Une neuroprothèse visuelle interactive pour mieux percevoir l'environnement

Julien Desvergnes, Axel Carlier, Vincent Charvillat, Christophe Jouffrais

Une neuroprothèse visuelle interactive pour mieux percevoir l'environnement

Julien Desvergnes
IRIT REVA - CNRS - INPT - IPAL
Toulouse, France

Axel Carlier
IRIT REVA - INPT - IPAL
Toulouse, France

Vincent Charvillat
IRIT REVA - INPT - IPAL
Toulouse, France

Christophe Jouffrais
IRIT ELIPSE - CNRS - IPAL
Toulouse, France

***Résumé*—Les neuroprothèses visuelles sont des dispositifs qui permettent de rétablir une perception visuelle limitée chez les patients non-voyants implantés. Certaines de ces neuroprothèses, implantées dans la rétine ou dans le cortex visuel, comprennent un implant, un dispositif de calcul informatique et une caméra externe pour capturer la scène. Une perception visuelle appauvrie est alors restaurée lors d'une micro-stimulation de la rétine ou du cortex visuel via l'implant, ce qui conduit à l'apparition de points blancs appelés phosphènes. Toutefois, la résolution des implants actuels (c'est-à-dire le nombre d'électrodes et leurs espacements) qui ont passé les phases d'essais cliniques reste faible. Cette faible résolution, associée au nombre limité de couleurs différentes rendues par les implants, limite les informations qui peuvent être transférées et donc perçues. Le processus de rendu utilisé par défaut sur les implants, appelé scoreboard, est insuffisant pour comprendre la scène visuelle. Pour permettre aux patients de mieux percevoir leurs environnements, les recherches en cours visent à maximiser la qualité et la quantité d'informations fournies par l'implant. Nous avons mis en place une étude comparative entre différents rendus et nous avons montré que le fait de donner aux non-voyants la possibilité d'alterner entre plusieurs modes de rendu en temps réel augmente significativement la compréhension de l'environnement.**

***Mots clés*—Neuroprothèses visuelles, rendu interactif, rendu adaptatif, implant rétinien, vision par ordinateur, non-voyants**

I. Introduction

Selon l'OMS [1], 253 millions de personnes souffrent de déficience visuelle : 36 millions d'entre elles sont aveugles et 217 millions ont une déficience visuelle entre modérée et sévère. Les personnes avec déficiences visuelles utilisent traditionnellement une canne ou un chien-guide pour se déplacer. Mais les cannes et les chiens-guides ont un rôle limité dans la compréhension de la scène visuelle.

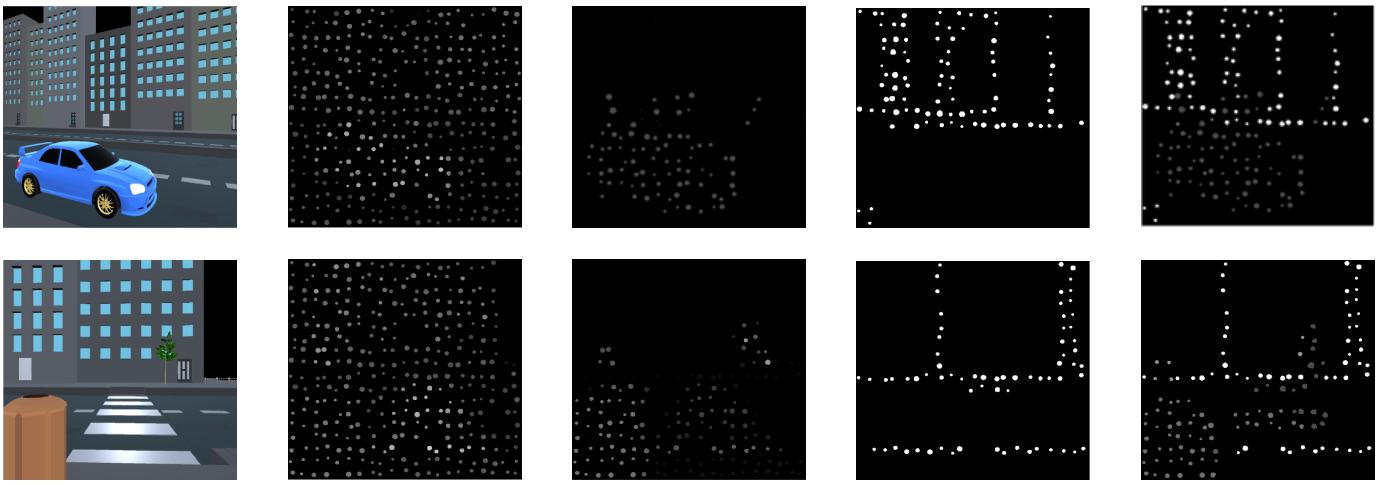
Les neuroprothèses visuelles sont apparues dans les années 60 [2] et au cours de la dernière décennie, elles sont apparues comme une technique prometteuse pour restaurer partiellement la vision chez les personnes atteintes de déficience visuelle [3]. Il existe à ce jour deux types de dispositif. Le premier est un dispositif implanté dans la rétine et basé sur des photorécepteurs, appelé MPDA pour microphotodiode-array. Les photons frappent ces derniers et stimulent alors les neurones sous-jacents (exemple avec l'implant ALPHA IMS [4]). Dans le second dispositif, une caméra externe capture la scène et envoie des signaux vers un implant situé dans la rétine ou le cortex pour restaurer des perceptions visuelles. Ce type d'implant peut être épirétinien (sur la rétine), sous-rétinien (juste en dessous de la rétine) ou encore supra-choroïdien (entre la choroïde et la sclérotique). Depuis les dix dernières années, plusieurs implants ont été placés sur des non-voyants et sont en phase d'essais cliniques [3]. Les neuroprothèses utilisant ce genre d'implants ne sont toutefois pas totalement fonctionnelles en raison d'une résolution limitée, c'est-à-dire du faible nombre d'électrodes utilisées (le système PRIMA en comptait 378 lors d'un essai clinique en 2020 [5]). Une forte limitation des niveaux d'intensité lumineuse applicables sur les phosphènes rend la tâche encore plus complexe. De plus, le bruit (imperfection dans le placement de la matrice d'électrodes, dysfonctionnement des électrodes) dans l'affichage des phosphènes peut également poser des difficultés de perception.

Dans cette étude, nous avons comparé trois modes de rendu prothétiques différents : (i) le mode scoreboard (Contrôle) qui est utilisée dans les implants actuels ; (ii) une combinaison de segmentation sémantique d'objets et de détection de structure de la scène appelé "Combiné", similaire à une méthode récente de l'état de l'art [6]) ; et (iii) un mode appelé "Switch" qui permet d'alterner entre le rendu Combiné et deux autres rendus où n'apparaissent respectivement que les objets ou que la structure. Notre hypothèse de travail est que le mode de restitution Switch permet de mieux comprendre la disposition des objets ainsi que la structure dans la scène extérieure que les modes de restitution Contrôle (scoreboard) et Combiné. Étant donné qu'il est impossible de faire de tels tests sur des patients implantés. L'étude a été réalisée avec un simulateur de vision prothétique inspiré de la réalité clinique. Nos résultats, obtenus sur 20 sujets, montrent l'intérêt du mode "Switch" dans la compréhension de scènes statiques (images).

II. État de l'art

A. Les neuroprothèses visuelles

Les neuroprothèses visuelles sont des dispositifs conçus pour restaurer la perception de la lumière chez les personnes atteintes de cécité partielle ou totale. Les implants épi et sous-rétiniens, se composent d'une caméra portable, d'un petit ordinateur et d'un réseau d'électrodes qui est implanté dans la rétine. Ces implants génèrent des micro-stimulations électriques qui provoquent l'apparition de points flous appelés phosphènes. Plusieurs dispositifs ont été mis au point et certains implants ont fait l'objet d'essais cliniques. Les systèmes commerciaux existants de neuroprothèses visuelles sont basés sur l'implantation rétinienne d'un nombre limité d'électrodes (6x10 pour l'Argus II développé par Second Sight [7], 21x18



(a) Scène 3D **(b)** Scoreboard **(c)** Segmentation **(d)** Structure **(e)** Combiné

FIGURE 1 – Exemples de rendus prothétiques : (a) montre l'image initiale dans un environnement virtuel. (b) montre comment (a) est rendu avec le rendu scoreboard. (c) montre comment (a) est rendu avec une méthode qui détecte uniquement les objets (segmentation sémantique des objets), et (d) montre comment (a) est rendu avec une méthode qui améliore les informations structurelles (amélioration de la structure). (e) est un rendu combinant (c) et (d) et appelé Combiné.

pour le système PRIMA développé par Pixium Vision [5]). Si ces neuroprothèses ont permis d'améliorer la vie quotidienne de nombreux non-voyants [8], la perception visuelle restaurée est trop faible pour permettre des processus perceptifs ou sensorimoteurs plus avancés tels que la navigation.

B. Différents rendus phosphéniques en fonction de la tâche

Les limitations des neuroprothèses visuelles actuelles posent des problèmes importants pour le rendu d'une scène de manière compréhensible. La méthode historique de rendu "scoreboard" consiste à réduire l'image capturée à la résolution de l'implant, puis à convertir l'image en niveaux de gris et à quantifier ces derniers au niveau d'intensité supporté par l'implant (voir figure 1b). Vergnieux et al. [9] ont proposé d'inclure certaines informations structurelles dans le rendu afin de permettre la navigation dans la scène. A l'aide d'un simulateur de vision prothétique, ils ont montré que la mise en évidence des bords des murs et des sols avec une intensité élevée permet d'améliorer les performances de navigation des sujets [10]. Un exemple de rendu de ce type est proposé sur la figure 1d. Li et al. [11] ont montré qu'une segmentation sémantique des objets (comme des chaussures ou des chaises) dans le rendu visuel améliore significativement la perception de la scène. Un exemple de rendu de ce type est proposé sur la figure 1c. Sanchez-Garcia et al. [6] ont proposé un rendu qui construit une représentation schématique des environnements intérieurs, mettant en évidence les bords informatifs structurels et les silhouettes des objets segmentés. Ils ont montré à l'aide d'un simulateur de vision prothétique (32x32 électrodes, 8 niveaux de luminance), que ce rendu est supérieur au rendu scoreboard pour la reconnaissance des objets et l'identification des pièces. Un exemple de rendu de ce type est proposé sur la figure 1e.

Cependant, comme la résolution des neuroprothèses est limitée, la combinaison de différents rendus pourrait renforcer l'encombrement visuel et altérer la perception. Une autre solution consisterait à donner à l'utilisateur la possibilité de passer d'un rendu à l'autre selon son désir (rendu interactif). Dans ce cas, l'encombrement visuel serait réduit et la perception pourrait être améliorée. Dans cette étude, nous avons conçu un simulateur de vision prothétique et une tâche visuelle consistant à identifier des objets et des configurations de carrefours routiers dans des scènes visuelles virtuelles. Les réponses de vingt sujets ont confirmé notre hypothèse selon laquelle le rendu interactif améliore significativement la compréhension de la scène.

III. SIMULATEUR DE VISION PROTHÉTIQUE

En raison de l'impossibilité d'utiliser des patients implantés pour tester différents rendus de façon systématique, l'utilisation d'un simulateur de vision prothétique est courante dans la littérature [6], [9], [10], [12]. Comme son nom l'indique, le simulateur permet de simuler différents implants (taille, position, résolution) dans des contextes variés (scènes visuelles 2D ou 3D, statiques ou dynamiques). Pour conduire cette étude, nous avons réalisé un simulateur de vision prothétique en utilisant le moteur Unity 3D développé par Unity Technologies (version 2019.4.19f1). Cet outil initialement dédié au développement de jeux vidéo inclut de nombreuses fonctionnalités (moteur de rendu, moteur physique, composants d'interface utilisateur) qui nous ont permis de publier notre simulateur en ligne [1].

A. Implant simulé

Nous avons choisi de simuler le système PRIMA développé par Pixium Vision (Paris, France), composé d'un réseau de 21x18 électrodes. Notre choix est motivé par le fait que c'est l'implant le plus récent ayant passé des phases d'essais cliniques satisfaisantes [3]. Il est aussi l'un de ceux qui permet

1. http://ubee.enseeiht.fr/Julien/WebV2-17/index.html

un rendu visuel avec la plus grande résolution à ce jour. En effet, les limitations techniques actuelles ne permettent pas d'avoir un nombre d'électrodes très élevé (l'Argus II possède seulement 6x10 électrodes) [13].

Un taux de défaillance (appelé "dropout" en anglais) de 10% était appliqué aux électrodes, afin de simuler des électrodes non fonctionnelles ou cassées. Un flou Gaussien était appliqué aux phosphènes générés, la taille des phosphènes au sein d'un même implant variait entre 0,235° et 0,275° du champ de vision. L'espacement entre deux phosphènes variait entre 0,55° et 0,825° du champ de vision. Chaque électrode pouvait générer un phosphène avec quatre niveaux de gris. Notre simulateur est inspiré de ceux proposés par Chen et al. [12] et Denis et al. [14].

B. *Rendus phosphéniques*

Sur la base de ces règles, une disposition différente des phosphènes a été générée de manière aléatoire pour chaque sujet, afin de modéliser le scénario réaliste où deux personnes implantées avec le même modèle de prothèse virtuelle ont des perceptions phosphéniques différentes.

Dans cette étude, nous avons utilisé le rendu scoreboard (Contrôle), utilisé par défaut dans certains implants commercialisés actuellement. Nous avons aussi utilisé un rendu mélangeant les segmentations d'objets et de structure (appelé "Combiné") récemment proposé par Sanchez et al. [6]. Pour tester notre hypothèse, nous avons conçu un troisième mode de rendu appelé "Switch" qui permet de changer le rendu à volonté parmi les trois suivants : Objets seuls, Structure seule, Objets + Structure. Les figures 2 et 3 illustrent respectivement la construction des rendus Contrôle et Combiné.

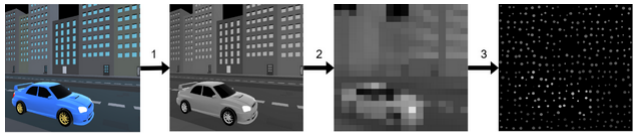


FIGURE 2 – Génération du rendu scoreboard (Contrôle). Chaque image est d'abord convertie en une image en niveaux de gris (étape 1). L'image est ensuite réduite à la résolution de l'implant simulé (dans notre cas 21x18) en faisant la moyenne des couleurs dans une fenêtre locale (on découpe l'image d'entrée en 21 x 18 = 378 zones, étape 2). Nous appliquons ensuite un filtre sélecteur d'intensité uniforme pour quantifier l'intensité en quatre niveaux de gris. Enfin, nous transformons chaque zone rectangulaire qui représente désormais une électrode de l'implant en un phosphène (étape 3).

Le mode de rendu Switch permet au sujet de changer de rendu à volonté. Il lui suffit, pour ce faire, de cliquer sur un bouton prévu pour cet effet. Dans ce mode, les rendus disponibles sont le rendu Combiné, le rendu Objets et le rendu Structure. Le rendu Objets est obtenu en transformant l'image issue de la figure 3 - 2A en version phosphénique. Le rendu Structure est obtenu en transformant l'image issue de la figure 3 - 2B en version phosphénique.

IV. MÉTHODE ET PROTOCOLE

A. *Hypothèse*

Nous avons formulé l'hypothèse suivante : la possibilité de changer de type de rendu en temps réel pour réaliser des tâches

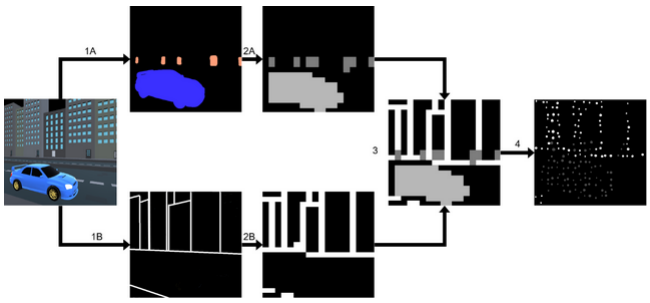


FIGURE 3 – Génération du rendu Combiné (objets + structure). Nous extrayons d'abord une carte de segmentation d'objets de l'image d'entrée (étape 1A). Afin d'obtenir cette carte de segmentation en temps réel, nous avons au préalable labélisé tous les objets dans notre scène. Nous avons mis en place un mécanisme dans la simulation qui permet à la caméra de modifier en temps réel la texture des objets en fonction du label associé. Cela nous permet d'obtenir des cartes des images de segmentation sémantique d'objets (1A). Comme pour le rendu Scoreboard, cette image est mise à l'échelle et quantifiée à l'étape 2A. En parallèle, nous extrayons également les contours de la structure de la scène (1B). Pour récupérer ces contours, nous appliquons un filtre de Sobel sur l'image de la scène sans les objets. On utilise dans ce cas une caméra qui ne détecte que les bâtiments et les routes. L'image des contours est ensuite mise à l'échelle et quantifiée (étape 2B). À ce stade, nous avons deux images qui représentent respectivement la segmentation Objets (2A) et la structure de la scène (2B). Ces deux images sont ensuite Combinées en une seule image (3), en conservant la valeur d'intensité de chaque pixel issue de l'extraction de structure puis en écrasant ces valeurs par celles chaque de la segmentation d'objets non nulles. Enfin, nous transformons chaque zone rectangulaire qui représente désormais une électrode de l'implant en un phosphène (4).

de compréhension de scènes visuelles extérieures permet aux sujets d'avoir de meilleures performances visuelles.

B. *Sujets*

Les sujets ont été recrutés d'une part dans une école d'ingénieur et d'autre part via les réseaux sociaux (LinkedIn, Facebook). 20 sujets, (11 hommes et 9 femmes âgés de 17 à 55 ans, moyenne : 25 ans, écart-type : 13 ans), ont participé à l'expérience.

C. *Conditions expérimentales*

Lors de cette étude nous avons inclus trois conditions expérimentales : le rendu Scoreboard (Contrôle) seul, le rendu Combiné (Objets + Structure) seul, et le mode Switch qui permet d'utiliser un ou plusieurs rendus (parmi Objets, Structure et Combiné) à volonté avant de répondre.

D. *Protocole expérimental*

Avant de démarrer l'expérience, les participants ont rempli un formulaire avec des données descriptives telles que leur âge, adresse email, état de la vision et un formulaire de consentement.

La session expérimentale contenait trois blocs correspondant aux trois conditions expérimentales. Chaque bloc était divisé en trois parties comprenant la familiarisation avec le rendu utilisé dans le bloc, la phase de test, puis un questionnaire concernant le rendu que le sujet vient d'utiliser dans le bloc.

Pendant la familiarisation, les sujets étaient libres d'utiliser le simulateur pour répondre à deux questions, sans contrainte

de temps. Pour la condition Switch, ils pouvaient changer de rendu à volonté avant de répondre à ces deux questions.

Pendant la phase de test, le sujet devait répondre à une série de 15 questions incluant 5 questions dans les trois catégories nommées "objets", "rues", "portes et passages piétons". Les questions de la catégorie "objets" portaient sur la présence d'objets qui peuvent gêner le déplacement du sujet. Les questions de la catégorie "rues" portaient sur la configuration des intersections. Les questions de type "portes et passages piétons" portaient sur la présence de portes et de passages piétons dans la scène visuelle. Une fois la réponse choisie, il passait à la question suivante (il ne connaît pas la validité de sa réponse). L'ordre des blocs et des questions était trié de manière pseudo-aléatoire pour limiter les biais inter-blocs ou inter-questions. Pour éviter que les sujets ne reconnaissent certaines questions d'un bloc à un autre, l'angle de vue de la caméra dans la scène était modifié. Ainsi la question restait la même, mais la réponse pouvait changer selon l'orientation de la caméra.

A la fin de chaque bloc, les sujets devaient répondre à des questions subjectives concernant la pertinence du rendu utilisé dans ce bloc. Les questions portaient sur le plaisir éprouvé et l'utilisabilité perçue concernant le rendu utilisé dans ce bloc. Il y avait 4 questions pour les conditions Scoreboard et Combiné, et 5 questions (dont 4 identiques aux précédentes) pour la condition Switch. Dans le cas de la condition Switch, on demandait également aux sujets s'ils trouvaient cette fonctionnalité pertinente. Pour toutes ces questions, ils devaient répondre sur une échelle de Likert allant de 1 à 7 (1 signifiant pas du tout d'accord et 7 tout à fait d'accord). Les questions posées sont disponibles à l'adresse suivante [2].

E. Variables analysées

Pour répondre à notre hypothèse, nous avons mesuré, pour chaque sujet, plusieurs variables :

- la validité de chaque réponse (correct / incorrect / je ne sais pas) ;
- le temps de réponse à chaque question ;
- pour la condition Switch : l'ordre de sélection des rendus, la durée d'utilisation de chaque rendu, le nombre de changements de rendu.

Pour analyser les résultats quantitatifs nous avons effectué des tests ANOVA à 2 facteurs selon le modèle (Condition * Tâche * Interaction), puis nous avons utilisé un test post-hoc de Tukey pour comparer les paires deux à deux. Dans les figures, nous avons utilisé des intervalles de confiance à 95%.

V. Résultats

A. Comparaison des trois rendus

Réponses : Une ANOVA à deux facteurs (Condition * Tâche * Interaction) a montré que le nombre de réponses correctes est significativement différent selon le rendu ($F(2,171) = 15{,}834$; $p < 0{,}0001$) et selon la tâche $F(2,171) = 81{,}730$; $p < 0{,}0001$). L'interaction est également significative ($F(4,171) = 5{,}728$; $p < 0{,}001$), ce qui montre que la condition Switch est très efficace pour identifier les objets, mais encore plus pour comprendre la configuration des rues (voir Figure 4).

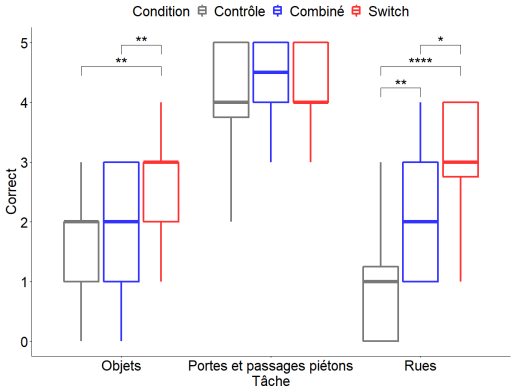


Figure 4 – Nombre moyen de réponses correctes par Rendu et par Tâche. La détection des passages piétons et des portes est équivalente pour toutes les conditions. L'identification des objets et des configurations de rues est significativement améliorée avec le rendu Switch. (N=20. Les barres indiquent un intervalle de confiance de confiance à 95%. *=0.05 ; **=0.01 ; ***=0.001 ; ****=0.00001)

Réponses "Je ne sais pas" : Une ANOVA à deux facteurs (Condition * Tâche * Interaction) a montré que le nombre de réponses "je ne sais pas" est significativement différent selon le Rendu ($F(2,171) = 10.543$; $p < 0{,}0001$) et selon la tâche $F(2,171) = 5.882$; $p < 0{,}001$). L'interaction n'est cependant pas significative ($F(4,171) = 1.909$; $p = 0.111$). Le nombre de réponses "je ne sais pas" est significativement plus élevé avec le rendu Scoreboard pour les tâches "objets" et "rues".

Temps de réponse : La distribution des temps de réponse n'étant pas normale, nous avons réalisé une transformation logarithmique des données. Nous avons réalisé une ANOVA à deux facteurs (Condition * Tâche * Interaction) sur les données transformées qui a montré que le temps de réponse est significativement différent selon la condition expérimentale ($F(2,891) = 17.245$; $p < 0{,}0001$) et selon la tâche $F(2,891) = 14.081$; $p < 0{,}0001$). L'interaction n'est cependant pas significative ($F(4,891) = 0.465$; $p = 0.762$). Le rendu Switch obtient des temps de réponses plus longs comparés au rendu Scoreboard, quelle que soit la tache réalisée.

B. Jugement subjectif

Pour rappel, les participants devaient répondre avec un score entre 1 et 7 à une série de questions à l'issue de chacun des trois blocs.

Une ANOVA à deux facteurs (Rendu * Tâche * Interaction) montre que le score est significativement différent selon les rendus ($F(2,228) = 11.878$; $p < 0.0001$) et selon la tâche ($F(3,228) = 5.604$; $p < 0.001$). L'interaction est significative ($F(6,228) = 7.537$; $p < 0.0001$). La Figure 5 présente les résultats obtenus.

Notons qu'à la question : "J'ai trouvé la possibilité de changer de rendu très utile", on obtient un score moyen de 6.2 +/- 1.2 (sur 7).

2. https://osf.io/dqvy8/?view_only=2807537abf8f437ebd4c33e45ecb67de

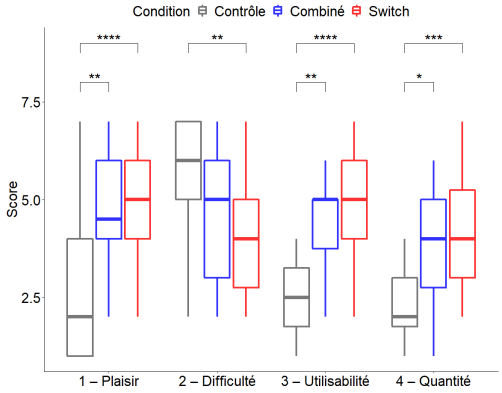


FIGURE 5 – Score moyen donné par les sujets aux trois rendus Scoreboard, Combiné et Switch. Cette figure montre des résultats significativement différents. Analyse réalisée sur les 20 sujets. Les scores sont compris entre 1 et 7 (1 signifiant pas du tout d'accord et 7 tout à fait d'accord).

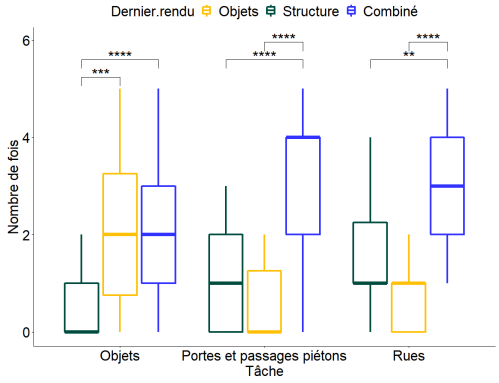


FIGURE 6 – Nombre moyen d'utilisation en dernière instance des trois rendus du mode Switch par Tâche. Analyse réalisée sur les 20 sujets.

C. Classement final

Les sujets devaient finalement classer les rendus par ordre de préférence en fonction de quatre dimensions : pour identifier les intersections et les coins de rues, pour identifier les objets, pour identifier les portes et passages piétons, et globalement. Afin de générer un score de préférence, nous avons attribué trois points si le rendu est classé premier, 2 points s'il est classé deuxième et 1 point s'il est classé troisième. On observe que l'ordre de préférence des rendus est Switch, Combiné, Contrôle, quelle que soit la question posée.

D. Analyse détaillée du comportement observé pour le rendu Switch

Dans le mode Switch, les utilisateurs pouvaient choisir d'utiliser librement un des trois rendus Objets, Structure, Combiné. Pour chaque essai réalisé par les sujets, nous avons analysé plus précisément le nombre de fois où chacun des rendus était utilisé, la durée d'utilisation de chaque rendu et le dernier rendu utilisé. La distribution des durées d'utilisation n'étant pas normale, nous avons réalisé une transformation logarithmique des données. Une ANOVA à un facteur (Tâche) a montré que les durées d'utilisation (après transformation) des rendus (Objets, Structure, Combiné) ne sont pas significativement différentes selon la tâche ($F(2,860) = 0.09$; $p = 0.914$). Une ANOVA à un facteur (Tâche) a montré que le nombre d'utilisations de chaque rendu (Objets, Structure, Combiné) n'est pas significativement différent selon la tâche non plus ($F(2,860) = 1.589$; $p = 0.205$). Une ANOVA à un facteur (Tâche) a montré que le nombre de fois ou un rendu est utilisé en dernier est significativement différent selon la tâche ($F(2,171) = 5.914$; $p < 0.001$).

VI. DISCUSSION

Nos résultats montrent que la détection des passages piétons et des portes est facile, indépendamment du rendu utilisé. Ceci n'est pas surprenant car pour les portes, le rendu consiste systématiquement en un rectangle blanc qui est facile à percevoir. Une étude de Moreno [15] utilisait une méthode différente pour détecter les portes. Leur objectif était différent du nôtre dans le sens où ils cherchent à donner un indice sonore sur l'emplacement de portes le long de murs dans un couloir. Les passages piétons sont moins faciles à percevoir que les portes mais plus faciles à percevoir que les autres objets car ils apparaissent comme une succession de rectangles parallèles.

Par contre, l'identification des objets et la compréhension de l'organisation des rues dépendent fortement du rendu utilisé. Les résultats montrent que le rendu Switch est significativement meilleur que le rendu Contrôle et le rendu Combiné. Nous avons observé que le rendu Combiné n'est pas significativement plus efficace que le rendu Contrôle pour la détection des objets. Ceci n'est pas cohérent avec l'étude de Sanchez et al. [6]. Cette divergence peut s'expliquer par la résolution des implants utilisés dans notre simulateur (21x18) qui est inférieure à celle utilisée dans le leur (32x32). En effet, selon Cha et al. [16], 600 électrodes avec un rendu de type scoreboard (Contrôle) sont suffisantes pour obtenir une perception déjà fonctionnelle. L'analyse des réponses "je ne sais pas" montre qu'il y a une différence significative selon les rendus et les tâches. Le rendu Contrôle, obtient plus de réponses "je ne sais pas" que le rendu Switch dans les taches portant sur les "objets" et les "rues".

En ce qui concerne les temps de réponses, nous obtenons des différences significatives. Le rendu Switch est systématiquement plus utilisé en termes de temps d'utilisation que le rendu Contrôle peu importe la tâche. Il est également plus utilisé que le rendu Combiné dans les tâches de type "Portes et passages piétons". Cela peut s'expliquer par le fait que changer de rendu rallonge automatiquement le temps de réponse. En contrepartie, comme le montrent les indices de préférence, la décision est plus sûre (les sujets jugent avoir suffisamment d'informations pour répondre à la question) et la satisfaction plus grande. On pourrait imaginer qu'avec de l'entraînement, la décision soit prise plus rapidement avec le mode Switch.

En ce qui concerne l'analyse des jugements subjectifs et du classement des modes de rendu, nous mettons en évidence que le mode de rendu Switch présente un net avantage sur le mode de rendu Scoreboard dans toutes les catégories. Cela dit, le mode de rendu Switch n'est pas significativement préféré

par rapport au mode de rendu Combiné. En revanche, le classement final montre quand même que le mode de rendu classé premier est le Switch devant le mode de rendu Combiné.

Les résultats obtenus en regardant de plus près le comportement des sujets lorsqu'ils utilisent le mode de rendu Switch sont également intéressants. La plupart des sujets utilisent souvent les rendus dans l'ordre d'affichage des boutons à l'écran. Il semble donc qu'il n'y ait pas de stratégie consciente d'utilisation des rendus de la part des sujets. Par contre, on observe que le dernier rendu choisi est significativement relié avec la tache réalisée. On observe notamment que les rendus Combiné et Objets sont plus souvent les derniers rendus utilisés pour l'identification d'objets. Les rendus Combiné et Structure sont plus souvent les derniers rendus utilisés pour comprendre la configuration des rues. Ceci confirme l'hypothèse selon laquelle les rendus appropriés à chaque tâche sont plus utilisables pour prendre la décision finale. Par contre, le rendu combiné est utilisé plus fréquemment en dernier ce qui va à l'encontre de notre hypothèse sur l'encombrement visuel. En effet, dans ce cas, le rendu Objets devrait être privilégié pour percevoir les objets et le rendu Structure devrait être privilégié pour comprendre les orientations des rues. Cependant ce résultat est à considérer en fonction du fait que le rendu Switch est largement préféré au rendu Combiné seul (en dehors de la condition Switch, quand il est présenté comme seul rendu disponible). Une interprétation possible est que le rendu Combiné est apprécié après avoir utilisé les rendus appropriés pour percevoir les items dans la scène visuelle. Il permettrait de confirmer la perception unitaire puis d'essayer de les replacer dans la scène globale afin de mieux la comprendre.

VII. Limitations de l'étude et perspectives

Dans un premier temps nous pouvons aborder la question de la prédiction des images de segmentation et de structure pour construire les différents rendus. Dans notre étude, ces cartes sont très faciles à obtenir car les objets de la scène sont labélisés et les caméras paramétrées pour ne capturer que les éléments qui sont pertinents (objets pour la caméra objets qui change la texture en fonction des labels, éléments de structure pour la caméra de structure). En condition réelle il faudrait trouver un autre moyen de récupérer ces informations. L'ajout de réseaux de neurones permettrait de résoudre ce problème. La prédiction de segmentation sémantique est un sujet très abordé depuis déjà quelques années avec la production de réseaux tels que EfficientDet [17] ou DeepLab [18]. En ce qui concerne la récupération de structure il existe également quelques réseaux entrainés à prédire ce genre d'informations comme par exemple PanoRoom [19]. Il reste deux problèmes à gérer, les erreurs dans les prédictions et la vitesse de calcul. Il faut que les prédictions ne soient pas trop loin de la réalité et que ces prédictions soient produite en temps quasi-réel.

Pour pousser le réalisme au maximum il pourrait également être intéressant de réaliser une expérience en condition réelle. Le dispositif pourrait être composé d'un casque de réalité virtuelle équipé de caméra qui filme la scène et d'un boîtier de calcul qui s'occupe de calculer le rendu visuel. Ce dispositif serait une extension du premier proposé par Cha en 1992 [16]. L'intérêt de faire une étude en conditions réelles est double : d'un côté on pourrait proposer un dispositif d'étude complet très proche de la réalité, ce qui permettrait aux futurs travaux d'être essayé sur un modèle réaliste. De plus nous pourrions lors de cette étude mesurer la réaction des sujets dans des tâches de navigation et non plus de perception et de compréhension.

Remerciements

Nous remercions le laboratoire Cherchons Pour Voir (CPV) pour l'aide fournie pendant ces travaux.

Références

[1] R. R. Bourne *et al.*, "Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment : a systematic review and meta-analysis," *The Lancet Global Health*, 2017.

[2] B. S. Wilson and M. F. Dorman, "Cochlear implants : current designs and future possibilities," *J Rehabil Res Dev*, 2008.

[3] E. Fernandez, "Development of visual neuroprostheses : trends and challenges," *Bioelectronic medicine*, 2018.

[4] K. Stingl *et al.*, "Subretinal visual implant alpha ims–clinical trial interim report," *Vision research*, 2015.

[5] D. Palanker *et al.*, "Photovoltaic restoration of central vision in atrophic age-related macular degeneration," *Ophthalmology*, 2020.

[6] M. Sanchez-Garcia *et al.*, "Semantic and structural image segmentation for prosthetic vision," *Plos one*, 2020.

[7] D. D. Zhou *et al.*, "The argus® ii retinal prosthesis system : An overview," in *IEEE international conference on multimedia and expo workshops (ICMEW)*, 2013.

[8] K. Stingl *et al.*, "Interim results of a multicenter trial with the new electronic subretinal implant alpha ams in 15 patients blind from inherited retinal degenerations," *Frontiers in neuroscience*, 2017.

[9] V. Vergnieux *et al.*, "Wayfinding with simulated prosthetic vision : Performance comparison with regular and structure-enhanced renderings," in *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2014.

[10] ——, "Simplification of visual rendering in simulated prosthetic vision facilitates navigation," *Artificial organs*, 2017.

[11] H. Li *et al.*, "Image processing strategies based on saliency segmentation for object recognition under simulated prosthetic vision," *Artificial intelligence in medicine*, 2018.

[12] S. C. Chen *et al.*, "Simulating prosthetic vision : I. visual models of phosphenes," *Vision research*, 2009.

[13] M. Farvardin *et al.*, "The argus-ii retinal prosthesis implantation ; from the global to local successful experience," *Frontiers in neuroscience*, 2018.

[14] G. Denis *et al.*, "Simulated prosthetic vision : improving text accessibility with retinal prostheses," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2014.

[15] M. Moreno *et al.*, "Realtime local navigation for the blind : detection of lateral doors and sound interface," *Procedia Computer Science*, 2012.

[16] K. Cha *et al.*, "Simulation of a phosphene-based visual field : visual acuity in a pixelized vision system," *Annals of biomedical engineering*, 1992.

[17] M. Tan *et al.*, "Efficientdet : Scalable and efficient object detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020.

[18] L.-C. Chen *et al.*, "Deeplab : Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE transactions on pattern analysis and machine intelligence*, 2017.

[19] C. Fernandez-Labrador *et al.*, "Panoroom : From the sphere to the 3d layout," 2018.