et gestuelle ou la capture de mouvements).

Oviatt et al. [10] ont proposé une classification des architectures pour les interactions multimodales : *fusion au plus tôt* et *fusion au plus tard*. Par *fusion au plus tôt* il faut comprendre que le système intègre immédiatement les résultats des divers processus de reconnaissance, sans traitement au préalable, ce qui est généralement approprié pour les modalités fortement couplées et synchronisées (geste et parole, par exemple). A l'inverse, un système de *fusion au plus tard*, va intégrer les informations ayant un haut niveau sémantique, donc va permettre d'exploiter en direct des systèmes de reconnaissance existants et faciliter du même coup l'extension et/ou la reconfigurabilité de l'interface. Bien qu'il existe de nombreuses autres plateformes pour la gestion multimodale, peu de ces systèmes intègrent la notion de co-référence passive signalée précédemment. Certains travaux s'intéressent aux caractéristiques requises pour la conception et l'implémentation de systèmes multimodaux (Ladry et al [11],

Mendonça et al [12]), d'autres présentent des revues complètes sur les interfaces multimodales (Dumas et al. [13], Lalanne et al. [14]). Ceci étant, les principales plateformes existantes destinées aux applications immersives sont celles de Flippo et al. [15], Sun et al. [16], Krahnstoeber et al. [17], Kaiser et al. [18], Pfeiffer et Latoschik [19] et Latoschik [20].

D'un autre côté, les premiers travaux sur les systèmes immersifs à utilisateurs multiples datent de 1997 pour des interactions en vis-à-vis sur des Workbench [21], puis pour des interactions dans des systèmes multi-écrans [22] ou CAVE [23]. Dans ces différents prototypes la séparation des utilisateurs restait de très mauvaise qualité. En fait, Fröhlich et al. [24] ont démontré en 2005 que la séparation active (obturateur ou shutter glass) des utilisateurs et la séparation passive (polarisation circulaire) pour perception binoculaire de la profondeur, est l'approche la plus prometteuse. Ces systèmes, généralement développés en collaboration étroite avec des industriels du domaine (tels que Barco, Infitec ou RealD) permettent d'ores-et-déjà à deux utilisateurs d'avoir chacun une stéréoscopie exacte tout en étant physiquement co-localisés dans le dispositif immersif. De plus cette technologie est extensible à plusieurs utilisateurs. Ainsi, dans le cadre d'un projet européen, l'équipe de B. Fröhlich est en train de mettre au point un prototype multiutilisateur. Au demeurant, sa solution n'est pas optimale en terme acoustique. Les projecteurs doivent être isolés dans des pièces spécifiques et séparés de la CAVE par un écran rigide, chose peu compatible avec la reconfigurabilité et les solutions ambisoniques dont voulait disposer l'équipe VENISE dans le système EVE.

D'un point de vue conceptuel, trois niveaux d'interaction collaborative ont été identifiés par Margery et al. [25] : coopération de base, tâche coopérative, manipulation coopérative. Pour permettre la coopération la plus élevée (niveau 3), des règles doivent être définies pour gérer les accès des utilisateurs. Quelques travaux ont été menés sur cette question par les mêmes auteurs, ou par exemple par Pinho et al. [26].

3.2 Supervision des interactions Multimodales et Collaboratives

Les travaux sur la multimodalité en Environnements Virtuels, initiés notamment dans le cadre des projets ANR Perf-RV et Perf-RV2, ont abouti à la création d'un moteur de fusion multimodale qui vise à transformer les événements issus de différents types de capteurs, donc de l'être humain, en information compréhensible par l'application. De façon générale, les événements des différentes reconnaissances (vocale et gestuelle) aussi bien que d'autres périphériques (tracking, flystick) sont injectés dans le moteur de fusion, qui les interprète et les transforme via une *fusion au plus tard*, selon des règles établies. L'interprétation des événements en entrée vise à les rendre indépendants du périphérique dont ils sont issus et à les traduire dans le contexte de l'application. Ces règles sont codées dans un fichier de configuration XML, et décrivent les interactions potentielles pouvant être générées par le moteur. Les commandes qui sortent du moteur sont elles transcrites dans un format adapté puis envoyées à l'application.

La mise en service du système EVE nous a permis de poursuivre et d'étendre ces travaux. De nouvelles modalités ont pu être intégrées au processus de fusion multimodale, en particulier l'haptique, dont les événements sont issus d'un bras à retour d'effort. Le système EVE nous permet de traiter les aspects propres aux CVE (Collaborative Virtual Environment) et également de valider la capacité de ce moteur de fusion

multimodale à gérer des situations multiutilisateur. Bien que faisant parti des spécifications initiales du moteur, la représentation multiutilisateur des événements, permettant donc à plusieurs personnes de réaliser des interactions multimodales, indépendamment ou non les uns des autres, n'avait ni été mise en place ni testée. Dès lors que plusieurs utilisateurs peuvent interagir de manière concurrente et par l'intermédiaire d'un même système, la notion de collaboration surgit. Ainsi le moteur rend possible la combinaison des actions des utilisateurs à différentes

étapes du processus de fusion (voir Fig. 2). Nos recherches sur les interactions multimodales et collaboratives se sont naturellement basées sur les travaux de Margery et al. [25] Il nous est rapidement apparu des similarités entre les processus multimodaux et collaboratifs, ce qui nous a amené à étendre la classification de proposée par ces auteurs, en ayant une approche centrée plutôt sur la notion de tâche/commande, que sur la notion de manipulation.

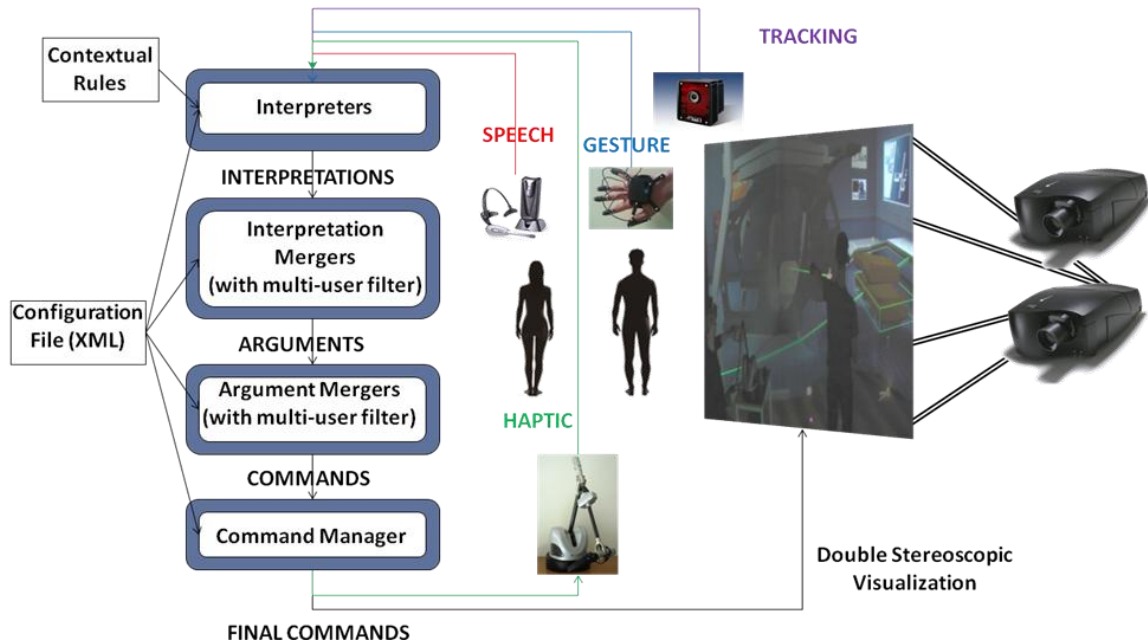


Figure 3. Architecture du système Multimodal et Collaboratif

3.3 Exemple de scénario d'usage dans l'industrie automobile

Ce démonstrateur donne un aperçu de l'état actuel de la plateforme logicielle développée dans l'objectif de cette étude, mais présente aussi à notre connaissance le premier système immersif gérant de façon intégrée les interactions multimodales et collaboratives.

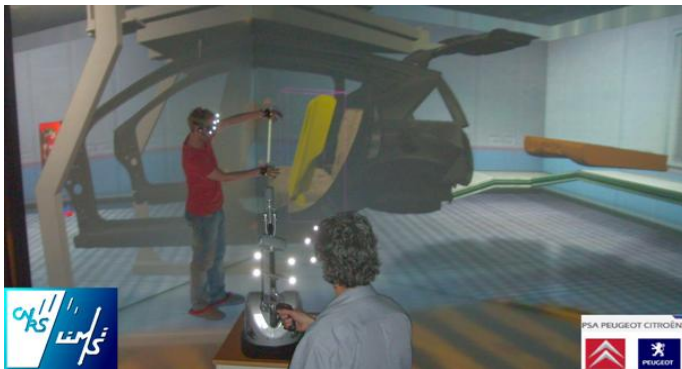


Figure 4. Tâche de positionnement d'un siège dans un véhicule dans un contexte multi-stéréoscopique (2 utilisateurs) et multimodal (haptique, voix et geste à 2 mains).

Du fait du contexte industriel, le moteur de supervision des interactions Multimodales et Collaboratives a été embarqué dans des composants Virtools. Une configuration spécifique du VRPack a été développée de sorte à pouvoir exploiter la double stéréoscopie du système EVE. Sont également utilisés, le système de reconnaissance gestuelle développé en interne et la reconnaissance vocale de Vecsys. La partie simulation physique est assurée par IPP (Interactive Physical Pack) de Haption, plugin intégré à Virtools, les manipulations haptiques se faisant via un Virtuose 6DoF.

Le scénario de la démonstration est une tâche de collaboration de deux utilisateurs pour des activités de formation ou de conception dans l'industrie automobile. Dans une chaîne d'assemblage virtuelle, les utilisateurs coopèrent pour définir la trajectoire d'un siège qui doit être inséré dans l'habitacle d'une voiture. Par exemple, un apprenti manœuvre un siège en utilisant les commandes vocales et le dispositif haptique afin d'analyser les trajectoires d'insertion possibles. Dans le même temps un tuteur contraint la tâche de l'apprenti en définissant des guides virtuels (par exemple, des axes ou des plans) à l'aide de commandes vocales et/ou gestuelles. Par ailleurs, la plateforme logicielle peut être utilisée dans une tâche de conception collaborative. Dans ce cas, le système gère la commutation d'un utilisateur à l'autre en détectant celui qui tient l'organe terminal du dispositif haptique. De plus, l'utilisateur du périphérique haptique peut définir des guides virtuels additionnels avec des commandes vocales et des gestes réalisés avec sa main libre. Ce démonstrateur a été réalisé

dans le cadre du projet Digiteo « SIMCoD » [4] en partenariat avec le CEA-LIST, Haption et PSA Peugeot Citroën.

3.4 Perspectives

Plusieurs types d'évaluation du système EVE et de ces travaux sur l'interaction multimodale et collaborative en immersion co-localisée doivent maintenant être menés : l'évaluation de la double stéréoscopie en elle-même, des études sur la perception métrique dans un environnement immersif disposant d'une dalle rétro-projetée (avec ou sans situation collaborative), et enfin l'évaluation de tâches collaborative dans une situation co-localisée (complexité de tâches, etc.).

De plus, la connexion future du système EVE du groupe VENISE avec la plateforme du CEA-LIST, va permettre de comparer ces deux classes de collaboration (distante et/ou co-localisée) dans un contexte immersif multiutilisateur (double stéréoscopie, multi audio), multi-sensorimoteur (parole, geste, haptique, audio 3D et multi-stéréoscopie), voire multimodale (supervision intelligent des canaux sensorimoteurs).

4 TELEPRESENCE ET VIRTUALITE AUGMENTEE

Un travail de recherche mené depuis quelques années au sein de l'équipe VENISE mobilise les concepts de Téléprésence et Virtualité Augmentée, avec pour objectif le pilotage à distance dans un contexte immersif. La notion de Virtualité Augmentée, symétrique à celle de Réalité Augmentée, consiste en la superposition d'un contenu réel capturé à distance dans une simulation de Réalité Virtuelle. L'enjeu de la Téléprésence immersive est l'occultation aussi complète que possible des éléments technologiques qui permettent de plonger un utilisateur dans une simulation numérique afin que réel et virtuel ne fassent plus qu'un, d'un point de vue perceptif.

4.1 Contexte et problématique

L'objectif du projet est de piloter un véhicule à distance grâce à un dispositif immersif. Ce dispositif recrée fidèlement l'environnement du véhicule en combinant, au sein d'une réalité mixte, des éléments réels (flux audio et vidéo) et virtuels (modélisations 3D). L'expérimentateur a ainsi l'impression de se trouver à la place du conducteur du véhicule, et il peut alors conduire sans être confronté à une interface informatique classique.

Dans ce cadre, il est nécessaire de s'assurer qu'un tel dispositif retranscrit le monde distant de façon satisfaisante, au sein d'une description virtuelle de l'environnement 3D du véhicule, vis-à-vis du processus de conduite. Par exemple, le pilote doit être capable de percevoir en temps réel ce qui se passe dans l'environnement du véhicule (obstacles, déformations de la chaussée, traversée d'un piéton etc.). Il est donc nécessaire de connaître avec précision la position du véhicule et des différents éléments de l'environnement, ce qui nécessite un recalage assez fin entre le réel et virtuel.

4.2 Téléimmersion : Etat de l'art

L'approche scientifique en termes de téléprésence en RV&A passe très souvent par l'apport d'une nouvelle modalité à un système existant (nouveau capteur, visualisation d'information originale), l'ajout d'un nouvel élément étant censé améliorer à chaque fois l'immersion ou les performances en termes de tâches à accomplir. Les systèmes concernés peuvent être supervisés à distance ou directement conduits. Dans le cas du projet VIRTUE [27], l'ajout d'une caméra stéréoscopique montée sur une tourelle de type Pan-Tilt renforce l'immersion du sujet sur un site de

construction, permettant ainsi à ce dernier d'être équipé d'une vue active (l'utilisateur et le robot regardent constamment au même endroit). Pettey et al. [28] mesure l'augmentation de l'immersion d'une projection vidéo par l'ajout du son ; Katz et al. [29] discutent l'apport d'un rendu sonore binaural et ambisonique dans un système de téléprésence et concluent pour un meilleur rendu sonore avec le système binaural. L'amélioration de l'efficacité sur une tâche téléopérée, lorsque l'on superpose un modèle virtuel sur un flux vidéo réel, est étudiée par Milgram [30].

En d'autres termes, les avancées technologiques et scientifiques dans le domaine des transmissions et du contrôle à distance ont sans cesse contribué à rendre meilleurs les dispositifs de téléprésence. Nous avons donc cherché à mettre en place un protocole expérimental qui permettrait de mettre en évidence la notion de téléprésence d'une part, et d'autre part de la mesurer. Pour cela nous nous appuyons sur plusieurs travaux élaborant des méthodes de mesure de présence. Le psychologue J.J. Gibson a introduit en 1977 le concept d'affordance : il s'agit de l'ensemble de toutes les possibilités d'action offerte par un environnement à un utilisateur. L'affordance d'un objet serait alors toutes les possibilités d'action qu'un utilisateur peut effectuer avec cet objet. L'expérience menée par Lepecq et al. [31] met en avant l'importance de l'affordance dans la mesure de la présence en RV : il est demandé à une vingtaine de sujets de passer d'une pièce virtuelle (dans un dispositif de RV) à une autre, le passage entre les deux pièces étant plus ou moins étroit. Il n'est pas précisé aux sujets de ne pas toucher le mur virtuel. La grande majorité des sujets tourment spontanément leurs épaules pour franchir l'obstacle (de façon similaire à la situation réelle), ce qui permet de quantifier dans une certaine mesure la sensation de présence.

Etant donné que la téléprésence est du domaine du ressenti, il n'existe pas de méthode autre qu'empirique pour son évaluation. De nombreux travaux de recherche utilisent pour cela des questionnaires soumis aux sujets d'une expérience. Les travaux de Lessiter et al. [32] concernant le questionnaire ITC-SOPI comprennent 4 échelles de mesure de la présence. On trouve la perception de l'espace, l'implication, la véracité environnementale et les effets négatifs liés à la technologie. Witmer et al. [33] ont une approche différente mais utilisent également des questionnaires comme moyen d'évaluation. Ils utilisent deux questionnaires pour d'une part, évaluer la présence et d'autre part afin de mesurer l'implication qu'un utilisateur peut avoir dans ses activités quotidiennes. Ce dernier questionnaire permet alors d'affiner le premier.

Cependant, le travail de Usuh et al. [34] a montré les limites de tels questionnaires. Les résultats de leurs expériences montrent que l'usage de questionnaires seuls dans des cas de comparaison réel/virtuel est statistiquement non-significatif.

Une approche originale a été proposée par Rétaux [35] en 2002, par l'évaluation de la sensation de présence en fonction de l'affinité du sujet avec la tâche. Cette expérience confirme notre analyse car elle met en évidence le fait que l'implication d'un sujet est variable en fonction de son affinité, ce qui renforce notre idée de l'utilisation de l'affordance pour l'évaluation.

Enfin, l'expérience conduite par Slater et al. [36] consiste à placer un utilisateur en immersion dans un dispositif en lui demandant de signaler explicitement les instants où il perd la sensation de présence. Les résultats sont analysés statistiquement pour construire un modèle de Markov des transitions entre

présence dans le réel et présence en immersion. Initialement, Slater et al. font l'hypothèse que le fait de rapporter la perte de la présence n'influe pas sur la sensation de présence. L'expérience montre cependant le contraire, mettant en doute le modèle employé. Ils concluent donc qu'il serait intéressant de trouver une solution alternative pour mesurer la perte de présence.

4.3 Application au pilotage immersif à distance

Les travaux de téléprésence immersive sont conduits au LIMSI sur une plateforme associant le dispositif EVE à un fauteuil roulant électrique muni de capteurs stéréoscopiques et d'un microphone 3D (voir Fig. 4).

Avec le démonstrateur associé à ces travaux, un utilisateur peut conduire à distance le véhicule robotisé grâce à un canal audiovisuel temps-réel 3D. Les qualités acoustiques uniques du système EVE et de sa halle permettent une télé-immersion sonore de type ambisonique. Ces données réelles seront recalées à terme sur une description virtuelle de l'environnement 3D du véhicule afin de faciliter le processus de conduite.



Figure 5. Véhicule robotisé utilisé dans l'expérience de télé-immersion.

L'expérimentation de téléprésence dans ce système en est encore à sa phase initiale. Il est maintenant nécessaire d'évaluer si l'immersion permise par le système EVE augmente efficacement la téléopération. La validation sur un panel de sujets sera effectuée prochainement en espace extérieur, sur un circuit aménagé. L'analyse des résultats comprendra un examen croisé des questionnaires en corrélation avec les résultats numériques expérimentaux mesurant la performance sur la tâche de conduite (évaluation des temps de parcours, régularité des trajectoires). Ainsi, on pourra juger du caractère statistiquement significatif des liens entre présence perçue et présence « réalisée ». Par ailleurs, les affordances spécifiques des éléments du circuit pourront être comparées dans les différentes situations expérimentales, afin de constater, par exemple, si un comportement d'exploration en situation de conduite réelle se retrouve ou non en situation de téléconduite, pour un sujet donné. Au regard de ces évaluations, les modalités de pilotage du système à distance pourront être affinées afin d'obtenir des indices de qualité plus grands, tant au niveau perceptif qu'au niveau quantitatif.

4.4 Perspectives

Dans un futur proche, nous souhaitons également développer l'autonomie du véhicule distant, donnant ainsi suite aux travaux

de Tarault et al. [37] dans le projet SACARI (Supervision of an Autonomous Car in an Augmented Reality Interface), dans une logique de supervision du système. Ainsi l'utilisateur pourra indiquer, en le pointant dans l'image, l'endroit où il désire se rendre, les modules de conduite embarquée se chargeant de calculer la trajectoire pour s'y rendre et les commandes pour la réaliser. Cette approche à base de commandes nous permettra également d'intégrer d'autres modalités d'entrée au système de téléprésence, comme la voix ou l'interaction haptique, qui se révéleront essentielles pour des tâches d'intervention plus complexes sur l'environnement.

Nous souhaitons enfin développer des modules de Réalité Augmentée, visant à améliorer encore la présence par l'ajout d'informations utiles à la tâche, comme la détection de piétons ou d'obstacles, ou la reconnaissance d'amers utiles à la navigation. Là encore, une attention particulière sera portée sur la mesure de la pertinence des éléments d'augmentation, ainsi que sur les modalités précises de leur mise en œuvre. Les problématiques de recalage 3D entre éléments virtuels et réels sont bien entendues aussi essentielles pour une tâche de Réalité Mixte. Le calcul de la position du véhicule téléopéré dans son environnement fera probablement appel à des techniques de recalage mixte combinant le traitement du signal visuel avec d'autres capteurs de position (inertiel, GPS).



Figure 6. L'interface de conduite télé-immersive SACARI

5 CONCLUSION

Dans ce papier nous avons présenté les principales caractéristiques du système EVE qui font de cet équipement un outil de recherche scientifique sans équivalent, à notre connaissance. Nous avons ensuite donné deux aperçus sur les axes de recherches engagés.

Le premier concerne l'extension de nos modèles de gestion des interactions multimodales en situation immersive à des activités collaboratives co-localisées. Outre l'usage d'une technologie d'immersion multiutilisateur qu'offre le système EVE, nos travaux tendent à prouver que la supervision multimodale en situation immersive est un processus quasiment identique à celui visant la gestion des interactions collaboratives co-localisée.

Le second éclairage concerne l'évolution de notre interface de Virtualité Augmentée visant des téléopérations multisensorielles. Là encore, au-delà de la performance technologique d'immersion visuelle et sonore que permet le système EVE, le dispositif expérimental de télé-conduite immersive construit autour de cette interface va permettre de caractériser et de mieux comprendre la notion de télé-présence.

REFERENCES

- [1] <http://www.barco.com/>
- [2] <http://www.haption.com/>
- [3] <http://www.ar-tracking.de/>
- [4] <http://www.digiteo.fr/>
- [5] <http://eurovr-eve-2010.limsi.fr/>
- [6] <http://afrv2010.limsi.fr/>
- [7] R. A. Bolt. "put-that-there": Voice and gesture at the graphics interface. In SIGGRAPH '80: Proc. of the 7th annual conference on Computer graphics and interactive techniques, pages 262–270, New York, NY, USA, 1980. ACM.
- [8] J.-C. Martin. Tycoon: Theoretical framework and software tools for multimodal interfaces. In John Lee (Ed.), *Intelligence and Multimodality in Multimedia Interfaces*. AAAI Press, 1998.
- [9] Y. Bellik. Media integration in multimodal interfaces. In Proc. of the IEEE Workshop on Multimedia Signal Processing, pages 31–36, 1997.
- [10] S. Oviatt, P. Cohen, L. Wu, J. Vergo, L. Duncan, B. Suhm, J. Bers, T. Holzman, T. Winograd, J. Landay, J. Larson, and D. Ferro. Designing the user interface for multimodal speech and pen-based gesture applications: state-of-the-art systems and future research directions. *Hum.-Comput. Interact.*, 15(4):263–322, 2000.
- [11] J.-F. Ladry, D. Navarre, and p. Palanque. Formal description techniques to support the design, construction and evaluation of fusion engines for sure (safe, usable, reliable and evolvable) multimodal interfaces. In ICMI-MLMI '09: Proc. of the 2009 international conference on Multimodal interfaces, pages 185–192, New York, NY, USA, 2009. ACM.
- [12] H. Mendonc,a, J.-Y. L. Lawson, O. Vybornova, B. Macq, and J. Vanderdonck. A fusion framework for multimodal interactive applications. In ICMI-MLMI '09: Proc. of the 2009 international conference on Multimodal interfaces, pages 161–168, New York, NY, USA, 2009. ACM.
- [13] B. Dumas, D. Lalanne, and S. Oviatt. Multimodal interfaces: A survey of principles, models and frameworks. pages 3–26, 2009.
- [14] D. Lalanne, L. Nigay, p. Palanque, P. Robinson, J. Vanderdonck, and J.-F. Ladry. Fusion engines for multimodal input: a survey. In ICMI-MLMI '09: Proc. of the 2009 international conference on Multimodal interfaces, pages 153–160, New York, NY, USA, 2009. ACM.
- [15] F. Flippo, A. Krebs, and I. Marsic. A framework for rapid development of multimodal interfaces. In ICMI '03: Proc. of the 5th international conference on Multimodal interfaces, pages 109–116, New York, NY, USA, 2003. ACM.
- [16] Y. Sun, F. Chen, Y. D. Shi, and V. Chung. A novel method for multisensory data fusion in multimodal human computer interaction. In OZCHI '06: Proc. of the 18th Australia conference on Computer-Human Interaction, pages 401–404, New York, NY, USA, 2006. ACM.
- [17] N. Krahnstoever, S. Kettebekov, M. Yeasin, and R. Sharma. A realtime framework for natural multimodal interaction with large screen displays. In ICMI '02: Proc. of the 4th IEEE International Conference on Multimodal Interfaces, page 349, Washington, DC, USA, 2002. IEEE Computer Society.
- [18] E. Kaiser, A. Olwal, D. McGee, H. Benko, A. Corradini, X. Li, P. Cohen, and S. Feiner. Mutual disambiguation of 3d multimodal interaction in augmented and virtual reality. In ICMI '03: Proc. of the 5th international conference on Multimodal interfaces, pages 12–19, New York, NY, USA, 2003. ACM.
- [19] T. Pfeiffer and M. E. Latoschik. Resolving object references in multimodal dialogues for immersive virtual environments. In VR '04: Proc. of the IEEE Virtual Reality 2004, page 35, Washington, DC, USA, 2004. IEEE Computer Society.
- [20] M. E. Latoschik. A user interface framework for multimodal vr interactions. In ICMI '05: Proc. of the 7th international conference on Multimodal interfaces, pages 76–83, New York, NY, USA, 2005. ACM.
- [21] M. Agrawala, A. C. Beers, I. McDowall, B. Fröhlich, M. Bolas, and P. Hanrahan. The two-user responsive workbench: support for collaboration through individual views of a shared space. In SIGGRAPH '97: Proc. of the 24th annual conference on Computer graphics and interactive techniques, pages 327–332, New York, NY, USA, 1997. ACM Press/Addison-Wesley Publishing Co.
- [22] K. Arthur, T. Preston, R. Taylor, F. Brooks, M. Whitton, and W. Wright. Designing and building the pit: a head-tracked stereo workspace for two users. In Proc. of 2nd International Immersive Projection Technology Workshop, 1998.
- [23] K. Blom, G. Lindahl, and C. Cruz-Neira. Multiple active viewers in projection-based immersive environments. *Immersive Projection Technology Workshop*, 2002.
- [24] B. Fröhlich, R. Blach, O. Stefani, J. Hochstrate, J. Hoffmann, K. Klüger, and M. Bues. Implementing multi-viewer stereo displays. In WSCG (Full Papers), pages 139–146, 2005.
- [25] D. Margery, B. Arnaldi, and N. Plouzeau. A general framework for cooperative manipulation in virtual environments. In *Virtual Environments*, volume 99, pages 169–178, 1999.
- [26] M. S. Pinho, D. A. Bowman, and C. M. Freitas. Cooperative object manipulation in immersive virtual environments: framework and techniques. In VRST '02: Proc. of the ACM symposium on Virtual reality software and technology, pages 171–178, New York, NY, USA, 2002. ACM.
- [27] A. Retik, G. Mair, R. Fryer, and D. McGregor, « Integrating Virtual Reality and Telepresence to Remotely Monitor Construction Sites : A VIRTUE Project », *Lecture Notes in Computer Science*, 1998, pp 459-453.
- [28] G. Pettey, C. Campanella Bracken, B. Rubenking, M. Buncher and E/ Gress, « Telepresence, soundscapes and technological expectation : putting the observer into the equation », *Virtual Reality*, mars 2010, pp 15-25.
- [29] B.F.G. Katz, A. Tarault, P. Bourdot et J.M. Vézien, « The Use of 3D-Audio in a Multi-Modal Teleoperation Platform for Remote Driving/Supervision », 30th International Conference: Intelligent Audio Environments, mars 2007.
- [30] P. Milgram, S. Yin and J.J.Grodski, « An Augmented Reality Based Teleoperation Interface for Unstructured Environments », 7th Topical Meeting on Robotics and Remote Systems, May 1997, pp 966-973.
- [31] J-C. Lepecq, L. Bringoux, J-M Pergandi, T. Coyle et D. Mestre, « Afforded Actions as a Behavioral Assessment of physical Presence », *Presence*, Vol. 7, No. 3, June 1998, pp 202 – 208.
- [32] J. Lessiter & J. Freeman, « Presence – A global media quality metric, « Immersive Projection Technology and Virtual Environments », 2001, pp 23-30.
- [33] B. G. Witmer et M. J. Singer. « Measuring Presence in Virtual Environments: A Presence Questionnaire », *Presence*, Vol. 7, No. 3, June 1998, pp 225–240.
- [34] M. Usoh, E. Catena, S. Arman et M. Slater, « Using Presence Questionnaires in Reality », *Presence*, Vol. 9, No. 5, Octobre 2000, pp 497-503
- [35] X. Rétaux, « A subjective measure of presence feeling : the "autoconfrontation" method », Fifth Annual International Workshop on Presence, 2002.
- [36] M. Slater, A. Steed, M. Usoh et R. Schroeder, « The Influence of Body Movement on Subjective Presence in Virtual Environments », *Presence: Teleoperators and Virtual Environments*, Vol. 9, pp. 37-51.
- [37] A. Tarault, P. Bourdot et J.M. Vézien « SACARI : An Immersive Remote Driving Interface for Autonomous Vehicles » *Lecture Notes in Computer Science*, 2005, pp 339-342.