THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)

Présentée et soutenue le jeudi 11 décembre 2014 par

Grégoire DENIS

Apport de la vision par ordinateur dans l'utilisabilité des neuroprothèses visuelles

---

École doctorale et discipline ou spécialité
ED MITT : Image, Information, Hypermédia

Unité de recherche
Institut de Recherche en Informatique de Toulouse – UMR 505

Directeur(s) de Thèse

Christophe Jouffrais	Chercheur, Université Paul Sabatier & CNRS, Toulouse
Corinne Mailhes	Professeur, ENSEEIHT, Toulouse

Jury

Mohamad Sawan	Professeur, École Polytechnique, Montréal (Canada)	Rapporteur
Ryad Benosman	Professeur, Institut de la Vision, Paris	Rapporteur
Isabelle Marc	Maître de Conférences, École des Mines, Alès	Examinatrice
Denis Kouamé	Professeur, Université Paul Sabatier, Toulouse	Examinateur
Christophe Jouffrais	Chercheur, Université Paul Sabatier, Toulouse	Directeur de Thèse
Marc Macé	Chercheur, Université Paul Sabatier, Toulouse	Encadrant

Invitée

Corinne Mailhes	Professeur, ENSEEIHT, Toulouse	Co-Directrice de Thèse

À mon père,

RESUME

L'Organisation Mondiale de la Santé estime, dans un rapport datant d'octobre 2013, que 285 millions de personnes dans le monde présentent une déficience visuelle, 39 millions d'entre elles étant aveugles. Avec le vieillissement de la population, ce chiffre est en constante progression car la cécité touche majoritairement les personnes âgées. Au quotidien, les conséquences de cette affection sont importantes puisqu'elle réduit considérablement l'autonomie des personnes qui en sont atteintes.

Pour aider les personnes non-voyantes à retrouver une certaine autonomie dans leurs tâches quotidiennes, de nombreuses aides techniques ont été développées. La canne blanche et les lecteurs d'écrans en sont deux exemples fonctionnels. Au-delà de ces aides techniques pour la compensation, les neuroprothèses visuelles ont pour objectif de restaurer la vision. Ces systèmes convertissent les informations de la scène visuelle en un train de microstimulations électriques. Celles-ci sont ensuite injectées dans des électrodes implantées dans l'un des relais du système visuel (rétine, nerf optique ou cortex visuel). Ces microstimulations électriques induisent la perception de phosphènes (des points lumineux souvent blanchâtres et circulaires). En étant très schématique, la perception visuelle générée correspond à une image en noir et blanc de très faible résolution (quelques dizaines de pixels assez espacés). Aujourd'hui on dénombre une vingtaine de projets en cours de développement, parmi lesquels un peu moins de la moitié est en test clinique. Bien qu'ils représentent un espoir important, ces systèmes restent, à ce jour, inutilisables dans un environnement naturel : l'information visuelle restituée est insuffisante pour que les personnes implantées puissent se déplacer en toute sécurité, localiser et reconnaître des objets ou lire confortablement.

Ces dernières décennies, la vision par ordinateur a connu d'énormes avancées, grâce à la mise au point de nouveaux algorithmes de traitement des images et à l'augmentation de la puissance de calcul des processeurs. Désormais, il est possible de localiser et reconnaître des objets, d'identifier des visages ou encore de lire du texte, et ceci en temps réel. Or, la plupart des neuroprothèses visuelles en test clinique intègrent une caméra qui peut facilement être associée à un module de traitement d'images.

Partant de ces constatations, nous proposons dans cette thèse d'améliorer l'utilisabilité des neuroprothèses visuelles à faible résolution (ce qui concerne les modèles actuels mais aussi les futurs systèmes testés), en utilisant des algorithmes de traitement d'image performants. Notre hypothèse de travail est qu'il est possible d'extraire puis de restituer la position d'une ou plusieurs zones d'intérêt dans une scène naturelle pour restaurer des comportements « visuomoteurs ». Pour mener nos expérimentations, nous avons dans un premier temps implémenté un simulateur de vision prothétique. Cet outil nous a permis d'évaluer cette approche dans trois tâches de la vie quotidienne : la localisation d'objets, de visages, et de texte.

Les résultats de ces travaux indiquent que malgré la faible résolution spatiale disponible dans les neuroprothèses visuelles actuelles et envisagées à court terme, il est dès à présent possible de restaurer, en s'appuyant sur des algorithmes temps réel de vision artificielle, la fonction visuelle primordiale de reconnaissance et localisation d'objets. Un algorithme temps réel de reconnaissance de texte permet, par exemple, la localisation de textes distants de plusieurs mètres. Dans cette thèse, nous montrons également l'intérêt d'imaginer différents types de rendu visuel en fonction de la tâche à réaliser. Cette multiplicité des rendus ouvre la voie à une meilleure utilisabilité des neuroprothèses à basse résolution, à condition de permettre le contrôle interactif par l'utilisateur du dispositif implanté. Dans ce cadre, nous proposons en preuve de concept un prototype de système interactif, dans lequel il est possible de contrôler le zoom de la caméra, et de changer de rendu visuel à la demande.

Mots-clés : non-voyant, neuroprothèses visuelles, implant rétinien, implant cortical, vision par ordinateur, vision prothétique simulée

ABSTRACT

In October 2013, the World Health Organization (WHO) estimated that 285 million people were suffering from visual impairment, among them 39 million were blind. With the ageing of the world population, this figure is constantly increasing because the elderly people are the most affected by blindness. The consequences of this ailment are really significant on a daily basis, as it reduces considerably the autonomy of blind people.

Several assistive devices such as the white cane or screen readers have been developed to help blind people to regain some autonomy in their daily tasks. More than just compensation techniques, visual neuroprostheses aim at restoring vision. These systems convert visual information into electrical microstimulations injected in electrodes implanted in the visual pathway (retina, optic nerve or visual cortex). These stimulations induce dots like perception called phosphenes. The restored visual perception corresponds to a black and white image with low resolution (tens of phosphenes with gaps between them). About twenty projects are currently in progress. Less than half are in human clinical trial phase. However, even if these systems give great hope to blind people, they are still inappropriate in a natural environment: the restored visual information is insufficient to enable the implanted people to navigate safely, to localize and recognize objects or to read correctly.

These last decades, computer vision field has reported important progresses, thanks to the development of new image processing algorithms as well as the increase of calculation power of processors. For example, this is now possible to localize and recognize objects, to identify faces or even to read texts in real time. Interestingly, most of the current visual neuroprostheses include an external camera, so it could be possible to process input images before the generation of phosphenes.

Given that observations, we propose the use of sophisticated real-time image processing to improve the usability of current and upcoming low resolution visual neuroprostheses. We extract high-level information from input images (such as the localization of an object), and render it with only few phosphenes to restore visuomotor behaviors. To support our experiments, we have implemented a simulator of prosthetic vision which allows us to evaluate our approach in three different daily tasks: objects localization, faces localization and text localization.

Our results suggest that implanted people could benefit from real-time computer vision algorithms, to localize objects with current and upcoming low resolution visual neuroprotheses. We also point out that task-specific renderings (for navigation, object and face localization, text localization, etc.) could all be integrated in one device, allowing a prosthesis user to switch between different modes. In this context we propose an interactive system prototype, in which the user can control the camera zoom and change the visual rendering on demand.

Keywords: blind people, visual neuroprostheses, retinal implants, cortical implants, computer vision, simulated prosthetic vision

TABLE DES MATIERES

Résumé **2**

Abstract **3**

Table des matières **4**

Introduction **7**

Classification des déficiences visuelles 7
Les principales causes de la cécité 8
Les handicaps liés à la cécité 9
Les technologies d'assistance pour les non-voyants 9
Problématique de la thèse et plan du manuscrit 11

CHAPITRE 1 Les neuroprothèses visuelles **13**

Introduction 14
Le système visuel humain 15
Un peu d'histoire 17
Les neuroprothèses corticales 20
Les neuroprothèses du nerf optique 27
Les neuroprothèses rétiniennes 29
Discussion 39

CHAPITRE 2 Traitement d'images et vision par ordinateur **47**

Introduction 48
Le traitement d'images 49
La vision par ordinateur 49
Vision par ordinateur et neuroprothèses visuelles 60
Discussion 66

CHAPITRE 3 Simulateur de vision prothétique **71**

Introduction 72
Caractéristiques des neuroprothèses visuelles et de la vision prothétique 73
Architecture du simulateur 79
Conclusion 84

CHAPITRE 4 Vision prothétique - Localiser et atteindre un objet **85**

Introduction 86
Expérience 1 89
Expérience 2 100
Discussion générale 107

CHAPITRE 5 Vision prothétique - Localiser un visage 113

Introduction 114
Expérience 117
Discussion 125

CHAPITRE 6 Vision prothétique - Localiser un texte 129

Introduction 130
Discussion 141

Discussion et perspectives 145

Verrous initiaux 145
Solutions proposées 145
Résultats 146
Localisation de points d'intérêt, scoreboard augmenté 148
Vision par ordinateur et neuroprothèses visuelles 150
Neuroprothèses visuelles contextuelles ou interactives ? 151
Perspectives 154

Bibliographie 155

Annexe 1 - Questionnaire post expérience localisation d'objets 165

Annexe 2 - Questionnaire post expérience localisation de texte 169

Figures 171

Tableaux 173

INTRODUCTION

La déficience visuelle est actuellement et depuis bien longtemps un des problèmes de santé majeurs. Selon l'Organisation Mondiale de la Santé (l'OMS), elle touche aujourd'hui près de 300 millions de personnes dans le monde, 15% d'entre eux étant aveugles. Malgré les améliorations sanitaires dans les pays en voie de développement, ce chiffre est en constante progression du fait du vieillissement général de la population et de la prévalence élevée des maladies cécitantes chez les personnes âgées.

CLASSIFICATION DES DEFICIENCES VISUELLES

Dans sa dixième révision de la classification internationale des maladies, l'OMS classe les déficiences visuelles selon l'acuité résiduelle et le champ visuel [WHO 2012]. L'acuité visuelle correspond à notre capacité à discerner un petit objet (un optotype) à la plus grande distance possible. Une mesure d'acuité visuelle peut être représentée par un quotient de deux nombres (système anglo-saxon) : le numérateur correspond à la distance à laquelle le sujet peut voir un objet donné, le dénominateur, la distance à laquelle une personne ayant une vision normale peut voir ce même objet. Ainsi, avec une acuité visuelle de 3/60 une personne déficiente visuelle perçoit lorsqu'il se trouve à moins de 3 mètres un objet qu'un sujet sans déficience discerne à 60 mètres. L'OMS définit cinq grandes catégories de déficiences visuelles numérotées de I à V. Les deux premières correspondent à la malvoyance :

Catégorie I : Acuité visuelle binoculaire corrigée inférieure à 6/18 et supérieure ou égale à 6/60 avec un champ visuel d'au moins 20°.

Catégorie II : Acuité visuelle binoculaire corrigée inférieure à 6/60 et supérieure ou égale à 3/60, avec un champ visuel compris entre 10° et 20°. En pratique, les sujets comptent les doigts de la main à trois mètres.

Les trois catégories qui suivent définissent la cécité légale :

Catégorie III : Acuité visuelle binoculaire corrigée inférieure à 3/60 et supérieure ou égale à 1/60 avec un champ visuel compris entre 5° et 10°. En pratique, le sujet compte les doigts d'une main à un mètre.

Catégorie IV : Acuité visuelle binoculaire corrigée inférieure à 1/60 ou champ visuel inférieur à 5°, mais perception lumineuse préservée. En pratique, le sujet ne compte pas les doigts à un mètre.

Catégorie V : Cécité absolue. Pas de perception lumineuse.

LES PRINCIPALES CAUSES DE LA CECITE

En 2010, parmi les 285 millions de personnes déficientes visuelles dans le monde, on estimait que 39 millions d'entre elles étaient aveugles. Avec 51% des cas (19,7 millions de personnes), la cataracte est la maladie provoquant le plus grand nombre de cécités. Elle correspond à une opacification du cristallin, ce qui entraîne une baisse progressive de l'acuité visuelle, jusqu'à la cécité complète si elle n'est pas traitée. Elle touche une personne sur cinq à partir de 65 ans. La deuxième des causes est le glaucome, une maladie engendrant une surpression intraoculaire qui se traduit par une compression du nerf optique. 4,5 millions de personnes dans le monde sont devenues aveugles suite à un glaucome (près de 12% des cas de cécités). La dégénérescence maculaire liée à l'âge (DMLA) est la première cause de déficience visuelle chez les personnes âgées de plus de 50 ans dans les pays industrialisés. C'est une maladie qui accélère le vieillissement de la macula (disparition progressive des cônes) et par conséquent prive la personne qui en souffre de vision centrale. Parmi les autres troubles de la rétine, la rétinite pigmentaire est une maladie génétique qui se manifeste d'abord par la perte de vision nocturne puis la réduction progressive du champ de vision jusqu'à la cécité. Cette pathologie touche une personne sur 4000 dans le monde. Elle est la cause la plus fréquente de cécité chez les personnes d'âge intermédiaire dans les pays industrialisés.

LES HANDICAPS LIES A LA CECITE

Le monde dans lequel nous vivons est par nature très visuel. La vision nous permet d'effectuer des tâches complexes comme nous déplacer dans notre environnement, lire, écrire, ou encore localiser et saisir des objets. Pour cette raison, les personnes aveugles subissent de profonds handicaps au quotidien. Leur cognition spatiale (à savoir la compréhension de leur environnement, la localisation d'objets et la navigation) en est fortement limitée. Il en est de même pour la communication écrite. Ces handicaps lourds se traduisent directement par une perte d'autonomie chez les personnes atteintes de cécité.

LES TECHNOLOGIES D'ASSISTANCE POUR LES NON-VOYANTS

Pour compenser l'absence de vision, de nombreux dispositifs techniques ont été imaginés pour aider les non-voyants dans leurs tâches de tous les jours. La canne longue, la plage braille ou encore le lecteur d'écran, en sont des exemples fonctionnels et totalement adoptés par la communauté des personnes non-voyantes. Depuis plusieurs décennies, de nombreuses aides électroniques ont également été développées. Elles se sont multipliées et démocratisées grâce à la miniaturisation et la baisse de coût des composants, à l'évolution de la puissance de calcul ou encore au développement de l'informatique embarquée. Parmi ces systèmes, nous pouvons distinguer deux types de dispositifs [Jouffrais 2011] : les systèmes « génériques » visant à substituer ou restituer la vision dans sa totalité, et les systèmes « spécifiques », dédiés à l'accomplissement d'une tâche. Les travaux que nous présentons dans ce manuscrit s'inscrivent dans le cadre des systèmes « génériques ».

Dans les systèmes « génériques », ou systèmes de vision artificielle, deux grandes approches se dégagent : les **systèmes de substitution de la vision** dans lesquels l'information visuelle captée est restituée au travers d'une autre modalité sensorielle (le toucher, l'audition), et les **systèmes de restauration visuelle**, aussi appelés neuroprothèses visuelles, dont l'objectif est de restaurer chez le non-voyant une forme de vision, par l'intermédiaire de percepts lumineux induits par microstimulation électrique du système visuel. Le principe général de fonctionnement de ces systèmes est identique. Il repose sur un module d'acquisition de l'information visuelle

(en général, une caméra), un module de transformation de cette information, et un module de restitution de l'information visuelle traitée.

Les premiers systèmes de substitution de la vision par le toucher ont vu le jour à la fin des années soixante lorsque Paul Bach-y-Rita a développé le système Tactile Vision Substitution System (TVSS) capable de convertir une image en sensations tactiles appliquées sur le dos des sujets [Bach-y-Rita et al. 1969]. Dans le même contexte, nous pouvons citer en exemple le Tongue Display Unit (TDU), qui restitue une image sous la forme d'impulsions électriques au niveau de la langue [Sampaio et al. 2001; Bach-y-Rita et al. 1998]. Ces systèmes ont révélé chez les sujets, après un long apprentissage, une capacité à reconnaître des formes simples.

Dans la même catégorie, les systèmes de substitution de la vision par l'audition transforment chaque image acquise par une caméra en un signal auditif. Le plus souvent, la position et l'intensité des pixels sont restituées en manipulant la fréquence, l'intensité et le délai interauraux des sons. La différence entre les systèmes développés réside dans l'encodage de l'information visuelle. Dans The vOICe [Meijer 1992], des images en niveaux de gris sont captées par une caméra toutes les secondes, et la position et l'intensité de chaque pixel sont convertis en signal auditif, en suivant les trois règles suivantes : 1/ la position verticale est codée en fréquences (plus le pixel est haut dans l'image, plus le son est aigu) ; 2/ la position horizontale est codée de manière temporelle, avec une résolution de 64 pixels en largeur (chaque colonne est sonifiée pendant 1/64ème de seconde) ; 3/ la valeur en niveau de gris d'un pixel est codé en intensité de son émis (plus le pixel est clair, plus le son est fort). Avec un temps d'apprentissage important, il a été montré que ce système pouvait être utilisé dans des tâches simples de reconnaissance et de discrimination d'objets [Auvray et al. 2007]. D'autres systèmes de substitution de la vision par l'audition ont été développés, reprenant le même principe que The vOICe, sans toutefois utiliser de composante temporelle [Auvray 2004; Capelle et al. 1998]. Cependant, là encore, les expérimentations visant à évaluer ces systèmes se limitent à des tâches et des environnements très simplifiés, et les améliorations de performance rapportées se font en contrepartie d'un apprentissage important.

Les systèmes de substitution de la vision par le toucher ou l'audition ont été déclinés et étudiés au travers de nombreux dispositifs [Bach-y-Rita & W. Kercel 2003]. Dans la majorité des cas, ils ont été évalués dans des tâches simples de localisation et de reconnaissance d'objets ou de formes. De plus, les environnements utilisés sont contrôlés, et fortement contrastés (des objets ou des formes noires sur fond blanc, ou inversement). Bien que l'ensemble des résultats mette en évidence la capacité à reconnaître des formes, les dispositifs développés restent des prototypes de laboratoire, et ne sont pas adoptés à grande échelle par la communauté non-voyante [Bach-y-Rita 1983]. Ceci peut s'expliquer par le long apprentissage nécessaire pour commencer à les utiliser, et pour comprendre l'environnement. De façon plus générale, il semble probable que ces systèmes ne soient pas fonctionnels dans des environnements naturels : la résolution spatiale de leur interface de sortie reste très limitée pour exprimer toute la richesse de l'information visuelle fournie par une image.

La deuxième grande catégorie de systèmes « génériques » concerne les neuroprothèses visuelles. Ces systèmes sont apparus dans les années soixante et visent à restaurer des perceptions visuelles chez une personne non-voyante. Ils sont, dans leur majorité, constitués d'une caméra qui acquiert l'information visuelle, d'un module de traitement de l'image, et d'une interface de sortie comprenant une matrice d'électrodes implantée dans l'un des relais du système visuel. Chaque image est codée en impulsions électriques, provoquant l'apparition de percepts lumineux chez la personne implantée. Actuellement, ces dispositifs ne sont pas matures mais ont montré des résultats très prometteurs. Un des principaux verrous réside dans la résolution spatiale proposée par ces dispositifs. Celle-ci est bien trop faible pour restaurer chez les non-voyants des fonctionnalités visuelles utiles dans des environnements naturels.

PROBLEMATIQUE DE LA THESE ET PLAN DU MANUSCRIT

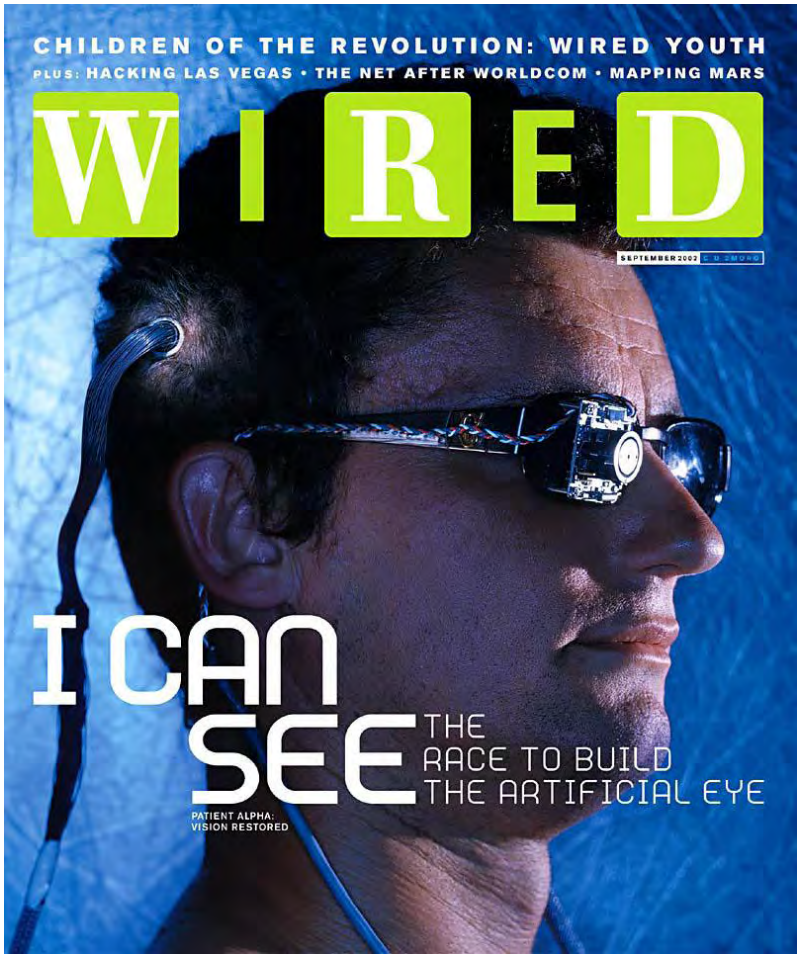
Les recherches que nous avons menées durant ces trois années portent sur l'amélioration de l'utilisabilité des neuroprothèses visuelles actuelles et à venir. Nous pensons que des algorithmes de vision par machine permettant des traitements de haut niveau et en temps réel (par temps réel on entend quelques dizaines de

millisecondes) sont intégrables dans la plupart de ces systèmes. Leur utilisation permettrait de développer et d'imaginer de nouveaux rendus fonctionnels, redonnant aux personnes implantées la possibilité d'effectuer des tâches « visuomotrices » dans leur quotidien.

Dans ce manuscrit, nous présentons, dans le chapitre 1, un état de l'art sur les neuroprothèses visuelles, de leur genèse jusqu'aux systèmes en cours de développement. Dans le chapitre 2, nous dressons un panorama général des possibilités offertes par la vision par machine, et des algorithmes effectivement embarqués dans les neuroprothèses visuelles. En fin de chapitre, j'introduis une approche qui consiste à utiliser des algorithmes de vision artificielle pour localiser des points d'intérêt (objets, visages, texte). Le chapitre 3 introduit la notion de simulateur de vision prothétique, outil nous permettant d'expérimenter à moindre coût cette nouvelle approche. Les chapitres 4 à 6 correspondent aux expérimentations menées dans trois contextes différents (localisation d'objet, de visages et de texte), afin d'évaluer cette approche le plus finement possible. La dernière partie dresse un bilan de l'ensemble des résultats obtenus. Une attention particulière est portée sur le besoin d'imaginer un ensemble de rendus adaptés à différents contextes. Cette multiplicité des rendus ouvre la voie vers plus de contrôle et d'interaction entre le dispositif implanté et l'utilisateur. Dans ce cadre, nous proposons en preuve de concept un prototype de système interactif, dans lequel il est possible de contrôler le zoom de la caméra, et de changer de rendu visuel à la demande. À la suite de ce bilan, nous ouvrons vers les perspectives qui résultent de ce travail.

CHAPITRE 1

LES NEUROPROTHESES VISUELLES



(Source : www.wired.com, septembre 2002)

INTRODUCTION

Une neuroprothèse est un dispositif connecté au système nerveux visant à améliorer ou remplacer une fonction motrice ou sensorielle défaillante. Pour les neuroprothèses sensorielles, il s'agit d'acquérir des informations à l'aide de capteurs artificiels pour se substituer au capteur biologique déficient et stimuler électriquement des relais sensoriels préservés (des fibres nerveuses, des neurones). Les implants cochléaires constituent l'exemple de succès le plus remarquable parmi toutes les neuroprothèses sensorielles développées à l'heure actuelle. Quatre fabricants[1] dans le monde ont déjà équipé plus de 300 000 personnes atteintes de surdité profonde[2]. Le principe général de ces implants est relativement simple (voir Figure 1-1). Un microphone capte les sons dans l'environnement de l'utilisateur. Ce son est filtré, transformé en signaux numériques, puis transmis par ondes radio à un récepteur placé sous la peau. Ce dernier décode les signaux et les convertit sous forme d'impulsions électriques envoyées à un faisceau d'électrodes implantées dans la cochlée. L'excitation des fibres nerveuses encore intactes du système auditif provoque alors la perception de sensations auditives qui peuvent s'apparenter à des sons.

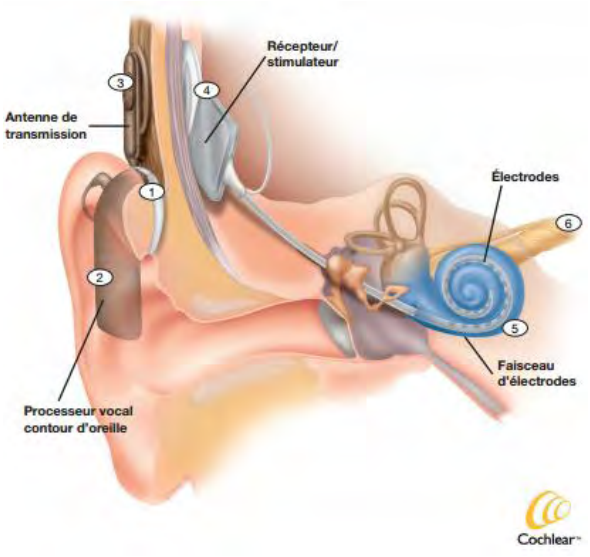


Figure 1-1 L'implant cochléaire.
Un microphone capte le son (1) qui est traité par un microprocesseur (2) et transmis via une antenne radio (3) à un stimulateur placé sous la peau (4). Le signal est convertit en impulsions électriques et envoyé aux électrodes (5) implantées dans la cochlée (6). Ces stimulations restituent une perception sonore. (Source : www.cisic.fr[3]).

[1] Advanced-Bionics (USA), Cochlear Limited (Australie), MED-EL (Autriche) et Neurelec (France).
[2] 324 000 en Décembre 2012 d'après la Food and Drug Administration américaine (FDA).
[3] Le CISC est le Centre d'Information sur la Surdité et l'Implant Cochléaire.

De manière analogue, il est possible de faire apparaître des percepts visuels, que l'on appelle des **phosphènes**, chez des personnes non-voyantes, en stimulant électriquement différents étages du système visuel. Sur le même principe que l'implant cochléaire, ces systèmes sont composés d'un dispositif d'acquisition, d'un microprocesseur et d'un stimulateur. Un capteur artificiel acquiert une image qui est traitée puis convertie en impulsions électriques. Ces impulsions sont envoyées à un ensemble d'électrodes. L'excitation des cellules stimulées par ces électrodes induisent la perception de phosphènes. Cet ensemble de phosphènes constitue ce qu'on appelle la vision prothétique.

Dans ce chapitre, nous décrirons tout d'abord le système visuel humain, puis nous retracerons la genèse des neuroprothèses visuelles. Nous mentionnerons l'ensemble des solutions proposées à ce jour en indiquant leurs avantages et leurs inconvénients. Enfin, nous conclurons par les contraintes techniques et biologiques que rencontrent actuellement ces implants.

LE SYSTEME VISUEL HUMAIN

Le système visuel humain achemine et traite les informations visuelles captées par la rétine. La lumière est captée par les récepteurs de la rétine où elle est transformée en signaux électriques. Ceux-ci transitent alors à travers différents relais nerveux (le nerf optique, le chiasma optique et le corps genouillé latéral) avant d'atteindre le cortex. C'est dans ce dernier que s'effectue l'interprétation du signal.

L'œil comme capteur

L'œil (Figure 1-2) est composé de quatre membranes principales : la cornée, la sclère, la choroïde et la rétine. La cornée est une couche transparente qui prolonge la sclère et laisse passer la lumière. La choroïde est une membrane intermédiaire qui permet d'alimenter en sang l'iris et la rétine. Derrière la cornée, se trouvent la pupille, l'iris et le cristallin. La pupille correspond à la zone au centre de l'iris par laquelle la lumière entre dans l'œil. L'iris est un diaphragme qui s'ajuste en changeant de taille, et qui régule la quantité de lumière qui pénètre dans l'œil. Enfin, le cristallin est une lentille qui focalise l'image sur la rétine pour la rendre nette. Avant d'atteindre la rétine, la lumière traverse l'humeur vitrée (ou corps vitré), une substance gélatineuse

maintenant sous pression le globe oculaire. Celle-ci absorbe les rayons ultraviolets et maintient la rétine contre la choroïde. Elle représente 90% du volume de l'œil.

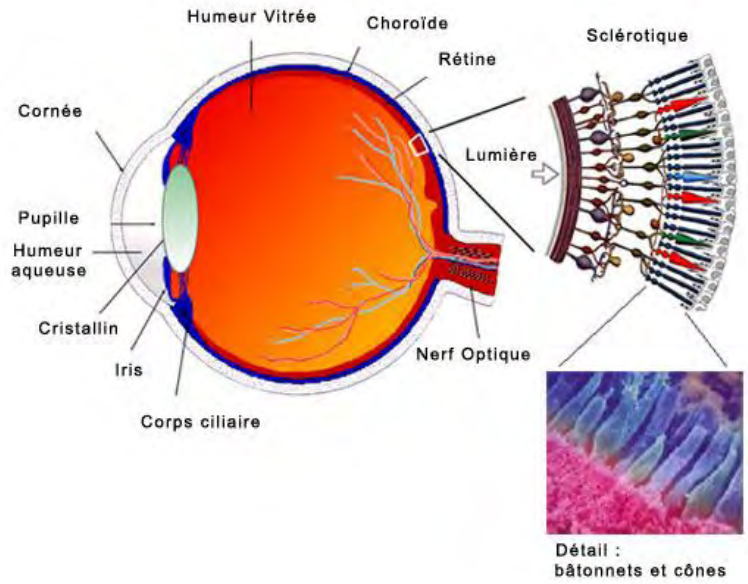


Figure 1-2 Anatomie et physiologie de l'œil humain (Source : http://www.savoirs.essonne.fr/).

Lorsque la lumière atteint la rétine, celle-ci est captée par les cellules photoréceptrices, les cônes et les bâtonnets. Les cônes permettent la vision des couleurs tandis que les bâtonnets permettent d'interpréter les signaux de faible luminosité. Ces cellules transforment l'information lumineuse en courants ioniques qui se transmettent aux cellules bipolaires et horizontales. Les cellules bipolaires propagent ces potentiels jusqu'aux cellules ganglionnaires. L'influx nerveux est généré dans les cellules ganglionnaires et se propage via leurs axones dont le rassemblement forme le nerf optique.

Le nerf optique comme transmetteur

Un nerf optique contient environ 1 million d'axones de cellules ganglionnaires et fait environ 3 à 4 mm de diamètre. L'information visuelle qui vient d'être transformée en impulsions électriques transite par les deux nerfs optiques (Figure 1-3). Ceux-ci se rejoignent pour former le chiasma optique. À cet endroit, les fibres provenant de l'hémi-champ visuel gauche (respectivement droit) forment le tractus optique qui mène à l'hémisphère droit (respectivement gauche). Avant d'atteindre le cortex

visuel, les tractus optiques se projettent sur le corps genouillé latéral (CGL) dans le thalamus.

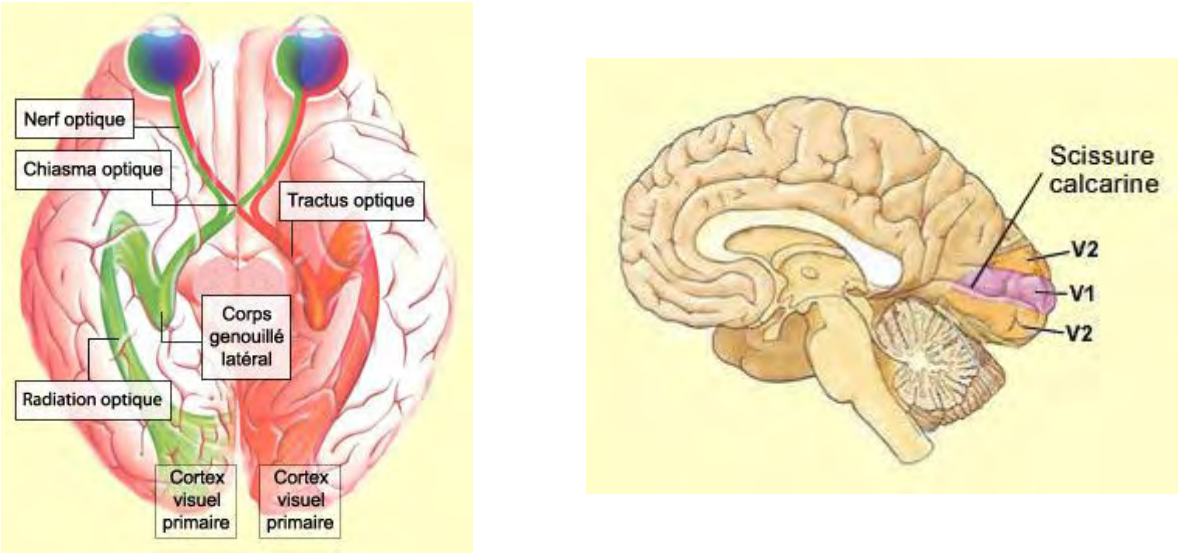


Figure 1-3 Les différents relais visuels.
À gauche, les différents relais visuels de l'œil jusqu'au cortex. À droite, coupe sagittale laissant apparaître les aires V1 et V2 du cortex visuel (Source : Le cerveau à tous les niveaux http://lecerveau.mcgill.ca/).

Le cortex comme interpréteur

Le cortex visuel primaire se trouve dans la partie postérieure du lobe occipital (Figure 1-3). Dans cette zone, le cortex visuel primaire (V1) reçoit directement l'information provenant des cellules du corps genouillé latéral. Un grand nombre des neurones de V1 se projettent dans le cortex visuel secondaire (V2). Ces étapes permettent une première interprétation des signaux visuels mais d'autres traitements complexes ont lieu dans le lobe temporal et le lobe pariétal.

UN PEU D'HISTOIRE

Un phosphène, du Grec *phos* (la lumière) et *phainein* (faire paraître), est une perception lumineuse qui peut être induite par stimulation mécanique, chimique, magnétique ou électrique. En 1896, le physicien français Arsène d'Arsonval, qui venait de créer 2 ans plus tôt l'École Supérieure d'Électricité, montra qu'il était possible d'évoquer des percepts visuels par stimulation magnétique transcrânienne [D'Arsonval 1896]. L'utilisation de l'électricité pour faire apparaître ces percepts est cependant une trouvaille plus ancienne. Il faut en effet remonter au XVIIIème siècle, en 1755, pour trouver trace des premiers écrits. Durant l'une de ses expériences,

Charles LeRoy, physicien et chimiste français, appliqua une décharge électrique à la surface oculaire d'un patient atteint d'amaurose. Ce dernier perçut alors des scintillements lumineux [LeRoy 1755]. En 1819, le neurophysiologiste tchèque Johannes Purkinje décrivit des phosphènes après avoir stimulé électriquement son front par le biais d'une électrode. Bien plus tard, en 1918, Löwenstein et Borchardt stimulèrent électriquement par accident le lobe occipital gauche d'un patient qu'ils étaient en train d'opérer. Le souffrant reporta des flashs lumineux dans son champ visuel droit [Löwenstein & Borchardt 1918]. D'autres cas identiques furent rapportés dans les années qui suivirent [Krause 1924; Foerster 1929]. En 1931, Krause et Schum stimulèrent électriquement le lobe occipital gauche d'un patient dont le champ visuel droit n'était plus fonctionnel depuis 8 ans [Krause & Schum 1931]. Il avait reçu une balle qui avait endommagé les radiations optiques gauches reliant le corps genouillé latéral au cortex visuel primaire. Le sujet reporta là aussi des flashs lumineux ce qui démontra que son cortex visuel était encore fonctionnel bien qu'il ait été privé de stimulus pendant plusieurs années, un résultat qui sera confirmé 30 ans plus tard [Button & Putnam 1962].

Les travaux sur la stimulation électrique du cortex visuel se sont intensifiés à partir des années 1950. Le neurochirurgien canadien Wilder Penfield l'utilisa pour notamment localiser les zones du cerveau impliquées dans la vision. Ses nombreuses expériences montrèrent qu'il était possible de générer des phosphènes en appliquant un courant électrique sur ces zones, et que ces perceptions étaient spatialement reproductibles [Penfield & Rasmussen 1950; Penfield & Jasper 1954].

Une fois acquise la possibilité de faire apparaître un phosphène par stimulation électrique du cortex visuel, l'étape suivante a été de savoir si la stimulation simultanée de différentes zones du cerveau par le biais d'un ensemble d'électrodes provoquerait l'apparition d'un ensemble cohérent de phosphènes. Avec l'idée, par la suite, de recréer artificiellement des motifs visuels, voire des images.

Les travaux de Brindley et Lewin sont aujourd'hui considérés comme précurseurs dans l'invention des neuroprothèses visuelles. En 1968, ils développent un premier prototype (Figure 1-4) et implantent 80 électrodes en platine, de 0.8 mm de diamètre, en contact avec la surface du cortex visuel droit d'une personne devenue aveugle six mois avant cette opération [Brindley & Lewin 1968]. Un dispositif extra-crânien

composé d'autant de récepteurs radio permet le contrôle individuel de chacune des électrodes.

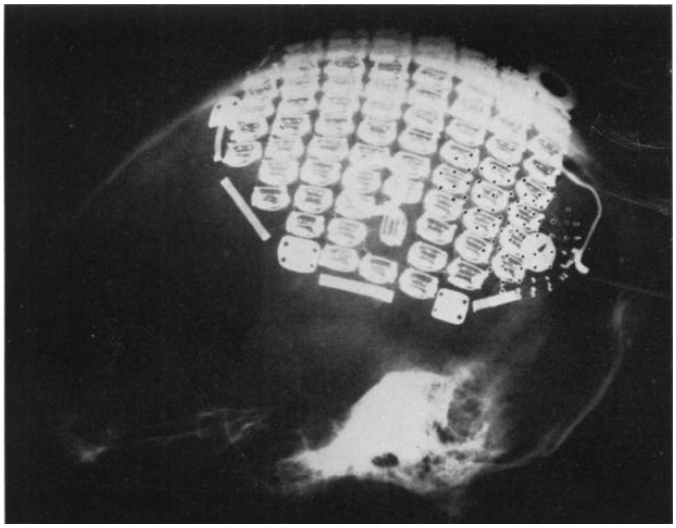


Figure 1-4 Prototype de neuroprothèse corticale de Brindley et Lewin.
Ce dispositif contient 80 électrodes de surface disposées sur le cortex occipital droit d'un patient aveugle. Chaque électrode est contrôlée par un récepteur radio (Source : Brindley & Lewin, 1968).

Deux conclusions essentielles résultent de cette expérience. La première conforte les résultats précédents : lorsqu'un courant est appliqué sur une seule électrode, le patient perçoit la plupart du temps un seul phosphène qui prend la forme d'une petite tâche lumineuse blanche et ce phosphène est reproductible dans sa localisation dans le champ de vision du sujet. De plus, si le patient bouge les yeux, le ou les phosphènes suivent ce mouvement. La deuxième conclusion permet une grande avancée pour ce champ de recherche : si un courant est appliqué simultanément à plusieurs électrodes, le sujet perçoit plusieurs phosphènes et ce motif visuel est ici aussi reproductible. D'autres conclusions non prévisibles sont rapportées. Les phosphènes générés scintillent et certaines électrodes induisent l'apparition de plusieurs phosphènes simultanément suivant la charge de courant injectée. D'autre part, si deux électrodes sont trop proches l'une de l'autre (séparées de moins de 3 mm), leur activation conjointe ne produit qu'un seul phosphène (de plus grande taille).

En 1974, Dobelle et Mladejovsky complètent ces résultats en faisant varier un ensemble de paramètres (taille et configuration des électrodes ; intensité du courant) sur 37 voyants volontaires [Dobelle & Mladejovsky 1974]. Là aussi ils montrent qu'une seule électrode peut provoquer l'apparition de plusieurs phosphènes et qu'un phosphène se déplace lorsque le sujet déplace son regard. Ils montrent également qu'en changeant l'amplitude ou la fréquence du courant électrique, il est possible de

modifier le niveau de luminance du phosphène généré. En revanche, à la différence de Brindley et Lewin, ils concluent qu'une stimulation continue ne permet pas de générer un phosphène dont l'apparence reste homogène dans le temps. Ils observent notamment que celui-ci peut même disparaître complétement au bout de 10 à 15 secondes de stimulation. C'est une découverte essentielle car elle implique de définir une stratégie de stimulation adéquate pour contrer ce phénomène d'adaptation.

Ces travaux de Brindley, Lewin et Dobelle sont à l'origine du développement des neuroprothèses visuelles. Dans les années 1980, on retrouve très peu d'études sur le sujet mais à partir des années 1990 et jusqu'à nos jours, les recherches se sont intensifiées d'abord chez l'animal, puis au travers d'essais cliniques chez l'homme. Outre le cortex visuel primaire, une stimulation au niveau de la rétine [Humayun et al. 1996] ou du nerf optique [Veraart et al. 1998] permet aussi d'évoquer des phosphènes. Les paragraphes suivants détaillent, par type d'implant, les différents travaux de recherche qui ont vu le jour.

LES NEUROPROTHESES CORTICALES

Dans cette catégorie de neuroprothèses, on doit distinguer deux approches : les implants à la surface du cortex [Brindley & Lewin 1968; Dobelle & Mladejovsky 1974], et ceux dits intra-corticaux, qui utilisent des électrodes pénétrantes. La seconde approche à l'avantage de requérir un courant électrique d'une intensité beaucoup plus faible pour générer des phosphènes [Schmidt et al. 1996]. Quelle que soit la technologie choisie, la matrice est positionnée sur le cortex visuel primaire (V1). Deux raisons principales à ce choix : son organisation rétinotopique est supposée maitrisée (Figure 1-5) et l'espace y est suffisamment grand pour y placer beaucoup d'électrodes. La taille des phosphènes générés est fonction de leur excentricité : ils sont très petits dans la fovéa, et de plus en plus large en périphérie [Tehovnik & Slocum 2007; Brindley & Lewin 1968], ce qui est en concordance avec ce que l'on appelle le facteur de magnification corticale, illustré par la Figure 1-5.

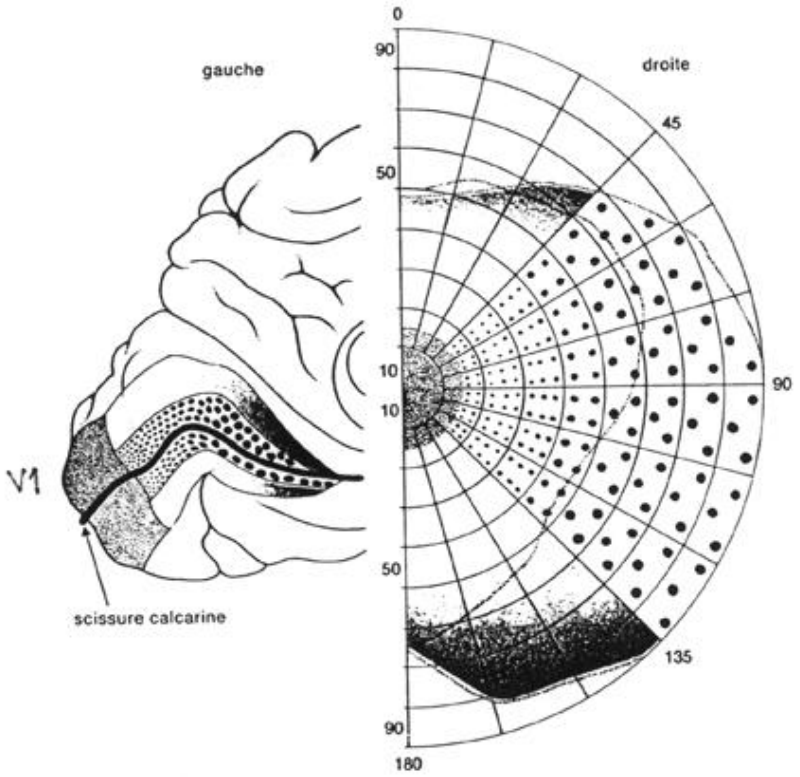


Figure 1-5 Organisation rétinotopique théorique de l'aire visuelle V1.
Un exemple ici de la projection du champ visuel droit sur l'aire visuelle primaire gauche (Source : http://acces.ens-lyon.fr/).

Bien que de nombreuses études aient porté sur la stimulation électrique du cortex visuel, il n'existe pour ainsi dire que le dispositif développé par William Dobelle qui soit un système complet de vision artificielle (par « complet » nous entendons capteurs de vision artificielle couplés à un implant cortical) et qui ait été implanté chroniquement chez l'homme. D'autres groupes comme le Monash Vision Group ont plus récemment porté leur intérêt dans le développement d'une telle prothèse corticale et envisagent dans un futur proche des essais cliniques chez l'homme.

Implants de surface

Dobelle Institute

En 2000, William Dobelle publie ses travaux pour ce qui sera considéré comme l'accomplissement d'une première neuroprothèse corticale fonctionnelle [Dobelle 2000]. Cet implant est constitué d'une matrice de 64 électrodes de surface et d'un paquetage externe incluant des câbles, une caméra et un micro-ordinateur. La matrice compte 8x8 électrodes de 1mm de diamètre. Les électrodes sont faites de

platine et disposées de façon hexagonale. La caméra noir et blanc est alimentée par une batterie de 9V. Elle a une résolution de 292x512 pixels pour un champ de vision de 69°. Elle est directement collée sur des lunettes (Figure 1-6). Cette caméra est reliée à un micro-ordinateur (120MHz, 32 MB de RAM) qui est attaché à une ceinture. Sur la ceinture on retrouve également un microcontrôleur en charge de piloter les électrodes. Le micro-ordinateur et le microcontrôleur sont eux aussi alimentés par une batterie externe.

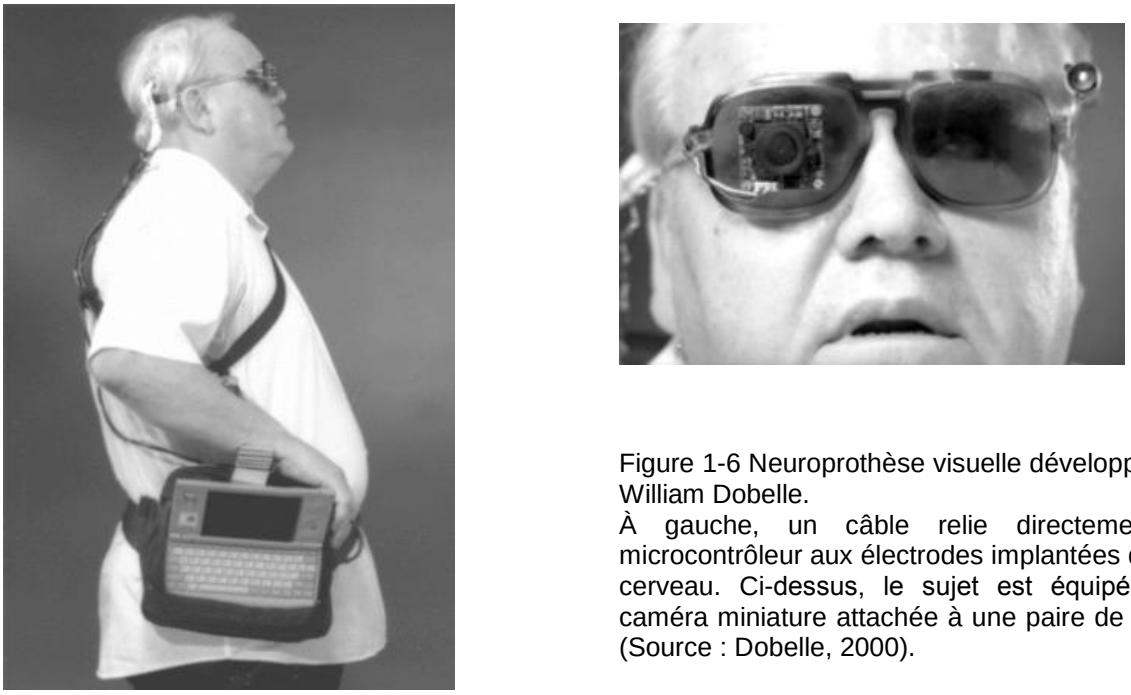


Figure 1-6 Neuroprothèse visuelle développée par William Dobelle.
À gauche, un câble relie directement un microcontrôleur aux électrodes implantées dans le cerveau. Ci-dessus, le sujet est équipé d'une caméra miniature attachée à une paire de lunette (Source : Dobelle, 2000).

D'un point de vue fonctionnel, chaque image acquise par la caméra est traitée par un programme du micro-ordinateur afin d'en réduire la résolution pour en obtenir une image de 8x8 pixels. Le microcontrôleur se charge de transformer cette image en trains de pulsations électriques (6 pulsations de 1 milliseconde chacune, à une fréquence de 30 Hz) qu'il délivre aux 64 électrodes. L'image est ainsi restituée sous forme d'un motif de phosphènes. Pour obtenir une utilisation optimale, l'ordinateur ne traite que 4 images par seconde.

Avant une utilisation optimale du système, il est nécessaire de cartographier la position de chaque phosphène. L'idée ici est d'appliquer un courant électrique, électrode par électrode, pour d'une part définir l'intensité nécessaire à la génération d'un phosphène et d'autre part localiser ce phosphène dans le champ de vision du patient. Le plus célèbre patient de cette expérience, connu sous le nom de « Jerry »,

est implanté en 1978. Il a alors 41 ans et a perdu la vue 5 ans auparavant. Après plusieurs jours d'utilisation du système, Jerry est capable de distinguer certains caractères (noirs sur fond blanc) de 15 centimètres de haut à une distance de 1,5 mètre. Après avoir soumis ce résultat, Dobelle améliore dans les mois qui suivent son système : un micro-ordinateur plus puissant traite désormais 8 images par seconde. La capacité de son ordinateur lui permet d'imaginer de nouveaux rendus. Il utilise les filtres de Sobel [Sobel 1970] pour ne restituer que les contours détectés dans l'image en entrée. Il imagine également ajouter un télémètre à ultrasons pour obtenir et restituer des informations de profondeur en modifiant par exemple l'apparence de certains phosphènes (luminosité, clignotement).

Bien que révolutionnaires et avant-gardistes, ces travaux sont très controversés. Premièrement, William Dobelle, qui avait créé le Dobelle Institute à New-York en 1980, a dû quitter les États-Unis en 1983 pour continuer à mener ses expériences à Lisbonne au Portugal. En effet, la Food and Drug Administration (FDA) n'avait pas donné son accord pour l'implantation de prothèses corticales. Deuxièmement, très peu de données scientifiques sont disponibles à ce jour sur l'évaluation de son système. Pourtant entre 2002 et 2004, seize autres patients vont recevoir cet implant d'un coût unitaire de 80000$. Jens Naumann est le premier de cette série. D'abord patient puis consultant pour le Dobelle Institute, Naumann délivre finalement l'information la plus exhaustive sur ce système dans son livre publié très récemment [Naumann 2012]. On y apprend notamment que le dispositif a eu un réel effet bénéfique chez certains sujets. En revanche, la plupart des patients ne voyaient plus de phosphènes quelques mois après leur opération. Suite au décès de William Dobelle survenu en 2004, aucun groupe à ce jour n'a entrepris de reprendre et poursuivre ses travaux.

Implants intra-corticaux

National Institute of Health (États-Unis)

Ce groupe créé au début des années 1990 était dirigé par le docteur Edward Schmidt. Il est le premier à avoir obtenu des résultats chez l'homme, sur la stimulation du cortex visuel primaire à l'aide d'électrodes pénétrantes [Bak et al. 1990]. Il montre que pour faire apparaître un phosphène, il suffit d'appliquer un

courant électrique, d'une intensité comprise entre 20 à 200 µA en fonction de la profondeur de l'électrode (entre 3 et 5 mm). Cette intensité est 10 fois moins importante que celle requise pour les électrodes de surface. La résolution spatiale est aussi améliorée : une distance entre électrodes de 0,7 à 1 mm suffit à générer des phosphènes distincts (contre 3 mm en surface). Cette distance correspond à l'espacement centre à centre des électrodes.

Avec ces premiers résultats encourageants, le groupe procède à une seconde expérience et implante, pour une période de 4 mois, une matrice de 38 électrodes intra-corticales à une femme de 42 ans atteinte d'un glaucome [Schmidt et al. 1996]. 34 des 38 électrodes produisent des phosphènes. L'intensité nécessaire pour les générer est le plus souvent en dessous de 20 µA. L'équipe conclut également que la luminosité des phosphènes est fonction de l'intensité, de la fréquence et de la durée de la stimulation. D'autre part, le sujet observe 2 phosphènes distincts lors de l'activation simultanée d'électrodes séparées d'à peine 500 µm, ce qui est 5 fois plus proche de ce qui avait été observé pour des électrodes de surface. La principale contrepartie par rapport aux électrodes de surface est un phénomène d'adaptation rapide à la stimulation : après 50 stimulations consécutives, l'intensité requise pour produire des phosphènes augmente en moyenne de plus de 50% et la luminosité observée baisse brutalement. Une période de repos de quelques secondes est alors nécessaire pour observer un retour à la normale.

Université d'Utah (États-Unis)

Depuis plus de 20 ans, ce projet, sous la conduite de Richard Normann, vise à développer une neuroprothèse intra-corticale. À ce jour, ils ont développé et testé l'implantation d'une matrice de 4x4 mm contenant 100 électrodes (Figure 1-7). Les électrodes sont longues de 1,5 mm et séparées de 400 µm [Normann et al. 1999].

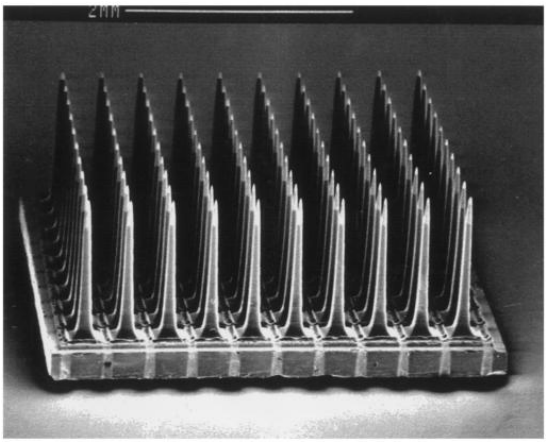


Figure 1-7 Matrice UEA (Utah Electrode Array). **(**Source : Normann et al. 1999)

Classiquement, le système envisagé transformera des images acquises par une micro-caméra en impulsions électriques délivrées à la matrice d'électrodes intra-corticales. Bien que des études aient été conduites chez le chat et le singe [Normann et al. 2009], aucun essai clinique n'est pour l'instant annoncé[4].

École Polytechnique de Montréal (Canada)

Le laboratoire de neurotechnologies PolySTIM de l'école polytechnique de Montréal a été fondé en 1994 par Mohamad Sawan. Son équipe a proposé également une solution d'implant intra-cortical [Coulombe et al. 2007]. Le système est composé d'une caméra et d'un processeur externe qui traite les images et transmet les instructions sans fil (alimentation et données) à l'implant (Figure 1-8). Le projet initial consistait à utiliser une unique matrice de 25x25 électrodes. Dans la solution actuelle, l'implant contient un ensemble de matrices plus petites (4x4 électrodes). Cette architecture permet de répartir les électrodes sur une plus grande surface, et de suivre la courbure du cortex visuel. Actuellement, les électrodes utilisées ont un diamètre de 50 µm, mesurent 1,5 mm, et sont séparées de 400 µm. L'équipe doit se concentrer prochainement sur les premières expérimentations in vivo.

[4] https://clinicaltrials.gov/

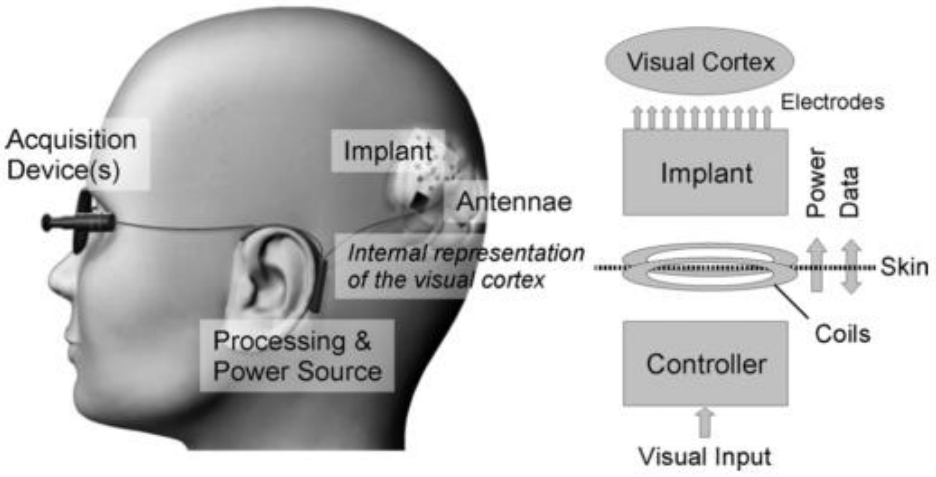


Figure 1-8 Neuroprothèse intra-corticale imaginée par PolySTIM.
L'acquisition de l'image se fait par l'intermédiaire d'une caméra externe. Un contrôleur s'occupe de transformer les images et les envoyer par ondes radio à l'implant (Source : Coulombe et al. 2007).

Monash Vision Group (Australie)

Monash Vision Group est un consortium Australien créé en 2010 ayant pour objectif de développer une neuroprothèse corticale (http://www.monash.edu.au/). Il regroupe des universitaires, des cliniciens et des ingénieurs du secteur privé. Le dispositif envisagé contient une caméra portée sur des lunettes et un microprocesseur capable de traiter le signal d'entrée (Figure 1-9). Aucune communication n'a été faite sur les éventuels traitements qui seront appliqués à l'image. Le signal visuel et l'alimentation seront transmis sans fil à un ensemble de contrôleurs (jusqu'à 11) implantés dans le cortex visuel. Chaque contrôleur pilotera 43 microélectrodes pénétrantes. L'implant pourra ainsi inclure jusqu'à 473 électrodes. Le groupe planifie les premiers essais cliniques chez l'homme en 2015.

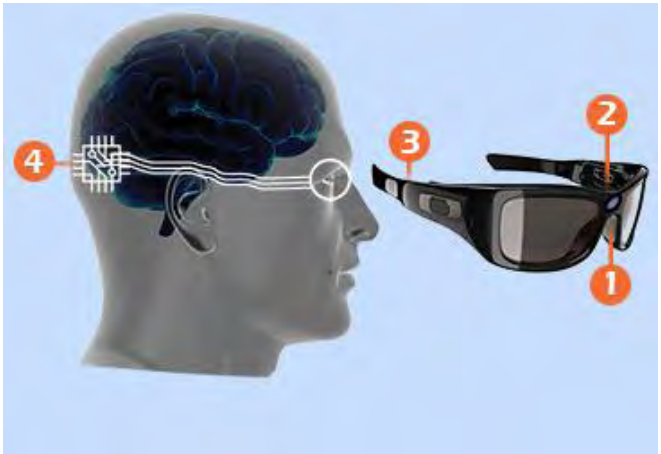


Figure 1-9 Neuroprothèse intra-corticale envisagée par le Monash Vision Group.
1) Caméra au centre des lunettes. 2) Composant permettant d'asservir la caméra aux mouvements des yeux. 3) Micro-processeur et antenne. 4) Contrôleurs implantés (Source : http://www.monash.edu.au/).

Autres projets

Sur le même modèle que la neuroprothèse corticale Australienne, on peut citer deux autres projets à l'état préclinique : celui de l'Illinois Institue of Technology aux États-Unis, dirigé par Philipp Troyk [Troyk et al. 2003] démarré au début des années 2000, et le projet européen CORTIVIS[5] mené par l'espagnol Eduardo Fernández [Fernández et al. 2005].

LES NEUROPROTHESES DU NERF OPTIQUE

Deux technologies d'implants sont en concurrence pour s'interfacer avec les axones des cellules ganglionnaires qui constituent le nerf optique. La première utilise une électrode spirale qui s'enroule le long du nerf optique et la seconde s'appuie sur une matrice d'électrodes pénétrantes.

Veraart (Belgique)

L'équipe de Claude Veraart, de l'université Catholique de Louvain à Bruxelles, a mis au point dans les années 1990 un implant constitué d'une électrode à manchon spiral avec 4 points de contact et qui s'enroule le long du nerf optique. Dans le cadre du projet européen « Microsystems based Visual Prosthesis » (MiViP), ils ont imaginé une neuroprothèse visuelle basée sur cette électrode. Une femme de 59 ans, atteinte de rétinite pigmentaire, participe en 1998 aux premiers essais précliniques [Veraart et al. 1998]. En faisant varier les paramètres de stimulation (amplitude, temps, fréquence et nombre de pulsations par stimulus), ils répertorient plus de 1000 phosphènes évoqués chez la volontaire, ce qui fera l'objet d'une modélisation [Delbeke et al. 2003]. Comme l'illustre la Figure 1-10, chaque contact est plus ou moins localisé dans un quadrant du champ de vision.

[5] http://cortivis.umh.es/

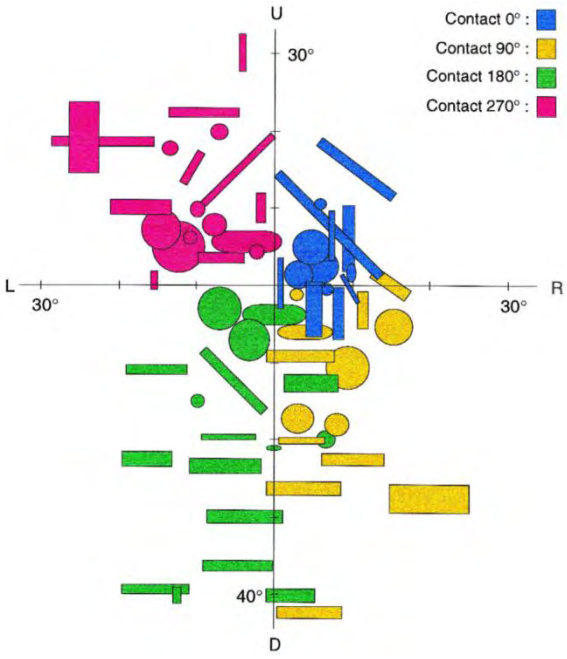


Figure 1-10 Carte des phosphènes évoqués par stimulation du nerf optique.
Exemple d'un sujet avec, pour chaque contact, 16 stimuli différents. (Source : Veraart et al. 1998)

Suite à ces travaux, l'équipe publie en 2005 les résultats d'une autre expérience avec le même sujet [Brelén et al. 2005]. Cette fois-ci, ils connectent une caméra en noir et blanc (25 images par seconde, 108° de champ de vision) au dispositif. Sur le millier de phosphènes qu'il est possible d'évoquer, ils n'en génèrent que 109, ceux considérés comme les plus stables et nécessitant une faible intensité électrique. Avec cette centaine de phosphènes distincts, le sujet doit reconnaître des motifs noirs sur fond blanc filmés par la caméra. Comme il n'est pas possible de faire apparaître plus d'un phosphène simultanément, la stratégie choisie est d'allumer aléatoirement, toutes les 80 ms, l'un des phosphènes qui coïncide avec le pattern filmé. Après une longue période d'entraînement, le sujet est capable d'identifier une vingtaine de motifs (85% correct) pour un temps de reconnaissance d'environ 1 minute par motif.

C-Sight (Chine)

Le projet C-Sight est le premier projet chinois de création d'une neuroprothèse visuelle [Chai et al. 2008]. Leur idée est de s'interfacer au nerf optique par le biais d'électrodes pénétrantes. L'approche semble intéressante car elle permettrait, contrairement au projet belge, l'évocation simultanée de plusieurs phosphènes [Sun et al. 2011]. À ce jour, aucun essai chez l'homme n'a été réalisé.

AV-DONE (Japon)

Tout comme C-Sight, le projet AV-DONE a pour objectif de stimuler le nerf optique avec des électrodes pénétrantes. En 2009, ils valident leur approche lors d'un essai préclinique : après avoir été implanté avec 3 électrodes au niveau du disque optique (point d'insertion du nerf optique dans l'œil), un sujet atteint de rétinite pigmentaire discerne des phosphènes [Sakaguchi et al. 2009]. Cette opération est renouvelée avec succès chez un autre sujet 6 mois après la première opération.

LES NEUROPROTHESES RETINIENNES

Les neuroprothèses visuelles qui connaissent les plus grands succès à ce jour sont celles qui stimulent la rétine. Elles sont destinées aux personnes qui souffrent de pathologies impliquant un dysfonctionnement des cellules photoréceptrices (la rétinite pigmentaire ou la dégénérescence maculaire liée à l'âge par exemple). Dans ces systèmes, la matrice d'électrodes peut être fixée à différents emplacements. Lorsqu'elle est en contact avec les cellules ganglionnaires, on parle d'implant épirétinien. Lorsqu'elle est insérée sous la rétine, en contact avec les cellules bipolaires, on parle d'un implant sous-rétinien. Enfin dans l'implant suprachoroïdien, la matrice est fixée sous la sclère, contre la choroïde.

Dans la fabrication de ces systèmes, deux approches sensiblement différentes se distinguent [Dowling 2009]. La première est basée sur l'optoélectronique. L'idée est d'utiliser des micro-photodiodes en charge de transformer, sous forme d'impulsions électriques, la lumière arrivant sur la rétine, en remplacement des cellules photoréceptrices endommagées. Dans cette solution, l'implant est en majorité du temps sous-rétinien. Dans la seconde approche, le courant électrique n'est pas généré directement par la lumière mais contrôlé par un stimulateur capable de piloter individuellement chacune des électrodes. Dans cette solution, c'est une caméra portée sur des lunettes qui capte l'information visuelle. Celle-ci est traitée puis transmise au stimulateur. Dans cette catégorie, l'implant peut être épirétinien, sous-rétinien ou suprachoroïdien (Figure 1-11).

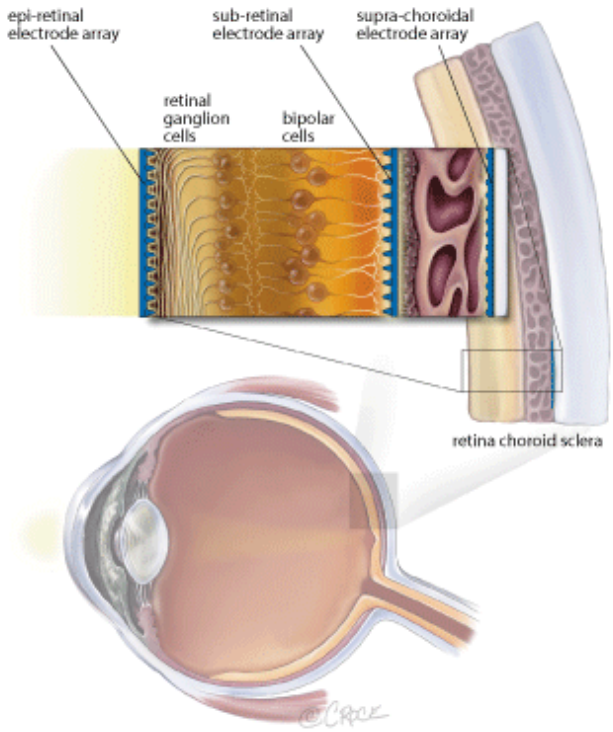


Figure 1-11 Les catégories d'implants rétiniens.
Ces différentes catégories sont dépendantes de la zone d'implantation de la matrice (Source : www.optometrists.asn.au).

Systèmes basés sur une caméra externe

Second Sight (États-Unis)

La société américaine Second Sight développe depuis le début des années 2000 une neuroprothèse épirétinienne. À sa tête, Mark Humayun est un des premiers à avoir étudié les microstimulations de la rétine et à avoir prouvé la faisabilité d'une telle approche [Humayun et al. 1999; Humayun et al. 1996]. Les composants de leur système sont illustrés dans la Figure 1-12.

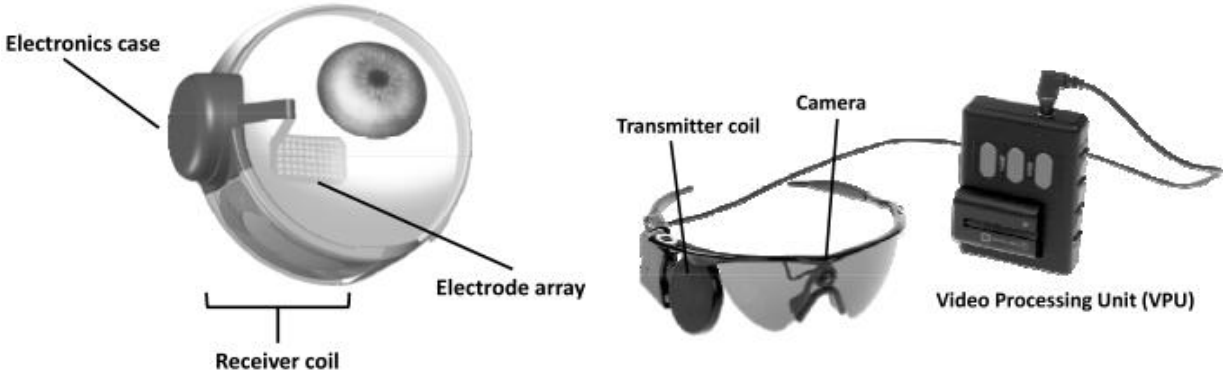


Figure 1-12 L'implant épirétinien développé par la société Second Sight.
Exemple ici avec l'Argus II, la seconde génération de leur système. À gauche, l'implant est constitué d'une matrice d'électrode fixée sur la rétine, d'un stimulateur (electronic case) en charge d'injecter du courant électrique dans les électrodes et d'un récepteur sans fil. À droite, la partie externe du système avec une petite caméra fixée à des lunettes, un contrôleur vidéo (VPU) et un transmetteur (Source : Zhou et al. 2013).

Deux versions de ce système ont vu le jour jusqu'à présent. La première est l'Argus I, dont la matrice contient 16 électrodes en platine (250-500 µm de diamètre) disposées de façon hexagonale, et séparées de 800 µm [Caspi et al. 2009]. Le stimulateur se situe derrière l'oreille. Cette version a été testée entre 2002 et 2004 sur 6 personnes atteintes de rétinite pigmentaire. Certains sujets étaient capables de détecter des mouvements et ont vu leurs performances améliorées dans des tâches de localisation et de discrimination d'objet [Yanai et al. 2007].

La seconde génération de l'implant, l'Argus II apporte des modifications importantes [Zhou et al. 2013]. La matrice est désormais composée de 60 (6x10) électrodes de 200 µm de diamètre séparées de 525 µm. Elle couvre environ 20° de champ de vision. Le stimulateur n'est plus installé derrière l'oreille mais directement fixé à la surface de l'œil (Figure 1-12). Entre 2007 et 2009, une campagne internationale (10

centres) d'essais précliniques est lancée à laquelle participent 30 patients [Humayun et al. 2012]. Plusieurs études récentes montrent une amélioration importante de la performance des sujets dans des tâches simples de localisation et de pointage [Ahuja et al. 2011] et de détection de mouvement [Dorn et al. 2012]. Certains sujets sont aussi capables de lire de grandes lettres blanches sur fond noir [da Cruz et al. 2013]. La meilleure acuité visuelle mesurée chez un sujet est estimée à 6/380.

Suite à l'agrément de la Food and Drug Administration (FDA) américaine et de son homologue européenne (marquage CE), l'Argus II est le premier implant rétinien à être autorisé à la vente aux États-Unis et en Europe pour un coût proche des 100 000 €. Il est prévu qu'à partir de cette année (2014), 36 personnes par an puissent bénéficier en France d'un remboursement par la sécurité sociale, suite à sa prise en charge dans le cadre du « forfait innovation »[6].

Boston Retina Implant Project (États-Unis)

Après avoir étudié pendant plus de 10 ans la faisabilité d'un implant épirétinien [Rizzo III et al. 2003b; Rizzo III et al. 2003a], Joseph Rizzo III et John Wyatt concentrent désormais leurs efforts sur une neuroprothèse sous-rétinienne [Shire et al. 2009; Rizzo III 2011]. Leur idée principale est de simplifier au maximum l'acte chirurgical. Ici seule la matrice d'électrode est à insérer dans l'espace sous-rétinien (Figure 1-13). Le stimulateur, placé à la surface de l'œil, est encapsulé dans un petit boîtier hermétique en titane. L'alimentation et les signaux proviennent de composants externes et sont transmis sans fil. Après avoir validé un prototype avec 15 électrodes chez l'animal, ils planifient des essais cliniques chez l'homme avec un système capable de piloter 200 électrodes.

[6] Arrêté ministériel du 4 Août 2014 (http://www.legifrance.gouv.fr/).

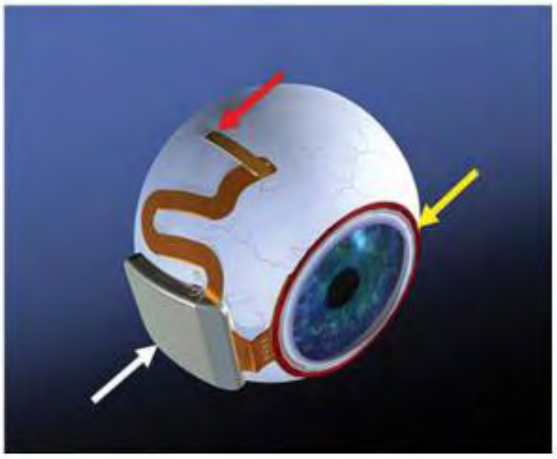


Figure 1-13 L'implant de seconde génération du Boston Retinal Implant Project.
La matrice d'électrode est insérée dans l'espace sous-rétinien (elle plonge dans l'œil sous la flèche rouge). L'alimentation et les données sont transmises sans fil (flèche jaune). La puce qui pilote la matrice d'électrode est encapsulée dans un boîtier étanche (flèche blanche) (Source : Rizzo III 2011).

IMI Gmbh (Allemagne) puis Pixium Vision (France)

La société allemande Intelligent Medical Implant a développé une neuroprothèse épirétinienne sur un modèle équivalent à celui de Second Sight. La particularité de leur système réside dans l'utilisation d'un encodeur (Retinal Encoder) permettant à l'utilisateur d'ajuster les paramètres de stimulation. L'implant, l'IRIS1, contient une matrice de 49 électrodes. Une étude démarrée en 2003, sur 20 sujets atteints de rétinite pigmentaire, montre que le dispositif permet de générer des phosphènes de différentes formes et de différentes couleurs [Hornig et al. 2008].

IMI a été rachetée en 2012 par la société française Pixium Vision, fondée en décembre 2011 par l'équipe du professeur José-Alain Sahel de l'institut de la vision à Paris. Celle-ci a lancé une campagne de recrutement dans trois centres européen (France, Allemagne et Autriche) pour tester cliniquement l'implant IRIS1 auprès d'une vingtaine de personne. Les premiers résultats sont attendus courant 2014. Pixium Vision projette également le développement de l'implant seconde génération, l'IRIS2, portant le nombre d'électrodes à 150.

Epiret GmbH (Allemagne)

Cet autre groupe allemand développe une neuroprothèse épirétinienne, EPIRET3, qui intègre une matrice constituée de 25 électrodes en or, de 100 µm de diamètre, séparées chacune de 500 µm, le tout couvrant à peu près 10° de champ visuel. Là aussi l'architecture du dispositif est identique à celle proposée par Second Sight. En 2006, 6 sujets atteints de rétinite pigmentaire participent à des tests précliniques pendant 4 semaines. Les résultats confirment ceux des autres solutions

épirétiniennes proposées : les sujets rapportent qu'ils perçoivent des phosphènes et qu'ils sont capables de discerner des motifs de points [Klauke et al. 2011]. Ils valident également la procédure d'implantation et d'explantation de l'implant [Roessler et al. 2009]. Un site regroupant la liste des projets en cours semble cependant indiquer que la société est désormais fermée[7].

Bionic Vision Australia (Australie)

Le consortium australien Bionic Vision Australia (BVA) comprend deux programmes de développement d'implant rétinien dirigés par le professeur Anthony Burkitt. Le premier projet (Wide acuity device) est dirigé par Gregg Suaning et consiste à créer un implant suprachoroïdien constitué de 98 électrodes. Des tests chez le chat ont validé l'approche [Villalobos et al. 2013]. Un premier prototype constitué de 24 électrodes a été développé et implanté en 2012 chez trois sujets atteints de rétinite pigmentaire. Les premiers résultats sont attendus courant 2014. Hamish Meffin est aux commandes du deuxième projet (High acuity device). Ce système comprendra un implant épirétinien de 256 électrodes en diamant et sera testé chez l'homme dans les deux à trois ans à venir.

Autres projets

Deux projets japonais sont aussi à mentionner. Le premier, en collaboration avec la société Nidek, développe un implant suprachoroïdien [Morimoto et al. 2011]. Le second, mené par Hiroyuki Kurino and Hiroshi Tomita de l'université de Tohoku, a choisi une approche épirétinienne.

[7] http://www.eye-tuebingen.de/zrenner/retimplantlist/

Systèmes basés sur l'optoélectronique

Optobionics Corporation (États-Unis)

Le premier implant utilisant l'optoélectronique, nommé Artifical Silicon Retina (ASR), a été créé et testé cliniquement par Alan et Vincent Chow au sein de leur société Optobionics [Chow et al. 2004]. Leur système consiste en une micro-puce (2 mm de diamètre et 25 µm d'épaisseur) contenant 5000 micro-photodiodes passives (c'est-à-dire alimentées uniquement par la lumière) et directement placée sous la choroïde. Une électrode en or est associée à chacune de ces petites cellules photovoltaïques. La lumière naturelle est captée par les photodiodes qui la transforment directement en impulsions électriques délivrées aux électrodes.

La faisabilité d'un tel système avait tout d'abord été validée chez le chat (Chow et al., 2001). Puis à partir de 2000, suite à l'accord obtenu auprès de la FDA, il est implanté chez 6 patients atteints de rétinite pigmentaire [Chow et al. 2004]. Les sujets rapportent des améliorations dans leur champ de vision à la fois dans la zone stimulée mais aussi dans des zones non stimulées. Il est rapidement prouvé que les photodiodes passives produisent un courant électrique trop faible pour activer les cellules bipolaires et donc générer des phosphènes [Zrenner 2002]. Les effets bénéfiques observés seraient dus à des facteurs neurotrophiques libérés par les cellules nerveuses lors des stimulations électriques sous-liminaires. La société sera liquidée en 2007 [Dowling 2009].

Retina Implant AG (Allemagne)

Le professeur Eberhart Zrenner du Eye Hospital de Tübingen est à la tête d'un consortium allemand créé en 1995 pour développer un implant rétinien. Tout comme les frères Chow, leur système est basé sur une matrice de micro-photodiodes. Les tests *in vitro* et *in vivo* sur leur premier prototype sont concluants mais Zrenner mentionne que, pour fonctionner correctement, le signal électrique en sortie des photodiodes doit être amplifié et qu'une alimentation externe de l'implant est donc nécessaire [Zrenner et al. 1999].

En 2003, naît de ce consortium la société Retina Implant AG pour continuer le développement et commercialiser un implant sous-rétinien fonctionnel. Ce dispositif

(Figure 1-14), tel qu'il existe aujourd'hui [Zrenner et al. 2009], est constitué d'une puce photosensible de 3x3 mm, épaisse de 70 µm et contenant 1500 cellules. Chaque cellule contient une photodiode (15x30 µm), un amplificateur et une électrode en iridium de 50 µm de diamètre. L'espace entre les électrodes est de 70 µm.

Dans le prolongement de cette puce, une petite matrice (1,2 x 1,2 mm) contient 16 plots, de 4 électrodes chacun, stimulables individuellement. Les électrodes sont identiques à celles montées sur la puce. Dans un plot, les quatre électrodes sont espacées de 20 µm. Les plots sont distants de 280 µm latéralement et 396 µm diagonalement. Un câble ressort derrière l'oreille et connecte la puce et cette petite matrice d'électrodes à un contrôleur externe. Ce dernier fournit l'alimentation nécessaire à la puce et permet de stimuler chacun des 16 plots.

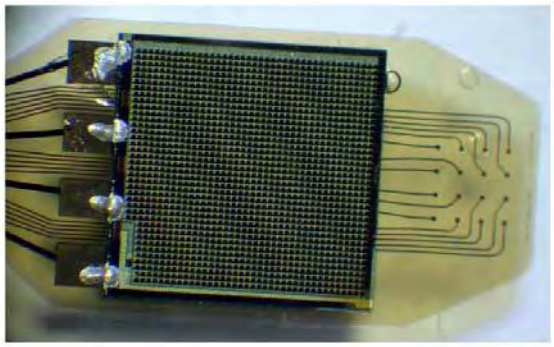


Figure 1-14 L'implant sous-rétinien développé par la société Retina Implant AG.
Au centre se trouve la puce de 3mm de diamètre sur laquelle sont gravées les 1500 cellules composées de photodiodes et de micro électrodes. Sur la droite se trouve la petite matrice contrôlable par stimulation directe et ses 16 électrodes.

Onze patients atteints pour la majorité de rétinite pigmentaire ont participé à une campagne d'essais cliniques qui a duré 4 semaines [Zrenner et al. 2009]. Les tests de contrôle sur la petite matrice d'électrode montrent que la stimulation directe engendre bien la perception de phosphènes. Certains sujets sont capables de distinguer des motifs visuels, notamment des lettres comprises entre 5 et 8° d'angle visuel. Lorsque la puce photosensible est activée, seules 6 des 11 personnes voient des phosphènes. Des tests plus poussés ont été effectués sur trois d'entre eux [Zrenner et al. 2011]. Les sujets sont capables de localiser, voire de nommer des objets clairs posés sur une table sombre.

Une version améliorée de l'implant, nommé Alpha-IMS, est testée cliniquement dans une nouvelle campagne internationale regroupant plusieurs centres [Stingl et al. 2013]. Dans cette version de l'implant, les signaux et l'alimentation sont transmis

sans fil. Les premiers résultats sont sensiblement identiques à ceux obtenus lors de la phase pilote. Lorsque la puce est activée, la plupart des sujets peuvent localiser une source de lumière (8/9) et détecter des mouvements (6/9). Trois des sujets lisent des lettres fortement contrastées d'une hauteur comprise entre 5 et 10°. Seuls deux des sujets sont capables de passer le test d'acuité visuelle. Le premier obtient 6/600 et le second 6/160 ce qui correspond à la meilleure des acuités rapportée à ce jour, toutes neuroprothèses visuelles confondues. En juillet 2013, l'Alpha IMS a reçu le marquage CE, ouvrant la voie à sa commercialisation prochaine.

Stanford University (États-Unis), Pixium Vision (France)

L'équipe de Daniel Palanker, de l'université de Stanford, développe un implant combinant optoélectronique et caméra externe. Cette solution projette une image par le biais de rayons infra-rouges pour activer des photodiodes [Loudin et al. 2007; Palanker et al. 2005]. La Figure 1-15 illustre le fonctionnement de ce système. Une petite caméra (640x480 pixels, 25-50Hz) attachée sur des lunettes capte la scène. L'image est traitée par un micro-ordinateur de la taille d'un téléphone puis affichée sur un écran LCD à l'intérieur des lunettes. L'écran est illuminé par lasers infra-rouges pulsés projetant directement sur la rétine une image d'une taille d'environ 30°. L'implant peut être constitué de plusieurs modules, chacun contenant un ensemble de cellules photovoltaïques passives qui convertissent cette énergie lumineuse en courant électrique alimentant directement les électrodes. L'illumination laser des photodiodes est suffisamment puissante pour que le courant électrique injecté dans les électrodes induise des potentiels d'actions dans les structures rétiniennes. Un module mesure environ 1x1 mm et peut contenir jusqu'à 300 couples photodiodes/électrodes.

En 2013, Daniel Palanker est entré dans le comité scientifique de la société française Pixium Vision, qui souhaite commercialiser un implant sous-rétinien s'inspirant de cette idée. Le système, nommé PRIMA, apportera des modifications à celui de Palanker comme l'utilisation d'électrodes 3D ou encore d'une caméra asynchrone. Le nouveau dispositif est prévu d'être validé et testé in-vivo d'ici fin 2014. Les premiers essais cliniques chez l'humain pourraient alors démarrer.

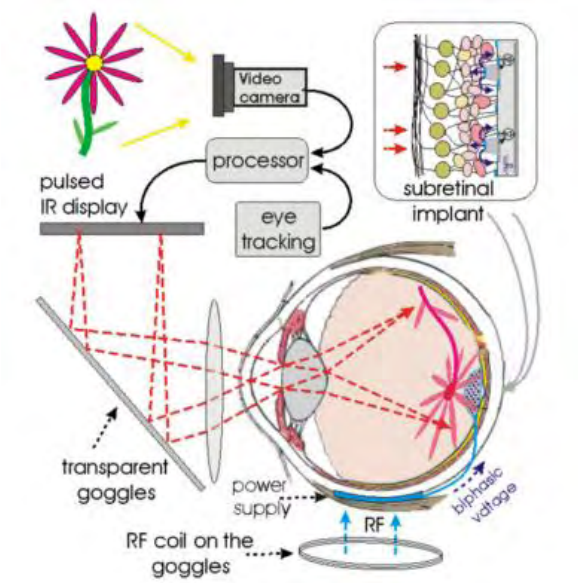


Figure 1-15 L'implant sous-rétinien développé à l'université de Stanford.
L'originalité de cette solution provient de l'utilisation de rayons infra-rouges pulsés captés et transformés en courant électrique par un ensemble de photodiodes placées dans l'espace sous-rétinien (Source : Loudin et al., 2007)

Nano Retina (Israël)

Bio-Retina est l'implant épirétinien que développe la société Israélienne Nano Retina. Dans l'idée, il est identique au système de Retina Implant, à la différence que la source d'énergie nécessaire aux photorécepteurs est fournie depuis les lunettes, par l'intermédiaire d'un faisceau infra-rouge. Ils promettent une opération chirurgicale sous anesthésie locale réduite à 30 minutes, en se limitant à l'insertion de l'implant dans le globe oculaire. La puce de première génération (3x4 mm) devrait contenir 24x24 (576) nano-électrodes puis près de 2000 pour la version suivante. Très peu de données techniques sont accessibles, hormis les informations provenant de leur site internet (et de leur plaquette commerciale) et le fait qu'ils soient dans une phase d'implants précliniques.

Autres solutions

Dans les autres projets en cours, nous pouvons citer un projet coréen démarré au début des années 2000 et porté par Jong-Mo Seo de l'université de Séoul [Kim et al. 2009] dont la spécificité est l'utilisation d'électrodes 3D.

DISCUSSION

La stimulation électrique de certains relais nerveux du système visuel induit des percepts lumineux, appelés phosphènes. Cette découverte est à l'origine des premiers travaux sur les neuroprothèses visuelles à la fin des années 1960. Ces systèmes ont pour objectif de restaurer partiellement la vision à des personnes non voyantes, par le biais de ces phosphènes. Les avancées technologiques des vingt dernières années, notamment la miniaturisation de l'électronique et les avancées sur la biocompatibilité des matériaux, ont permis à de nombreux systèmes d'être développés. Suivant le site de stimulation, les neuroprothèses sont qualifiées de rétiniennes, corticales ou du nerf optique.

Avantages et inconvénients des différentes solutions

Comparées aux autres types d'implants, les neuroprothèses corticales peuvent potentiellement couvrir un spectre de pathologies plus larges. Des maladies comme le glaucome ou la rétinopathie diabétique, qui détruisent trop profondément la rétine et le nerf optique, ne peuvent pas être traitées avec des implants rétiniens par exemple. Les systèmes composés d'implants de surface ont été pour le moment abandonnés du fait des trop grands risques de crises épileptiques engendrées par l'intensité électrique nécessaire à la génération de phosphènes. Aujourd'hui, toutes les solutions en cours de développement intègrent un implant intra-cortical dont la ou les matrices d'électrodes sont positionnées dans l'aire visuelle primaire (V1). Cette aire présente l'avantage d'être une zone relativement étendue dans laquelle il est envisageable d'implanter des centaines d'électrodes. En revanche plusieurs raisons freinent l'évolution des neuroprothèses corticales. Premièrement, il est difficile de placer des électrodes dans la zone correspondant à la vision centrale, car celle-ci est enfouie dans un sillon (scissure calcarine). De plus, la stimulation directe des neurones du cortex visuel court-circuite les différents traitements effectués normalement par la rétine et le corps genouillé latéral. Bien qu'il n'y ait aucune étude approfondie sur ce sujet, on peut supposer que l'absence de prétraitements peut engendrer des problèmes perceptifs. L'organisation rétinotopique très particulière de V1 pose également le problème de la reconstruction de l'image sous forme de phosphènes : contrairement aux implants rétiniens, la disposition des phosphènes générés ne correspond pas à celle des électrodes. L'image reconstituée est donc

fortement dépendante de la cartographie des phosphènes (correspondance électrode/phosphène), qu'il est difficile de mettre en place chez un non-voyant. Enfin, c'est une chirurgie relativement lourde qui peut mener à d'éventuelles complications.

Le nerf optique contient une représentation compacte du champ de vision et les neuroprothèses qui le ciblent en tirent des propriétés très intéressantes. À ce jour, c'est avec ce type d'implants que le plus grand nombre de phosphènes distincts ont été évoqués (1465 exactement [Veraart et al. 1998]). De plus, le champ de vision restauré est aussi très large (60° en horizontal et 80° en vertical [Veraart et al. 1998]). En revanche, bien que les sujets distinguent plusieurs niveaux de luminance, ceux-ci ne sont pas contrôlables pour un même phosphène par modification des paramètres de stimulation, contrairement aux implants corticaux et rétiniens. Pour les implants utilisant une électrode à manchon spiral, un autre inconvénient majeur est qu'il est impossible d'évoquer plusieurs phosphènes simultanément. En effet dans cette solution, la diffusion d'un courant électrique dans différentes zones du nerf optique contribue à la génération d'un unique phosphène. La reconstitution d'une image apparaît dès lors moins évidente qu'avec les autres types d'implants, car elle se fait phosphène après phosphène. L'approche par électrodes pénétrantes pourrait résoudre ce problème. Enfin, l'organisation topique, par rapport au champ visuel est certainement celle, parmi les 3 emplacements de stimulation proposés, qui est la plus complexe et la moins étudiée [Eiber et al. 2013].

La majorité des essais cliniques en cours portent sur les implants rétiniens. Ce sont actuellement les plus prometteurs. Les trois types d'implants (épirétinien, sous-rétinien, suprachoroïdien) ont leurs avantages et leurs inconvénients. Les implants suprachoroïdiens simplifient au maximum l'acte chirurgical. L'espace est facile à atteindre pour y insérer la matrice d'électrodes, qui est mécaniquement maintenue entre la choroïde et la sclère [Villalobos et al. 2013]. Dans l'implant épirétinien, un micro clou est nécessaire pour fixer la matrice dans la rétine, contre les cellules ganglionnaires. Avec l'implant sous-rétinien, l'opération peut engendrer des risques post-opératoires non négligeables. Les implants sous-rétiniens ont en revanche pour avantage d'être au plus proche des cellules bipolaires ; ils viennent donc plus directement en remplacement des cellules photoréceptrices endommagées. Le signal généré est plus proche d'une stimulation naturelle, car il utilise les premiers

relais restés fonctionnellement intacts. Contrairement aux implants suprachoroïdiens, les implants épirétiniens et sous-rétiniens ont l'avantage d'être placés plus proches des cellules qu'ils stimulent, ce qui implique une plus faible intensité électrique requise pour générer des phosphènes. C'est un paramètre important car cela permet d'envisager pour ces implants des matrices d'électrodes d'une plus grande densité.

Parmi les implants rétiniens, les solutions basées sur l'optoélectronique permettent de conserver le lien naturel entre le mouvement des yeux et la perception visuelle. Dans les solutions basées sur une caméra externe, le lien est moins naturel car le sujet doit scanner l'environnement avec sa tête et les phosphènes générés se déplacent au gré des mouvements oculaires. En revanche, toutes les neuroprothèses visuelles utilisant une caméra externe, qu'elles soient rétiniennes, corticales ou sur le nerf optique, permettent de contrôler finement l'image à restaurer et les stimulations à appliquer. Avant d'être transformée en impulsions électriques, l'image peut être modifiée en temps réel par des algorithmes de traitement d'image.

Résultats fonctionnels

En regardant de plus près les résultats fonctionnels proposés par les trois grandes catégories de neuroprothèses visuelles, on peut s'apercevoir que les fonctions restituées aux personnes implantées sont aujourd'hui encore très limitées.

Pour les neuroprothèses corticales, seul le système développé par William Dobelle a été installé chez des non-voyants (16 personnes). Cependant, aucun résultat concret sur les capacités perceptives des patients (études psychophysiques) n'a été publié. L'un de ses patients, Jens Naumann, relate son expérience personnelle dans un livre dont il est l'auteur [Naumann 2012]. À cette époque, le système n'est qu'à l'état de prototype, et Naumann est capable de différencier des formes très contrastées.

Tout comme les implants corticaux, aucune neuroprothèse visuelle du nerf optique n'a été à ce jour commercialisée. Cependant quelques résultats fonctionnels sont rapportés. Un sujet capable de distinguer une centaine de phosphènes, identifie de gros motifs noirs apposés sur un fond blanc après un long entraînement. Quasiment une minute lui est nécessaire pour reconnaître un motif. Ce même sujet localise et discrimine de gros objets blancs disposés sur un fond noir, après un entraînement là aussi conséquent [Duret et al. 2006].

Concernant les implants rétiniens, il est intéressant de comparer les résultats obtenus sur les deux systèmes les plus avancés à ce jour : l'Argus II de Second Sight et l'Alpha IMS de Retina Implant AG. Le premier est basé sur une caméra externe et contient 60 électrodes, le second utilise des photodiodes et 1500 microélectrodes. La meilleure acuité visuelle rapportée est de 6/380 avec l'Argus II et 6/160 avec l'Alpha IMS. Bien que ces résultats soient encourageants, la correction est encore loin du seuil légal de cécité (6/120), d'autant plus que dans les deux cas, seul un ou deux sujets obtiennent une telle performance. D'autres améliorations sont rapportées pour les sujets équipés de ces deux implants : certaines personnes parviennent à saisir, voire décrire des objets, ou encore lire de grosses lettres. Dans tous les cas, les environnements utilisés pour faire ces évaluations sont extrêmement simplifiés et ne sont pas représentatifs de situations réelles. Avec trente fois plus d'électrodes, les résultats obtenus avec l'Alpha IMS sont finalement très proches de ceux obtenus avec l'Argus II. Ce résultat s'explique par le phénomène de diaphonie [Wilke et al. 2011]. Le courant injecté sur une électrode peut se propager aux électrodes voisines si celles-ci sont trop proches les unes des autres. Concrètement, l'Alpha IMS, malgré ces 1500 électrodes, n'induit pas la perception de 1500 phosphènes distincts.

Tous ces résultats indiquent que les améliorations fonctionnelles sont très faibles, et limitées à des environnements contrôlés. Les neuroprothèses visuelles, et tout particulièrement les implants rétiniens, démontrent tout de même une réelle fiabilité. Pour preuve, l'Argus II et l'Alpha ont tous les deux reçus le marquage CE respectivement en 2011 et 2013. Second Sight est également autorisé à commercialiser l'Argus II sur le sol américain suite à l'accord obtenu auprès de la FDA.

Verrous et défis

Pour obtenir une meilleure acuité visuelle et une meilleure restauration fonctionnelle, le principal défi est d'améliorer la résolution spatiale des implants. Pour cela il est nécessaire d'augmenter le nombre d'électrodes. Cependant plusieurs facteurs viennent contraindre cette augmentation de résolution [Meffin 2013] : pour les neuroprothèses rétiniennes, la taille de la matrice est limitée par le site d'implantation et la procédure chirurgicale associée. Aujourd'hui les implants suprachoroïdiens

semblent autoriser les matrices les plus larges (8x13 mm contre 3x3 mm pour une matrice sous-rétinienne). La seconde contrainte est l'espacement centre à centre entre électrodes. En effet, quel que soit le type d'implant, une trop grande densité, et donc un espacement trop faible entre les électrodes, pose le problème de la diaphonie comme expliqué précédemment pour l'Alpha IMS. Il existe aussi un lien étroit entre l'espacement des électrodes et leur proximité avec les neurones à stimuler : plus les électrodes sont proches des cellules cibles, plus l'intensité du courant électrique requise pour évoquer un phosphène est faible, et par conséquent, plus l'espacement entre électrodes peut être réduit. Pour augmenter le nombre d'électrodes, en conservant un espacement minimum, une possibilité est d'utiliser des électrodes moins larges. Cependant, pour ne pas endommager les tissus nerveux, la densité de charge électrique par électrode ne doit pas excéder un certain seuil (350 $\mu C.cm^{-2}$), ce qui impose un diamètre minimum à respecter [Eiber et al. 2013].

Quelques pistes sont à l'étude pour dépasser les contraintes actuelles. L'une d'elle est d'utiliser de nouveaux matériaux pour la fabrication des électrodes comme le diamant en raison de ses propriétés de semi-conducteur et de son excellente biocompatibilité. C'est le choix fait par le consortium Bionic Vision Australia pour le développement de leurs futurs implants. Certaines équipes réfléchissent à de nouvelles formes d'implants. L'Institut de la Vision à Paris s'intéresse par exemple au développement d'implants tridimensionnels [Djilas et al. 2011]. L'idée est de « piéger » les neurones dans des puits au fond desquels se trouvent les électrodes, de telle sorte que celles-ci stimulent électriquement une quantité limitée de neurones. Le contact entre les neurones et les électrodes peut aussi être amélioré en modifiant la surface des électrodes, par exemple en y déposant des matériaux conducteurs ayant une structure tridimensionnelle. Cette nanostructuration des électrodes peut augmenter le rapport signal sur bruit et le transfert de charge jusqu'à plusieurs ordres de grandeur [Castagnola et al. 2014]. Enfin, des techniques innovantes comme la stimulation directe des neurones de la rétine par laser infra-rouge, pourraient ouvrir la voie à des systèmes restaurant une meilleure résolution spatiale [Bec 2010; Bec et al. 2012].

Pour le moment le développement d'implants haute-résolution n'est pas envisageable dans les cinq à dix prochaines années. À ce propos, il est intéressant de regarder la courbe de prévision des implants épirétiniens publiée début 2008 (Figure 1-16). À l'époque, l'Argus II était en cours de test chez l'homme. On prévoyait dès 2009 une version à plus de 200 électrodes, et une version contenant plus de 1000 électrodes en 2014. Huit ans après les premiers tests chez l'homme, l'Argus II est toujours la version la plus évoluée, et les premiers essais cliniques avec l'Argus III (270 électrodes) ne sont pas d'actualité.

D'autres implants rétiniens sont en cours de développement, mais d'une part les résolutions proposées restent faibles (entre 200 et 300 électrodes), et d'autre part, si on suit cette tendance, il faudra entre cinq et dix années avant que les systèmes ne soient validés pour être commercialisés. Pour les neuroprothèses corticales, le constat est le même, si ce n'est plus sévère : la seule neuroprothèse ayant été testée chez l'homme est celle de Dobelle. Depuis, les implants de surface ont été abandonnés au profit des solutions intra-corticales. Malgré des travaux débutés il y a maintenant trente ans, aucune neuroprothèse intra-corticale n'est en cours de test chez l'homme.

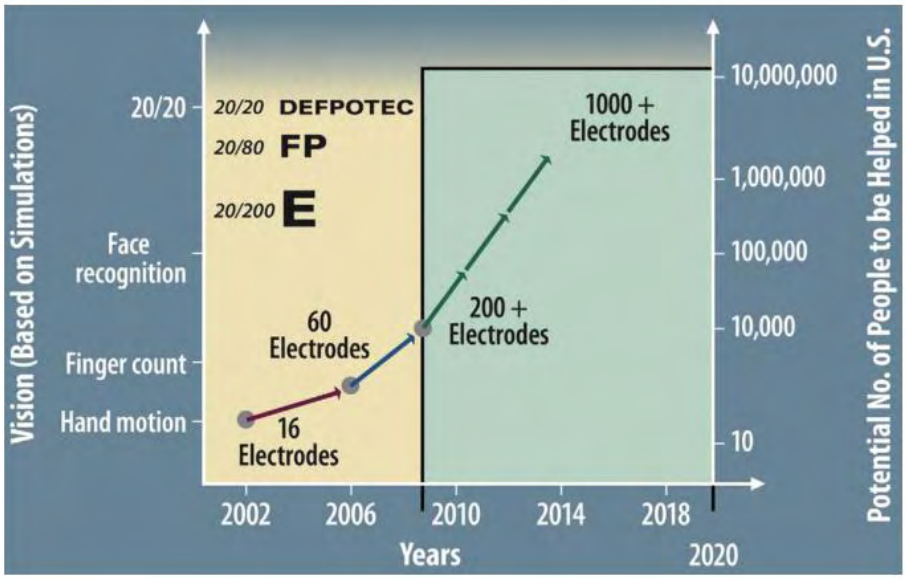
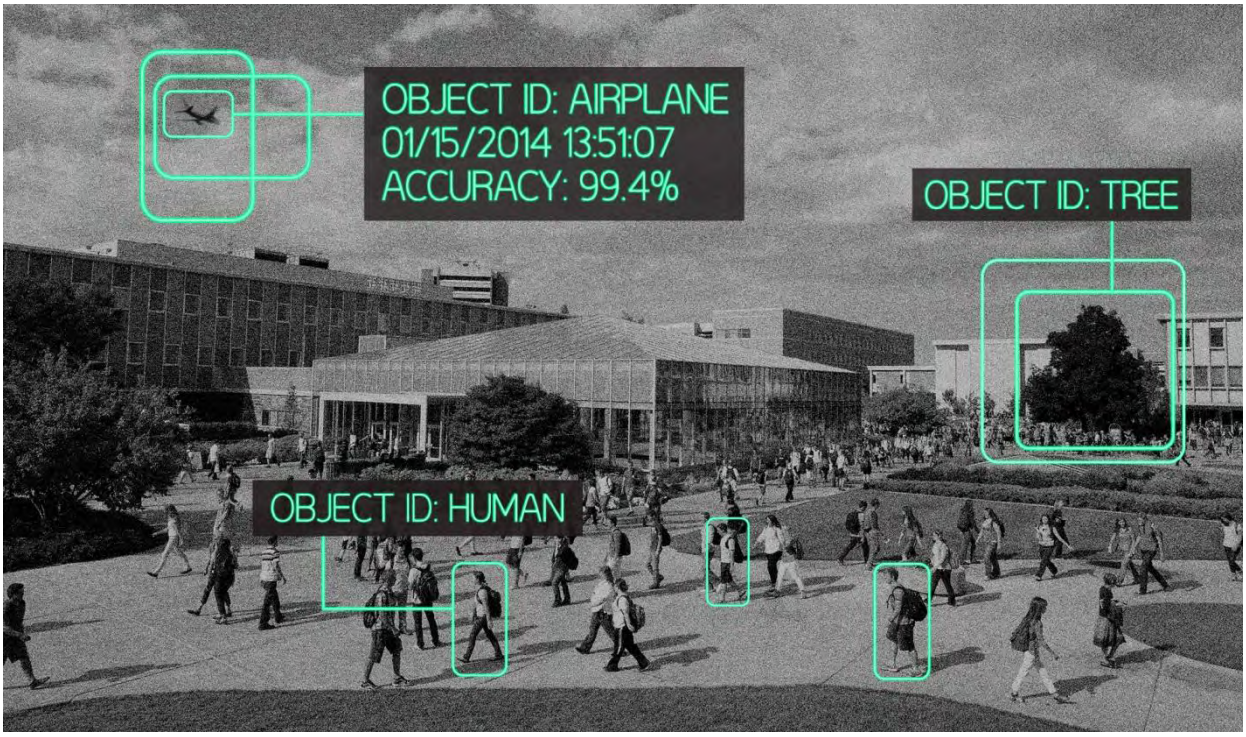


Figure 1-16 Prévision des résolutions d'implants.
Historique et prévision des résolutions d'implants épirétiniens en Janvier 2008.
En beige, les systèmes existants à l'époque, et en vert les prévisions. L'échelle de gauche indique les fonctions visuelles retrouvées. L'échelle de droite correspond au nombre de patients potentiellement concernés/équipés par ces systèmes (Source : Chader, Weiland, and Humayun 2009).

Pour outrepasser les limitations dues à la faible résolution des implants actuels et à venir, une opportunité à moindre coût existe pour toutes les neuroprothèses équipées d'une caméra externe. Pour ces systèmes, il est possible d'imaginer des traitements en temps réel sur l'image, avant de la restituer sous forme d'impulsions électriques. La problématique de cette thèse porte sur l'amélioration de l'utilisabilité des neuroprothèses visuelles, par l'usage d'algorithmes haut niveau issus de la vision par ordinateur. Par exemple, nous pensons que ces algorithmes offrent la possibilité d'extraire en temps réel la localisation d'objets d'intérêt, et que cette localisation peut être restituée par l'intermédiaire de quelques phosphènes seulement. Ainsi, il deviendrait possible, dans un environnement naturel, de localiser un objet spécifique, un visage, ou encore du texte, malgré la très faible résolution spatiale offerte par les solutions actuelles. Dans le chapitre suivant, nous dressons un panorama des différents traitements qu'il est possible de réaliser sur une image par l'intermédiaire d'algorithmes de vision par ordinateur. Puis, nous détaillons les traitements effectués ou imaginés dans le cadre des neuroprothèses visuelles. Enfin, nous présentons l'approche par localisation d'objets d'intérêt, dont la spécificité est l'extraction en temps réel de zones d'intérêt grâce à des algorithmes haut niveau de vision par machine.

CHAPITRE 2

TRAITEMENT D'IMAGES ET VISION PAR ORDINATEUR



(Source Lillywhite et al. 2013)

INTRODUCTION

Exception faite d'une partie des implants rétiniens qui utilisent directement la lumière entrant dans l'œil comme source d'informations visuelles à restaurer [Stingl et al. 2013; Chow et al. 2004], la grande majorité des neuroprothèses visuelles, qu'elles soient rétiniennes, corticales ou du nerf optique, récupèrent cette information par l'intermédiaire d'une caméra externe. Ces systèmes ont alors la possibilité d'appliquer des traitements sur les images captées. Ces traitements doivent être temps réel pour ne pas perturber le comportement visuomoteur de la personne implantée. Ces neuroprothèses sont dites actives car elles ont un contrôle total sur le pattern de phosphènes qu'elles génèrent. Il est possible d'envoyer indépendamment à chacune des électrodes les impulsions électriques adéquates pour induire l'apparition d'un ensemble de phosphènes aux positions voulues. Avec ces systèmes, il est donc envisageable de traiter en temps réel les images captées par la caméra pour définir par le biais d'instructions électriques le pattern de phosphène présenté.

Partant du constat que les résolutions fournies par les neuroprothèses visuelles sont extrêmement limitées, des équipes de recherche se sont très rapidement intéressées aux techniques de traitement d'image, afin de simplifier celles-ci (par exemple en améliorant le contraste) avant leur restauration par l'intermédiaire de phosphènes. De manière similaire, de nombreux algorithmes de traitement du signal ont été développés pour les implants cochléaires, pour filtrer du mieux possible le signal auditif recueilli.

La contribution de cette thèse ne porte pas directement sur le domaine de la vision par ordinateur, mais les solutions que nous proposons s'appuient dessus. Le premier objectif de ce chapitre est de présenter dans les grandes lignes les principes et les avancées les plus pertinents de ce vaste domaine. Nous nous intéressons particulièrement aux algorithmes qui privilégient la vitesse, un critère fondamental pour être embarqués dans les neuroprothèses visuelles. Puis, nous étudions dans le détail les traitements d'images qui sont utilisés dans les neuroprothèses réelles, et ceux qui ont été imaginés et testés au travers de simulations. Enfin, notre discussion

porte sur les limitations actuelles et ouvre sur notre proposition, qui constitue la contribution principale de cette thèse.

LE TRAITEMENT D'IMAGES

Le traitement d'images vise à développer des outils pour transformer une image numérique (par exemple la compresser, ou modifier son apparence), ou pour détecter et extraire de cette image une information particulière (une intensité, une couleur, des régions d'intérêt, la présence d'un élément, etc.). D'un point de vue informatique, une image est une matrice à deux ou trois dimensions dans laquelle sont stockées les valeurs d'intensité/couleur de chacun des pixels. Ces valeurs correspondent soit à un niveau de gris (souvent compris entre 0 = noir et 255 = blanc), soit à une couleur (un triplé rouge, vert, bleu, dont chaque composante varie aussi classiquement entre 0 et 255). D'un point de vue mathématique, une image peut être considérée comme une fonction discrète à deux arguments, qui associe à une coordonnée donnée, une intensité/couleur. Comme pour n'importe quelle fonction mathématique, on peut y appliquer des opérations. Le traitement d'une image peut être vu de façon simplifiée comme l'application d'un ensemble d'opérations mathématiques.

LA VISION PAR ORDINATEUR

La vision par ordinateur (également appelée vision artificielle ou vision par machine) est un domaine de recherche pluridisciplinaire qui s'appuie principalement sur les mathématiques, le traitement d'images ou encore l'intelligence artificielle (notamment l'apprentissage automatique). Il est extrêmement difficile de parler de vision par machine sans parler de traitement d'images, tant les deux disciplines sont liées. Cependant, leurs objectifs fondamentaux diffèrent. Dans de nombreux cas, la vision par machine essaye de reproduire les capacités du système visuel (humain), par exemple l'interprétation d'une scène visuelle. Pour y parvenir elle s'appuie sur des algorithmes de traitements d'images. Le traitement d'images est donc l'un des outils de cette discipline. Par exemple pour répondre à la question « où se situe le sol dans cette image ? », il conviendra d'utiliser un algorithme de segmentation de l'image pour déterminer les emplacements des différentes surfaces et leurs dispositions relatives.

Depuis plusieurs années, les progrès dans le domaine de la vision artificielle sont impressionnants. Les algorithmes de traitement d'images sont toujours plus performants (tâche à accomplir, vitesse), et l'augmentation régulière dans la puissance de calcul disponible sur un ordinateur ou un smartphone permet de les exécuter de plus en plus rapidement. À notre connaissance, il n'existe pas vraiment de consensus sur une taxonomie des techniques développées dans le cadre de la vision par machine. Pour cette présentation, notre choix est de les regrouper en deux grandes catégories :

les **traitements de bas niveau**, qui à partir d'une image renvoient une image modifiée, ou un ensemble de zones d'intérêt (des points, des lignes ou des régions).

les **traitements de haut niveau**, qui font souvent appel à une combinaison de traitements bas niveau, pour extraire de l'image des informations plus intégrées (la position d'un objet, l'identité d'un visage).

La présentation que nous proposons pour ces deux catégories n'a pas pour ambition d'être exhaustive. Pour les traitements de haut niveau, l'étude se limite à la détection rapide d'objets (au sens large du terme) dans une scène naturelle. Pour une vision complète et approfondie, on pourra se référer à [Szeliski 2011].

Les traitements bas niveau

Les filtres

Un filtre est un outil qui permet de transformer une image en appliquant, à chaque pixel, une fonction tenant compte des pixels les plus proches. Un filtre est dit linéaire lorsque chaque pixel de l'image est remplacé par une combinaison linaire (une somme pondérée) de ses pixels voisins. On trouve dans cette catégorie, le filtre moyenneur et le filtre Gaussien, qui sont des filtres de lissage et qui permettent notamment de diminuer le bruit ou encore de flouter l'image (Figure 2-1). Les filtres peuvent également servir à détecter les contours dans une image [Canny 1986; Sobel 1970; Prewitt 1970].

Figure 2-1 Exemples de filtres.
De haut en bas, (1) le filtre gaussien atténue le bruit d'une image en la lissant, (2) l'égalisation d'histogramme améliore le contraste d'une image, et (3) le filtre de Sobel met en évidence les contours présents dans l'image.

Les points ou les régions d'intérêt

En vision par machine, un point ou une région d'intérêt est une zone de l'image qui a une propriété locale discriminative. Parmi ces zones remarquables (Figure 2-2), on retrouve les points saillants comme les coins (changement brutal de direction) [Rosten & Drummond 2006; Harris & Stephens 1988], et plus généralement, les régions d'intérêt dont les propriétés locales sont remarquables [Alahi et al. 2012; Bay et al. 2006; Lowe 1999; Matas et al. 2004; Nistér & Stewénius 2008; Rublee et al. 2011]. Les contours décrits dans le paragraphe précédent peuvent être également considérés comme des régions d'intérêt.

Deux grandes classes d'algorithmes se distinguent : ceux permettant de **détecter** les points ou régions d'intérêts dans une image, et ceux permettant de les **décrire** au moyen d'une signature unique (vecteur caractéristique). Certains algorithmes se chargent à la fois de la détection et de la description, comme par exemple SIFT (Scale-Invariant Feature Transform) qui détecte un ensemble de points remarquables

(points-clés) et les décrit au moyen d'un vecteur caractéristique à 128 dimensions [Lowe 1999]. Comme nous le verrons par la suite, ces points et régions d'intérêt sont souvent le point de départ d'algorithmes de plus haut niveau.

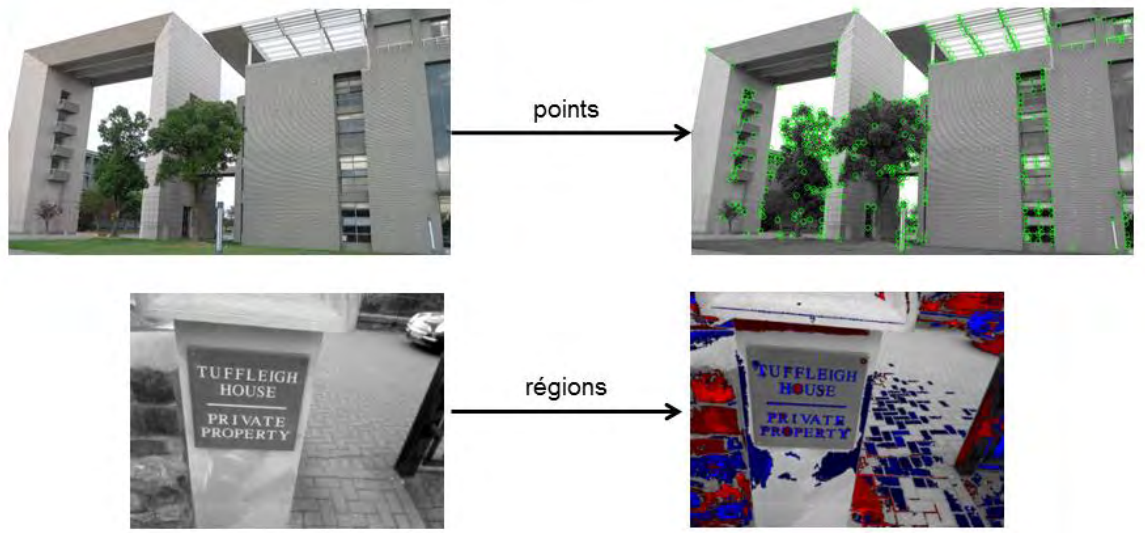


Figure 2-2 Exemples de points et de régions d'intérêt.
De haut en bas, (1) détection de coins de Harris (Source http://sse.tongji.edu.cn), et (2) détection de régions (MSER) (Source Merino-Gracia, Lenc, & Mirmehdi, 2012).

La carte de profondeur

En vision par machine, une carte de profondeur indique la distance métrique de chaque point de l'image par rapport à un capteur. Pour obtenir une telle carte, une première technique consiste à récupérer en parallèle deux points de vue différents issus d'une même scène par l'intermédiaire d'une caméra binoculaire [Scharstein & Szeliski 2002]. Les deux images sont comparées pour déterminer les zones de correspondance et un algorithme calcule les disparités verticales et horizontales en chaque point de l'image. En s'appuyant sur une calibration préalable, il est possible, à partir de cette carte de disparité, de déduire une carte de profondeur. Une autre possibilité est de construire cette carte à partir d'images acquises avec une caméra monoculaire [Hartley & Zisserman 2003]. Le principe général est d'extraire d'une image des points d'intérêt (par exemple les coins), et de suivre leurs trajectoires dans les images issues du mouvement de la caméra afin d'en déduire la profondeur de ces points. Enfin, notons l'arrivée de nouveaux capteurs de vision particulièrement performants. Par exemple, les caméras 3D « temps de vol » sont capables de

construire une carte de profondeur à raison de 50 images par seconde. Le capteur projette des impulsions laser sur la scène. Les profondeurs sont déduites du temps de vol aller-retour de ces impulsions. De nombreuses sociétés embarquent désormais ces capteurs 3D dans leurs dispositifs (Kinect 2.0 de Microsoft, projet Tango[8] de Google, caméra RealSense[9] d'Intel…).

La carte de saillance

Une carte de saillance permet de prédire dans une image le focus de l'attention visuelle, en d'autres termes prédire les mouvements oculaires. La construction automatique de ces cartes repose sur des modèles apparus à la fin des années 1990. L'un des plus célèbres est celui proposé par Itty et al. [Itti & Koch 2001] : la carte de saillance associée à l'image est calculée à partir d'une combinaison de zones remarquables (des intensités, des orientations, des couleurs, etc.) obtenues à plusieurs échelles spatiales. Comme nous le verrons par la suite, ces cartes ont fait l'objet d'études en vision prothétique.

La segmentation

La segmentation a pour objectif de partitionner une image en un ensemble de régions de même nature. Nous pouvons distinguer deux approches : l'approche bottom-up et l'approche top-down. L'approche top-down consiste à utiliser des connaissances sur la représentation de la ou des régions à segmenter. Elle sera abordée dans le paragraphe sur les algorithmes de détection d'objets. Dans l'approche bottom-up, aucune information à priori n'est fournie sur la nature des régions à extraire. Seules les caractéristiques de l'image permettent d'effectuer des regroupements (intensité des pixels, texture, contours). Il existe énormément de techniques de segmentation bottom-up. L'une d'entre elles consiste à faire croître un ensemble de petites régions, en y incorporant les régions adjacentes similaires selon un certain critère [Vincent & Soille 1991]. À l'inverse, il est possible de partir d'une image entière et de la subdiviser en petites régions jusqu'à que celles-ci deviennent homogènes. Ces deux techniques peuvent être également mixées [Horowitz & Pavlidis 1976]. Une autre possibilité est la segmentation par regroupement supervisé

[8] https://www.google.com/atap/projecttango/#project

[9] http://www.intel.fr/content/www/fr/fr/architecture-and-technology/realsense-depth-technologies.html

(algorithme des k-moyennes) ou non supervisé [Comaniciu & Meer 2002]. On peut enfin citer les techniques considérant la segmentation d'image comme un problème de partitionnement de graphe [Felzenszwalb & Huttenlocher 2004; Shi & Malik 2000]. Le principe est de modéliser l'image comme un graphe pondéré non orienté, puis de le partitionner. Les nœuds du graphe sont des groupes de pixels de l'image, et le poids entre deux nœuds correspond à la similarité entre les deux groupes de pixels.

Figure 2-3 Exemple de segmentation d'image.
Segmentation par un algorithme de regroupement non supervisé (Source Comaniciu & Meer 2002).

Les traitements haut niveau

Dans cette partie, seuls les traitements permettant une détection rapide (et à fortiori une localisation) d'éléments d'intérêt dans une scène naturelle (un objet, des visages ou du texte) sont couverts. En vision par ordinateur, la problématique de la détection d'un objet dans une image a été très étudiée notamment en raison de son vaste potentiel d'application (la vidéo surveillance, la robotique, la recherche d'image par le contenu…). Ce problème est loin d'être trivial. La difficulté principale tient au fait qu'un objet peut avoir une apparence très différente en fonction de nombreux paramètres comme l'illumination de la scène, la variation de point de vue, la taille apparente de l'objet, les occlusions potentielles, le mouvement de la caméra, les changements de forme pour les objets déformables… Regrouper dans une même catégorie des objets ayant parfois une apparence très différente se situe encore à un niveau de difficulté supérieure, mais de nouveaux algorithmes toujours plus performants apparaissent à un rythme soutenu et commencent à résoudre ces défis d'une manière convaincante. Des compétitions associées aux grandes conférences du domaine permettent de confronter ces algorithmes sur des jeux de données de plus en plus riches et réalistes.

La localisation d'objets

Pour localiser des objets dans un flux vidéo, trois grandes approches ont été proposées [Jafri, Ali, Arabnia, et al. 2013] : l'approche à base d'étiquettes, celle utilisant des algorithmes 2D de reconnaissance de formes, et enfin une approche utilisant la construction de modèles 3D. Cette dernière ne sera pas abordée ici, car elle est aujourd'hui inadaptée à une utilisation temps réel en raison de sa complexité algorithmique.

Approche à base d'étiquettes

Cette approche est spécifiquement apparue dans les systèmes développés pour les personnes malvoyantes. Elle consiste à apposer sur les objets concernés une étiquette comportant des caractéristiques unique (un code-barres, un QR code, des formes particulières …). De cette façon, la détection des objets dans l'image se résume à la détection des étiquettes qui ont été conçues pour faciliter cette reconnaissance. Cette solution a par exemple été mise en place dans le système Badge3D [Iannizzotto et al. 2005]. Les étiquettes sont composées d'un code-barres simplifié, lui-même entouré d'un rectangle noir. Un filtre de Canny permet de détecter ce rectangle, d'extraire le contenu du code-barres, et ainsi identifier et localiser l'objet. Bien qu'efficace, ces solutions sont contraignantes car il n'est pas envisageable, mis à part pour les objets personnels les plus courants, de procéder à un étiquetage massif de tous les objets nous entourant.

Approche à base d'algorithmes 2D

Comparée à l'approche précédente, l'approche à base d'algorithmes 2D est beaucoup plus générique et nécessite de ce fait l'implémentation de systèmes plus complexes, et des temps de calcul plus importants. Il convient de distinguer deux problématiques différentes : la détection d'instances d'une même catégorie (je cherche « la locomotive jaune »), et la détection de catégorie d'objets dans son ensemble (je cherche « les chaises »). Cette seconde problématique est bien plus difficile à mettre en œuvre car le plus souvent, il existe une très forte variabilité des instances au sein d'une même catégorie.

Pour détecter et reconnaître des instances d'objets, les premières solutions focalisaient sur l'extraction de lignes et de contours, afin de les mettre en correspondance avec un ensemble d'objets connus [Szeliski 2011]. Les solutions actuelles sont plus robustes et rapides. Elles reposent sur une approche à base d'apprentissage supervisé de modèles d'objet. Leur principe est le suivant : (1) construire en premier lieu une base d'objets de référence (les objets à rechercher) à partir de descripteurs de points d'intérêt représentant ces objets, (2) en phase de détection, extraire des nouvelles images les points d'intérêt, (3) essayer de les mettre en correspondance avec les objets stockés en base, et (4) si un nombre suffisant de correspondances est trouvé avec l'un des objets de la base, vérifier que leur structure géométrique s'aligne correctement.

Même si il n'est pas temps réel, SIFT est l'algorithme qui a rendu populaire ces méthodes [Lowe 1999]. SIFT (Figure 2-4) détecte des points d'intérêt (points-clés) dans une image et caractérise chacun d'entre eux au moyen d'un vecteur contenant 128 caractéristiques. Ces points-clés sont invariants aux rotations, aux changements d'échelle et d'illuminations, et à moindre mesure aux modifications de point de vue (jusqu'à 30°). Depuis, de nouveaux algorithmes bien plus rapides que SIFT ont été décrits. C'est le cas par exemple de SURF [Bay et al. 2006], ou plus récemment de FREAK [Alahi et al. 2012].

Figure 2-4 Détection d'objets reposant sur SIFT.
Ici un exemple avec des détecteurs/descripteurs SIFT, pour une application en reconnaissance d'objet. Les deux objets à gauche sont reconnus dans la scène très chargée, par la mise en correspondance de leurs descripteurs SIFT (Locomotive en vert et en jaune, grenouille en rouge).

Avec cette approche, la base d'objets de référence peut vite devenir très grande. Pour éviter d'avoir à comparer un nouvel objet à toute la base, des techniques efficaces, adaptées du domaine de la recherche d'information, ont été proposées [Sivic & Zisserman 2009]. De façon très simplifiée, ces techniques reposent sur le principe du « sac de mots » (modèle « bag of words ») où chaque image est exprimable par un ensemble de mots visuels prédéfinis. Un algorithme de similarité permet de retrouver pour une image donnée, les objets de la base les plus semblables, et de filtrer ces derniers par un algorithme plus classique à base de mise en correspondance de descripteur de points d'intérêt.

Différentes méthodes ont également été proposées pour détecter et reconnaître des catégories d'objet [Szeliski 2011], mais aujourd'hui encore les solutions se limitent à la reconnaissance efficace de quelques classes[10] (visage, voiture, avion...). Les approches les plus classiques reposent sur l'utilisation de « sac de mots », sur de la segmentation (top-down puisque elle s'appuie ici sur des informations concernant le ou les objets à extraire) ou encore sur un découpage par parties. Dans cette dernière approche, l'idée est de modéliser une classe en la partitionnant (un vélo est constitué de deux roues, d'un guidon et d'un cadre), et d'indiquer la relation géométrique entre ces parties. La détection de l'objet se fait ensuite par l'extraction et la mise en correspondance des différentes parties, et la conformité de leur configuration spatiale.

Enfin, certaines solutions très prometteuses reposent sur une approche inspirée du système visuel humain. Dans cette catégorie un algorithme développé par la startup toulousaine SpikeNet Technology, en étroite collaboration avec le laboratoire CerCo[11], a retenu notre attention [Dramas et al. 2010]. Leur solution repose sur un moteur de reconnaissance de formes bio-inspiré. En raison de sa robustesse aux transformations et de son fonctionnement temps réel (quelques dizaines de ms pour reconnaître plusieurs dizaines d'objets), cet algorithme est un excellent candidat pour de la détection d'objets. La reconnaissance de forme se fait ici par apprentissage supervisé : chaque objet est décrit par un ensemble de modèles que l'outil est capable d'apprendre. L'algorithme est ensuite capable de reconnaître ces objets

[10]Résultat de la compétition « Visual Objects Classes 2012»

[11] Centre de Recherche Cerveau & Cognition - UMR5549 - Toulouse

dans de nouvelles images.

La localisation de visages

La détection de visages dans une image compte parmi les problèmes les plus étudiés en vision par machine. En exemple de leur succès, aujourd'hui des algorithmes très rapides sont embarqués dans la grande majorité des appareils photos numériques dans le but d'automatiser la mise au point des visages. La première méthode de détection temps réel a été décrite en 2001 (revue en 2004) par Viola et Jones dans une approche basée sur l'apparence [Viola & Jones 2004; Viola & Jones 2001]. Ces approches se sont démocratisées grâce aux constantes avancées en matière de capacités de calcul et de stockage. Elles sont aujourd'hui toujours les plus performantes. Elles reposent principalement sur un apprentissage supervisé, où un classifieur est entrainé sur un très grand nombre d'images positives (des visages) et négatives. Le classifieur est ensuite capable d'indiquer la position des visages dans une image quelconque. Dans ce processus, deux points sont essentiels : (1) l'extraction de zones d'intérêt pour modéliser les visages, et (2) l'algorithme d'apprentissage. Les avancées les plus récentes en matière de détection de visages concernent l'amélioration de ces deux points [Zhang & Zhang 2010].

La localisation de texte

Tout comme la détection d'objets et de visages, la détection de texte dans une scène naturelle est une tâche complexe en vision par ordinateur. Elle doit s'adapter à de nombreuses variations comme les différentes polices de caractères, les différentes tailles et couleurs du texte, les distorsions géométriques, les occlusions partielles ou encore les multiples conditions de luminosité. Depuis les quinze dernières années, de très nombreux algorithmes ont été rapportés [Sharma et al. 2012; Liang et al. 2005]. Il n'y a pas vraiment de consensus sur une classification des différentes techniques utilisées, mais une façon simple est de les regrouper en deux grandes catégories : celles basées sur la texture, et celles basées sur les régions remarquables.

Approche basée sur la texture

Dans l'approche basée sur la texture, le texte est vu comme une texture particulière distinguable du fond de l'image. Classiquement, ces méthodes découpent l'image en petite zones, et classifient chacune d'entre elles en tant que texte ou non texte, en s'appuyant sur des propriétés particulières de ces zones comme la forte densité de contours, la variabilité importante de l'intensité, etc. [Sharma et al. 2012]. Plusieurs techniques classiques d'apprentissage automatique sont utilisées pour la phase de classification : SVM [Ji et al. 2009; Ye et al. 2007], k-means [Shivakumara et al. 2011; Shivakumara et al. 2009], dictionnaires discriminants [M. Zhao et al. 2010]… Les solutions basées sur cette approche ne sont pas encore temps réel en raison du balayage important à effectuer sur l'image.

Approche basée sur les régions

Par opposition à l'approche basée sur la texture, l'approche basée sur les régions extrait directement de l'image des zones, appelées composantes connexes (un ensemble de pixels voisins au profil similaire). Un ensemble de caractéristiques (principalement des propriétés géométriques) est calculé sur chacune des composantes, et permet de rejeter celles qui ne correspondent pas à des lettres. Cette approche offre l'avantage de détecter directement les caractères quel que soit leur police et leur taille apparente. Les méthodes proposées se distinguent principalement sur la construction des composantes connexes. Dans certaines solutions, les composantes sont obtenues simplement par l'extraction des contours [Shivakumara et al. 2008]; d'autres construisent ces régions en regroupant ensemble les bords d'épaisseur équivalente [Jung et al. 2008; Epshtein et al. 2010]. L'extracteur de régions d'intérêt MSER[12] [Matas et al. 2004] s'avère être aussi un très bon candidat à la construction de composantes connexes pour la détection de texte [Chen et al. 2011; Merino-Gracia et al. 2012]. Aujourd'hui, les dernières solutions reposant sur cette approche sont temps réel ou très proche de l'être.

[12] Maximally Stable Extremal Regions

VISION PAR ORDINATEUR ET NEUROPROTHESES VISUELLES

Le fonctionnement des neuroprothèses visuelles utilisant une caméra externe comme source d'information est schématisé par la Figure 2-5.

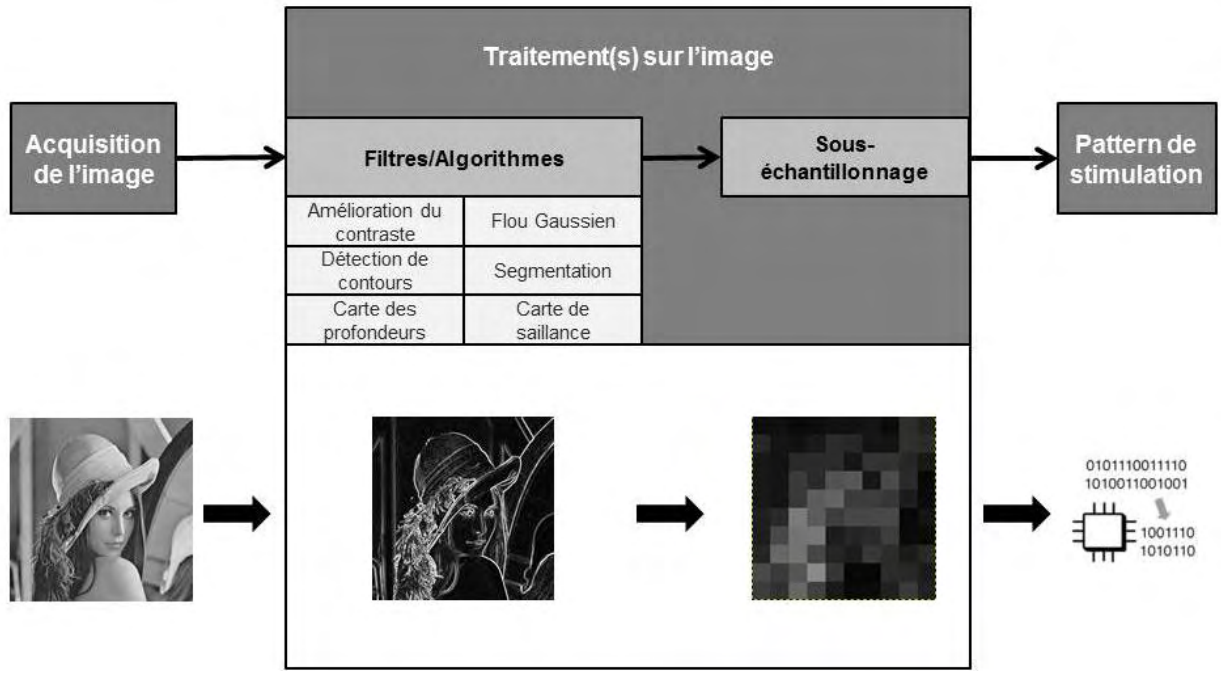


Figure 2-5 Principe général du module de traitement de l'image.
Ce module est inclus dans les neuroprothèses utilisant une caméra externe pour capter l'information visuelle.

Lorsqu'une image est acquise par la caméra, celle-ci est transformée en niveau de gris, modifiée par un ou plusieurs filtres/algorithmes, avant d'être sous-échantillonnée à la résolution de l'implant, pour finalement être transformée sous forme d'impulsions électriques qui vont stimuler le tissu nerveux. Comme nous l'avons vu précédemment, ces systèmes proposent aujourd'hui des implants d'une résolution très faible. L'Argus II de Second Sight possède 60 électrodes, et l'IRIS1 de Pixium Vision en contient 49. Toutes les autres solutions en cours de développement prévoient des implants avec un maximum de 200 à 300 électrodes dans les cinq à dix prochaines années. Il est évident que cette résolution est insuffisante pour restituer toute l'information d'une image composée de centaines de milliers de pixels. En revanche, le module de traitement d'images peut modifier le signal pour en extraire et restituer l'information la plus utile à l'utilisateur en fonction du contexte. Nous allons détailler dans un premier les traitements effectués dans les

neuroprothèses visuelles existantes ou ayant existées, puis nous ferons un état de l'art des solutions imaginées en simulation.

Le traitement d'images dans des systèmes réels

Pour protéger leur propriété intellectuelle, les entreprises qui fabriquent des implants limitent au minimum les informations sur le fonctionnement de leurs systèmes. Ainsi, les traitements d'images effectués dans les neuroprothèses visuelles sont très peu détaillés (Tableau 2-1). Il existe cependant une certitude : à ce jour, seul des traitements de bas niveau sont appliqués aux images.

Tableau 2-1 Les traitements d'images effectués dans les neuroprothèses visuelles.

Implant	Traitement d'image	Références
Prototype (rétinien)	Transformation géométrique Filtres spatio-temporel (Différence de gaussienne)	[Asher et al. 2007]
Dobelle (cortical)	Détection de contour	[Dobelle 2000]
IMI (rétinien)	Filtres spatio-temporel	[Hornig et al. 2008]
Argus II (rétinien)	Amélioration de contraste Détection de contour	[Second Sight Medical Products 2013]
Prototype (cortical)	Détection de contour	[Srivastava & Troyk 2005]
Prototype (rétinien)	Filtre moyenneur, Filtre gaussien, Zoom	[Tsai et al. 2009]

Parmi les solutions proposées, seules trois ne sont pas des prototypes. Très rapidement dans ses recherches, William Dobelle a porté un intérêt à la détection de contours (filtre de Sobel) afin de limiter la quantité d'information à restaurer dans sa neuroprothèse corticale [Dobelle 2000]. Jens Naumann, l'un des tous premiers patients de Dobelle, confirme que ce traitement a bien été implémenté dans son système, et qu'il s'avère plus fonctionnel qu'avec l'image brute [Naumann 2012].

L'implant développé par la société IMI Gmbh (depuis racheté par Pixium Vision) comporte un composant en charge de traiter l'image avant sa restitution sous forme de phosphènes. Ce composant, le « Retina Encoder », implémente un ensemble de

filtres spatio-temporel (Figure 2-6). Plutôt que d'appliquer un filtre sur toute l'image, chaque filtre est découpé en sous-filtre (autant que d'électrodes) en charge de traiter une partie de l'image. Plusieurs images consécutives permettent de prendre en compte des traitements temporels. En sortie, chaque sous-filtre envoie les instructions électriques adéquates à l'électrode qui lui est associée. [Hornig et al. 2008].

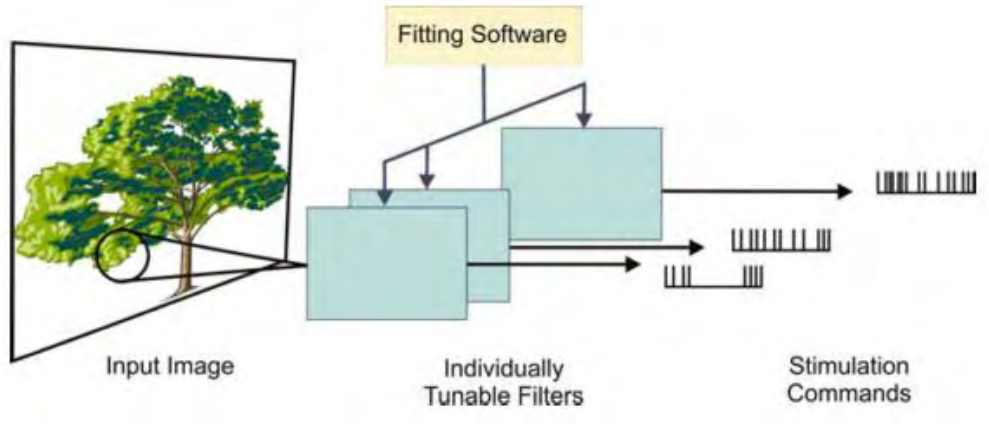


Figure 2-6 Les traitements d'images effectués dans l'implant de la société IMI Gmbh.
Un ensemble de filtres spatio-temporel sont appliqués à l'image en entrée avant de générer les impulsions électriques (Source Hornig et al., 2008).

Pour l'Argus II, quelques traitements effectués sur les images sont brièvement décrits dans le manuel dédiés aux chirurgiens [Second Sight Medical Products 2013]. Ce système inclut un petit composant portable par le biais duquel l'utilisateur interagit. Le Video Processing Unit (VPU cf. Figure 2-7) permet notamment de changer à la volée le traitement effectué sur l'image. Il est ainsi possible de passer d'un mode « standard », à un mode « amélioration des contrastes », ou encore « détection des contours ». Aucune étude à notre connaissance ne permet de mesurer concrètement les améliorations fonctionnelles apportées par ces différents traitements.

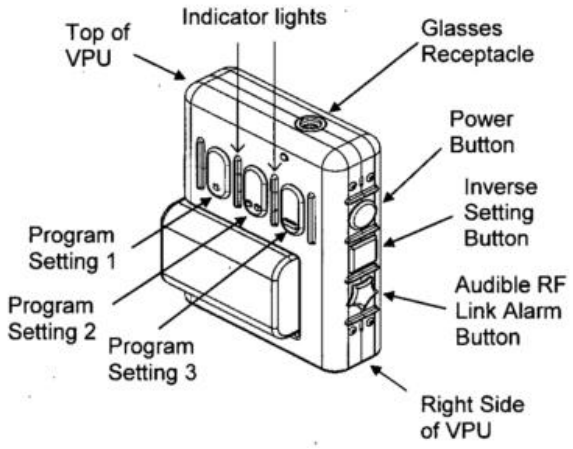


Figure 2-7 Schéma du Video Processing Unit (VPU) du système Argus II.
L'utilisateur peut modifier à la volée le traitement effectué sur l'image par l'intermédiaire des trois boutons « Program Setting ».

Le traitement d'images imaginé en simulation

Il est possible de simuler chez des voyants une vision prothétique par le biais d'un simulateur de neuroprothèse visuelle affiché dans un casque de réalité virtuelle (voir Chapitre 3 pour une description détaillée). Bien que très idéaliste, cet outil donne la possibilité d'imaginer et d'évaluer de nouveaux traitements sur l'image ou de nouveaux implants, et donc de nouveaux rendus, en les testant au préalable chez des personnes voyantes. En 2009, Chen et al. listent l'ensemble des techniques de traitement d'images qui ont été utilisées ou imaginées en simulation de 1992 à 2008 [Chen et al. 2009]. Le Tableau 2-2 complète cette revue, et ajoute les contributions supplémentaires parues entre 2008 et 2014.

La majorité des travaux portent sur l'application d'algorithmes bas niveau sur toute l'image, avant que celle-ci ne soit restituée à l'utilisateur par l'intermédiaire de phosphènes. À l'image de ce qui est implémenté dans l'Argus II, de nombreuses études suggèrent d'améliorer le contraste de l'image et/ou d'utiliser une détection de contours, avant sa restitution.

Tableau 2-2 Les différents traitements d'images imaginés en simulation.

Traitements d'images	Références
Filtres	[Boyle et al. 2002; Boyle 2008; Boyle et al. 2003; Buffoni et al. 2005; Chang et al. 2012; Dowling et al. 2004; Fink et al. 2005; Guo, Qin, et al. 2010; Guo, Wang, et al. 2010; Lu et al. 2013; Lui et al. 2012; Pelayo et al. 2003; Pelayo et al. 2004; Morillas et al. 2007; Vurro et al. 2006; Wang, Lu, et al. 2014; Y. Zhao et al. 2010]
Carte de profondeur	[Wang, Wu, et al. 2014; Boyle 2008; Boyle et al. 2003; Boyle et al. 2002; Feng & McCarthy 2013; Li et al. 2012; McCarthy et al. 2011; McCarthy et al. 2013; McCarthy & Barnes 2012; Mohammadi et al. 2012; Lieby et al. 2011]
Segmentation	[Buffoni et al. 2005; McCarthy et al. 2013; Li et al. 2012; McCarthy et al. 2011; Feng & McCarthy 2013; Horne et al. 2012; Li et al. 2011; Lui et al. 2012; Wang, Lu, et al. 2014; Y. Zhao et al. 2010]
Carte de saillance	[Boyle 2008; Boyle et al. 2002; Boyle et al. 2003; Kiral-Kornek et al. 2011; Parikh et al. 2010; Parikh et al. 2013; Weiland et al. 2012; Stacey et al. 2011]
Zoom[13]	[Stingl et al. 2013; He et al. 2012; Rheede et al. 2010; Wang, Wu, et al. 2014]
Détection de visages	[He et al. 2012; Lui et al. 2012; Wang, Wu, et al. 2014]

Plus récemment, plusieurs travaux provenant de l'équipe de Nick Barnes, proposent un ensemble d'algorithmes dédié aux développements de rendus visuels adaptés à la navigation [Li et al. 2012; Feng & McCarthy 2013; McCarthy et al. 2011; McCarthy et al. 2013]. Ces différentes solutions reposent sur l'utilisation conjointe de la segmentation du sol dans l'image, et de la carte de profondeur obtenue à l'aide d'une caméra stéréoscopique. Les rendus proposés semblent permettre de mieux discerner la structure de l'environnement qui les entoure, et ce malgré la faible résolution des implants simulés. Sur des principes similaires, Mohammadi et al. ont

[13] Dans ces simulations, le zoom est une opération de bas niveau qui consiste à ne restituer qu'une sous-partie de l'image de départ.

proposé un algorithme permettant de calculer très rapidement une carte de profondeur en utilisant une caméra monoculaire [Mohammadi et al. 2012]. L'idée est de pouvoir par la suite concevoir des rendus visuels mettant en valeur les objets les plus proches.

Quelques travaux se sont intéressés à l'extraction d'une information de plus haut niveau, comme la position d'un objet dans une scène, dans le but de restaurer celle-ci à l'utilisateur. Dans cette optique, plusieurs équipes ont proposé d'utiliser une carte de saillance, basée sur le modèle d'attention visuelle de Itti [Itti & Koch 2001]. Stacey et al. détectent automatiquement les obstacles les plus notables, et les mettent en surbrillance par le biais de phosphènes [Stacey et al. 2011]. Parikh et al. extraient l'emplacement des objets les plus saillants, et utilisent quatre phosphènes (nord, sud, est et ouest) pour prévenir l'utilisateur de la présence et la direction de ces objets [Parikh et al. 2013]. Dans un contexte de navigation, cette approche bottom-up (aucune connaissance à priori) peut s'avérer pertinente car elle permet d'indiquer automatiquement aux sujets où se situent les objets se trouvant sur leur chemin. En revanche, elle n'aide pas l'utilisateur à chercher un objet d'intérêt spécifique.

Pour détecter ou identifier un objet d'intérêt, une approche top-down est nécessaire. C'est le choix qui a été fait dans trois études portant sur de la détection de visages [He et al. 2012; Lui et al. 2012; Wang, Wu, et al. 2014]. Constatant que les résolutions actuelles ne permettent pas de distinguer les visages, Lui et al. proposent de les détecter automatiquement en utilisant l'algorithme développé par Viola-Jones, et de les restaurer sous forme d'avatars (Figure 2-8). Pour permettre d'identifier des personnes, He et al. s'intéressent à la détection et au suivi de visages afin de zoomer automatiquement sur eux, et de restituer aux utilisateurs cette partie grossie de l'image (Figure 2-9). Sur un principe identique (zoom automatique), Wang et al. montrent que l'identification d'un visage peut être encore améliorée en élargissant légèrement la zone retournée par l'algorithme de Viola-Jones afin d'inclure des indices visuels supplémentaires, comme la totalité des cheveux. À noter que pour ces deux dernières solutions, même si l'image est zoomée, une résolution importante (500 phosphènes) est tout de même nécessaire pour identifier les visages.

Figure 2-8 Transformative Reality.
Lorsqu'ils sont détectés, les visages sont restaurés par un ensemble de phosphènes représentant un avatar très simplifié [Lui et al. 2012].

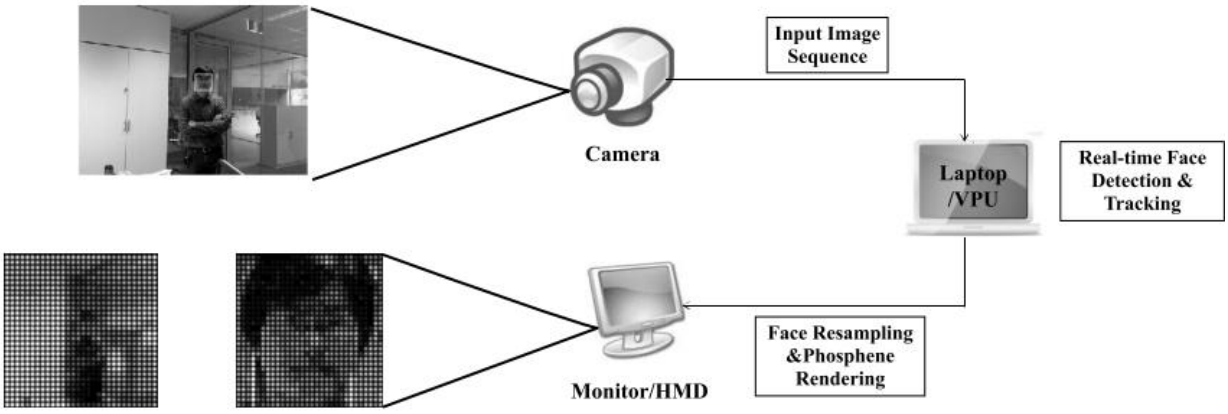


Figure 2-9 Système de fixation automatique de visages.
Le visage est détecté (Viola-Jones) et suivi en temps réel. Il est zoomé numériquement pour être restitué à l'utilisateur dans la totalité de son champ visuel afin de faciliter son identification [He et al. 2012].

DISCUSSION

La vision par ordinateur est une discipline extrêmement dynamique, et en constante progression. Elle repose notamment sur de nombreux algorithmes de traitement d'images toujours plus innovants, robustes et efficaces. L'innovation se fait également au niveau des dispositifs d'acquisition : par exemple, les caméras 3D s'appuyant sur le temps de vol des photons permettent d'obtenir la distance des objets presque instantanément. Les capteurs d'image biomimétique (capteurs d'image asynchrone) sont aussi très prometteurs [Lichtsteiner et al. 2008]. Ils permettent la mise en place d'algorithmes très performants (mise en correspondance stéréo, suivi d'objets…), mais aussi de s'affranchir des problèmes d'éblouissement et de flou que rencontrent les caméras « classiques » [Benosman et al. 2012; Rogister et al. 2012].

La grande majorité des neuroprothèses visuelles inclut une caméra externe, qui capte l'information visuelle avant de la restituer sous forme de phosphènes. Tous ces systèmes souffrent actuellement d'une limitation majeure : ils offrent une résolution spatiale très faible limitée à quelques dizaines de phosphènes. Pour pallier ce problème, les différentes équipes de recherche travaillant sur les implants visuels se sont rapidement intéressées aux algorithmes de traitement d'images. Effectivement, pour faciliter la compréhension des images, ces systèmes procèdent à un prétraitement des images en entrée avant leur restitution sous forme de phosphènes.

L'approche scoreboard

Aujourd'hui, tous les systèmes existants ou en phase de prototypage, sans exception, utilisent un traitement de l'image basé sur un ensemble d'algorithmes/filtres bas niveau (amélioration des contrastes, détection de contours …), dans le but de restituer à l'utilisateur les informations les plus saillantes. À la suite des différents traitements, l'image modifiée est sous-échantillonnée à la résolution de l'implant, puis restaurée dans sa totalité par l'intermédiaire de quelques phosphènes. C'est l'approche que nous nommerons **approche scoreboard** (Figure 2-10) dans la suite de ce manuscrit, en référence aux panneaux d'affichage que l'on peut trouver dans les stades [Dobelle & Mladejovsky 1974].

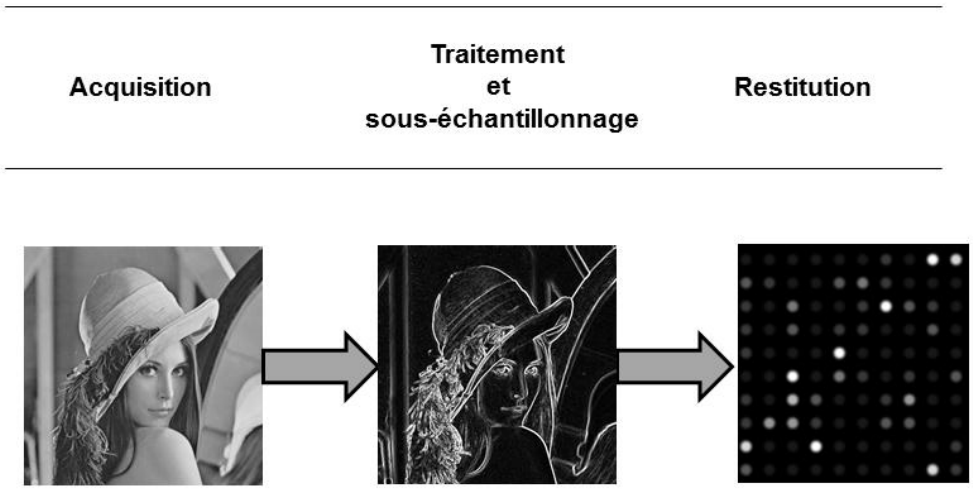


Figure 2-10 L'approche scoreboard.
Dans les neuroprothèses visuelles équipées d'une caméra, les images sont (1) captées par la caméra, (2) traitées en temps réel (ici une détection de contour), puis sous-échantillonnées à la résolution de la matrice d'électrodes (ici 10x10). Enfin (3), l'image résultante de ces modifications est rendue sous forme de stimulations électriques évoquant des phosphènes pour l'utilisateur.

Dans les simulations de neuroprothèses visuelles, la plupart des solutions imaginées proposent aussi un rendu basé sur l'approche scoreboard : l'image est modifiée par un ensemble de traitements de bas niveau, avant d'être entièrement restituée sous forme de phosphènes.

L'approche scoreboard a ses limites : elle n'est pas adaptée lorsque l'utilisateur cherche à localiser précisément un objet d'intérêt. En effet, en situation naturelle, les objets de notre quotidien, les personnes qui nous entourent ou encore les blocs de texte, se trouvent le plus souvent situés à quelques mètres de nous. Dès lors, quels que soient les traitements de bas niveau appliqués à l'image, il est difficilement concevable que quelques phosphènes suffisent à localiser ces objets d'intérêt dont la taille apparente est relativement faible à cette distance.

Pour ces situations précises, et pour aider les utilisateurs dans leurs tâches quotidiennes, nous pensons qu'il est possible d'extraire des images des informations de haut niveau (par exemple la localisation d'un objet), et de restituer cette information par l'intermédiaire de quelques phosphènes, voire d'un unique phosphene selon la tâche à accomplir. En ce sens, nous rejoignons les solutions proposées par Lui et al. et Parikh et al. [Lui et al. 2012; Parikh et al. 2013]. L'idée n'est pas avant tout de vouloir restituer toute l'information contenue dans une image, mais au minimum celle nécessaire à l'accomplissement d'une tâche afin de restaurer des comportements visuo-moteurs utiles pour les personnes implantées.

Proposition : l'approche par localisation de points d'intérêt

Par l'usage d'algorithmes de haut niveau disponibles dès à présent dans le domaine de la vision par ordinateur, nous proposons de développer et tester des rendus visuels fonctionnels en environnement naturel. Cette proposition est fondée sur deux principes inhérents aux neuroprothèses visuelles basées sur une caméra :

un ensemble de traitements est applicable en temps réel sur les images captées par la caméra,

on peut contrôler le pattern de phosphènes à restituer, ainsi que la luminance (jusqu'à 8 niveaux) de chaque phosphène.

Les algorithmes de localisation d'objets sont aujourd'hui performants et s'exécutent très rapidement. Il est par exemple possible d'extraire d'une image, en temps réel, l'emplacement d'une tasse, d'un visage ou encore d'un bloc de texte. S'il est concevable de récupérer la ou les positions de points d'intérêt dans une scène, nous suggérons que cette information soit restaurée par le biais de quelques phosphènes. Cette approche sera nommée **approche par localisation de points d'intérêt** (Figure 2-11) dans la suite du manuscrit.

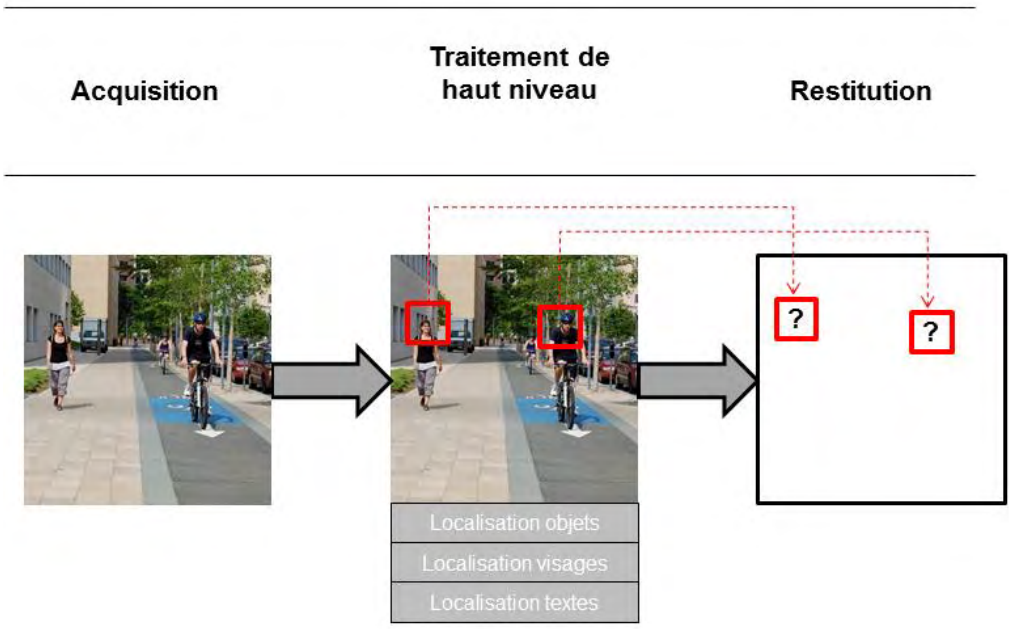
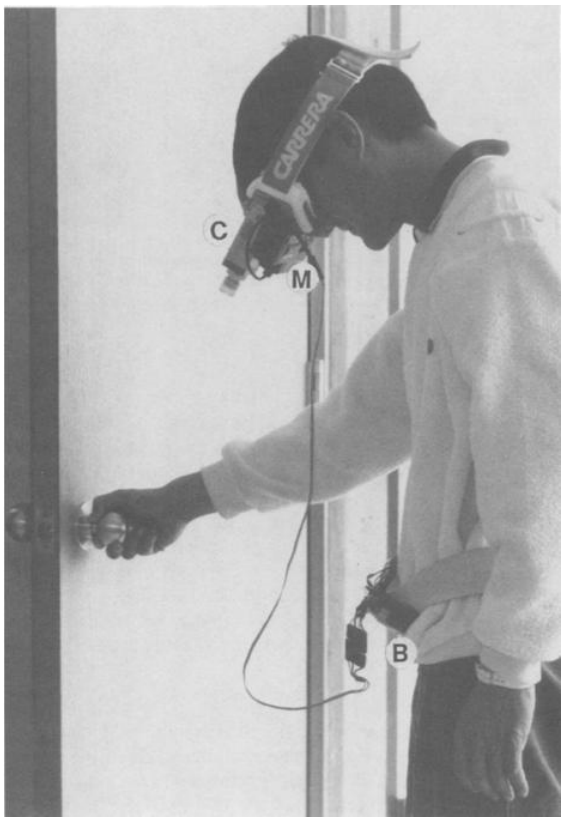


Figure 2-11 L'approche par localisation de points d'intérêt.
(1) L'image est captée par une caméra, (2) un traitement haut niveau permet de récupérer la localisation d'une chaise, et (3) cette information est restaurée par l'intermédiaire d'un ou plusieurs phosphènes.

La contribution principale de cette thèse, est de concevoir et d'évaluer en situation réelle, des rendus visuels basés sur cette approche. Dans un premier temps, le chapitre 3 introduira la notion de simulateur de vision prothétique, outil support à nos expérimentations. Les chapitres 4 à 6 se concentreront sur le développement et l'évaluation de l'approche par localisation de points d'intérêt dans trois contextes : la localisation d'objets du quotidien, la localisation de visages, et la localisation de blocs de texte.

CHAPITRE 3

SIMULATEUR DE VISION PROTHETIQUE



(Source K. Cha, K. W. Horch, et al. 1992)

INTRODUCTION

La simulation de vision prothétique (en anglais SPV pour Simulated Prosthetic Vision) est un outil qui permet de reproduire chez une personne voyante les sensations visuelles perçues par une personne ayant été implantée. Bien qu'ils ne modélisent pas parfaitement les neuroprothèses visuelles, ces simulateurs sont un premier support pour l'évaluation de l'utilisabilité de ces systèmes. Ces évaluations sont notamment bien plus économiques que celles réalisées chez des personnes implantées. L'utilisation de SPV permet également d'investiguer différentes questions de recherche avec une grande souplesse : on peut par exemple faire varier certains paramètres comme le nombre de phosphènes ou la taille du champ de vision restitué, et mesurer les performances obtenues dans des tâches comportementales ou cognitives. On peut également imaginer et évaluer de nouvelles formes de rendu à partir de techniques de traitement d'images.

Les premiers simulateurs ont vu le jour au début des années 90. Ils permettent à Cha et al. de mener des expérimentations sur la lecture et la navigation en vision prothétique simulée [Kichul Cha et al. 1992; K. Cha, K W Horch, et al. 1992]. Ils concluent que plus de 600 phosphènes restaurés dans un petit champ de vision (1,7°) sont nécessaires pour réaliser ces deux tâches. Depuis, d'autres équipes de recherche développent et utilisent ces simulateurs dans des contextes variés. En plus de la lecture et de la navigation, les SPV ont permis de récolter de nombreuses données autour de la reconnaissance d'objets, de visages, ou encore de la coordination œil-main [Barry & Dagnelie 2011].

Pour être les plus réalistes possible, les SPV doivent reproduire au mieux les caractéristiques des neuroprothèses visuelles. Ces caractéristiques sont détaillées au début de ce chapitre. Nous les regroupons en deux grandes catégories : les caractéristiques ou paramètres techniques, inhérents aux matériels et logiciels utilisés dans le cadre des neuroprothèses visuelles, et les caractéristiques liées à la vision prothétique, qui sont issues d'observations chez l'animal et chez l'homme et qui sont propres aux sensations visuelles restituées. Dans un second temps, nous présentons l'architecture générale du simulateur de vision prothétique développé et utilisé lors de nos différentes expériences.

CARACTERISTIQUES DES NEUROPROTHESES VISUELLES ET DE LA VISION PROTHETIQUE

Dans ce paragraphe, toutes les caractéristiques décrites sont celles qui sont communes à l'ensemble des neuroprothèses visuelles existantes ou ayant existé, (corticales, rétiniennes, nerf optique) et équipées d'une caméra externe.

Caractéristiques techniques

Ces caractéristiques sont directement liées aux matériels (la caméra, la matrice d'électrodes de l'implant, …) ou logiciels (les algorithmes de traitement d'images) utilisés dans les neuroprothèses visuelles. Elles sont résumées dans le Tableau 3-1.

Tableau 3-1 Principales caractéristiques techniques d'une neuroprothèse visuelle.
Les dimensions ou unités sont entre parenthèses.

Caméra	Résolution (pixels), Fréquence (Hertz), Champ de vision acquis (degrés)
Traitements	Algorithmes (temps d'exécution), Champ de vision traité (degrés)
Matrice	Résolution (nombre et espacement des électrodes), Taille (millimètres), Disposition (rectangulaire ou hexagonale), Champ de vision couvert (degrés)
Électrodes	Diamètre (micromètres), Distance centre à centre (micromètres), Électrodes non fonctionnelles (pourcentage)

L'acquisition et le traitement des images

L'information visuelle est captée à l'aide de micro-caméras. Celles-ci enregistrent des images à une résolution donnée (nombre de pixels horizontaux * nombre de pixels verticaux) et à une fréquence d'acquisition donnée (exprimée en Hertz ou en nombre d'images par seconde). La zone de la scène couverte par la caméra correspond au champ de vision acquis (nombre de degrés horizontaux * nombre de degrés verticaux).

Les neuroprothèses visuelles actuelles traitent et restituent la totalité de l'image, ou seulement une sous-partie de celle-ci. Dans ce dernier cas, le champ de vision traité est donc restreint à une portion prédéfinie. Par exemple, avec l'Argus II, Second Sight extrait de chaque image une zone couvrant la même taille que celle de l'implant à savoir 17,9° x 10,8° (Figure 3-1). Un ou plusieurs algorithmes sont appliqués à l'image avant que celle-ci soit envoyée au stimulateur. Pour une vue détaillée des traitements effectués on se reportera au chapitre précédent.

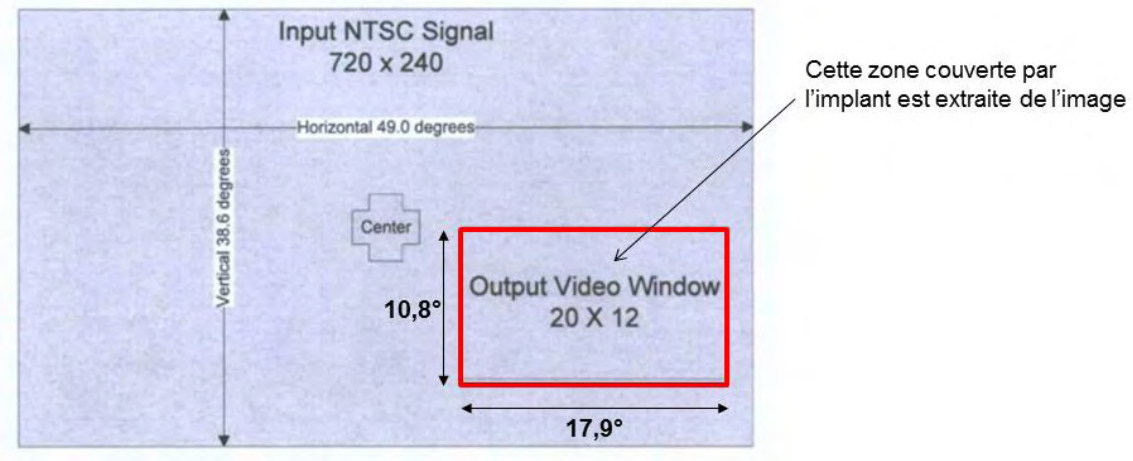


Figure 3-1 Paramètres d'acquisition, de traitement et de restitution d'image pour l'Argus II.
La caméra a une résolution de 720 x 240 pixels. Elle couvre un champ de vision de 49° x 38,6°. Une zone correspondant à la surface couverte par l'implant (17,9 x 10,8°) est extraite de l'image (cadre rouge). Cette partie est sous échantillonnée en 20 x 12 pixels, traitée par différents filtres avant d'être à nouveau sous échantillonnée à la résolution de l'implant (6 x 10 pixels) pour restitution finale (Source Second Sight Medical Products 2013).

La matrice d'électrodes

L'implant comporte une matrice dont la résolution dépend du nombre d'électrodes (Figure 3-2). Celles-ci sont arrangées de façon rectangulaire ou hexagonale. Une configuration hexagonale permet d'obtenir une densité d'électrodes par unité de surface plus élevée qu'une disposition rectangulaire. Les électrodes ont un certain diamètre, et sont séparées l'une de l'autre par une distance (la distance centre à centre) qui est dépendante de l'arrangement choisi. Quel que soit le type de neuroprothèse visuelle, la taille et l'emplacement de la matrice ont un lien direct avec le champ de vision restitué aux personnes implantées.

Figure 3-2 L'implant Argus II.
Photo montrant la matrice d'électrodes fixée dans la rétine d'un patient. Cet implant épirétinien développé par la société Second Sight, mesure 4 x 5 mm et contient une matrice de 6 x 10 électrodes arrangées de façon rectangulaire. Les électrodes ont un diamètre de 200 µm. Elles sont séparées horizontalement et verticalement de 525 µm (centre à centre). La matrice couvre une diagonale d'environ 20° de champ de vision. (Source Zhou, Dorn, & Greenberg, 2013)

Après implantation, certaines des électrodes restent non fonctionnelles : elles n'induisent pas la perception de phosphène. Les premières observations faisaient état d'un nombre important d'électrodes défaillantes. Ce nombre a tendance à diminuer avec les progrès accomplis dans le domaine des traitements de surface et de la connectique. Pour les implants actuels, environ 10% des électrodes ne fonctionnent pas ou n'évoquent pas de phosphènes à la suite de l'implantation.

Caractéristiques visuelles

La vision prothétique correspond à la restitution de l'information visuelle sous la forme de phosphènes. Le phosphène est cette perception lumineuse induite par stimulation électrique de cellules au niveau de la rétine, du nerf optique ou du cortex visuel. De nombreuses expérimentations chez l'homme permettent de caractériser plus précisément leur apparence. On applique un courant électrique via une ou plusieurs électrodes, et le sujet décrit la perception qui en résulte. Chen et al. regroupent de façon détaillée ces résultats dans une revue publiée en 2009 [Chen et al. 2009]. Ils sont résumés dans le Tableau 3-2 et décrits dans les sections suivantes.

Tableau 3-2 Principales caractéristiques d'un phosphène pour les trois types d'implant.
En gras, ce qui est observé le plus souvent.

	Implants corticaux	**Implant du nerf optique**	**Implants rétiniens**
Taille	0,1° - 2,5°	1° - 7°	0,1° - 2°
Forme	**Rond**, Trait, Carré	Ensemble de points, Trait, Triangle	**Rond**, Trait, Ensemble de points, Beignet
Couleur	**Blanc**, **Jaune**, **Gris**, Bleu, Rouge, Marron	**Blanc**, Jaune, Bleu, Rouge	**Blanc**, **Jaune**, Gris, Vert, Bleu, Rouge
Luminance	Jusqu'à 12 niveaux	Jusqu'à 9 niveaux	Jusqu'à 10 niveaux
Carte des phosphènes	Pseudo rétinotopique	Dépendant des paramètres de stimulation	Pseudo rétinotopique
Seuil de fusion temporel	Variable	8-10Hz	40-50Hz
Effet d'adaptation	Oui	Oui	Oui

Forme et taille des phosphènes

Pour les implants corticaux et rétiniens, les phosphènes sont majoritairement des petits points de forme ronde [Dobelle & Mladejovsky 1974; Humayun et al. 2003; Zrenner et al. 2007], mais ils peuvent apparaître sous d'autres formes comme par exemple un trait (grain de riz), un triangle, un carré... Cette forme n'est généralement pas dépendante des paramètres de stimulation. C'est en revanche le cas pour les phosphènes générés en stimulant le nerf optique, où les motifs observés sont plus complexes [Delbeke et al. 2003; Veraart et al. 1998].

Pour avoir une idée de la taille apparente des phosphènes générés, les expérimentateurs demandent aux sujets de comparer la taille perçue avec celle de différents objets portés à bout de bras (environ 60 cm). Il est alors possible d'estimer la taille des phosphènes perçus en degrés d'angle visuel. La plupart du temps celle-ci varie entre 0,5 et 2°. Là encore une plus grande variabilité est observée avec une stimulation du nerf optique [Veraart et al. 1998]. Les phosphènes induits avec des implants corticaux ont une particularité : leur taille est fonction du lieu de stimulation et suit le facteur de magnification corticale, à savoir qu'ils s'accroissent avec l'excentricité [Brindley & Lewin 1968; Dobelle & Mladejovsky 1974].

De façon plus générale, la taille des phosphènes générés est notamment liée au diamètre des électrodes. Plus ce diamètre est petit, plus le phosphène est fin, car l'électrode stimule une zone de tissus nerveux moins large. En revanche, les électrodes doivent respecter une taille minimum pour garder une stimulation efficace tout en restant sous le seuil de sécurité en termes de densité de courant, sous peine d'abimer les cellules ciblées. Pour que les électrodes soient les plus fines possibles, tout en restant efficaces, elles doivent être placées au plus proche des neurones à stimuler, car l'intensité électrique requise pour induire la perception de phosphènes sera minimisée. Une autre voie possible est d'améliorer le transfert de charge entre l'électrode et le tissu, par exemple en nanostructurant la surface de l'électrode [Castagnola et al. 2014].

Couleur et luminance du phosphène

Quel que soit le type de neuroprothèse visuelle, les phosphènes sont le plus souvent blancs ou jaunes [Humayun et al. 2003; Zrenner et al. 2007; Dobelle & Mladejovsky 1974; Veraart et al. 1998]. Mais d'autres couleurs (rouge, bleu ou encore marron) ont été rapportées. À l'heure actuelle, il n'est pas possible de contrôler l'apparition de ces couleurs. En revanche, on peut faire varier la luminance des phosphènes en modulant l'intensité de la microstimulation électrique [Nanduri et al. 2012; Greenwald et al. 2009]. Certains sujets distinguent jusqu'à dix niveaux avec un implant épirétinien [Humayun et al. 2003]. Douze niveaux sont même rapportés dans les premières expérimentations concernant les implants corticaux de surface [Schmidt et al. 1996]. À noter que les résultats sont très variables entre les sujets quel que soit le type d'implant.

Carte des phosphènes

L'application d'un courant électrique via une électrode induit l'apparition d'un phosphène à un emplacement dépendant de la position du regard de la personne implantée. La carte des phosphènes correspond au référencement, électrode par électrode, de l'ensemble des phosphènes générés, lorsque la personne implantée regarde droit devant elle. Elle varie en fonction du type de l'implant et est propre à chaque individu. Dans tous les cas elle ne couvre qu'une partie restreinte du champ de vision. Pour les implants corticaux et rétiniens, ces cartes sont dites pseudo rétinotopique car les phosphènes n'apparaissent pas parfaitement aux emplacements théoriques correspondant au site de stimulation. Concernant les implants du nerf optique, la correspondance électrodes/phosphènes est encore plus complexe car elle dépend fortement des paramètres de la stimulation électrique [Delbeke et al. 2003]. De façon intéressante, les implants rétiniens sont ceux qui produisent des cartes de phosphènes dont le pattern est le plus proche de celui formé par les électrodes (Figure 3-3). La constitution des cartes de phosphènes est une étape essentielle au bon fonctionnement des neuroprothèses visuelles : la connaissance des emplacements de chacun des phosphènes est un prérequis à la reconstruction de l'image.

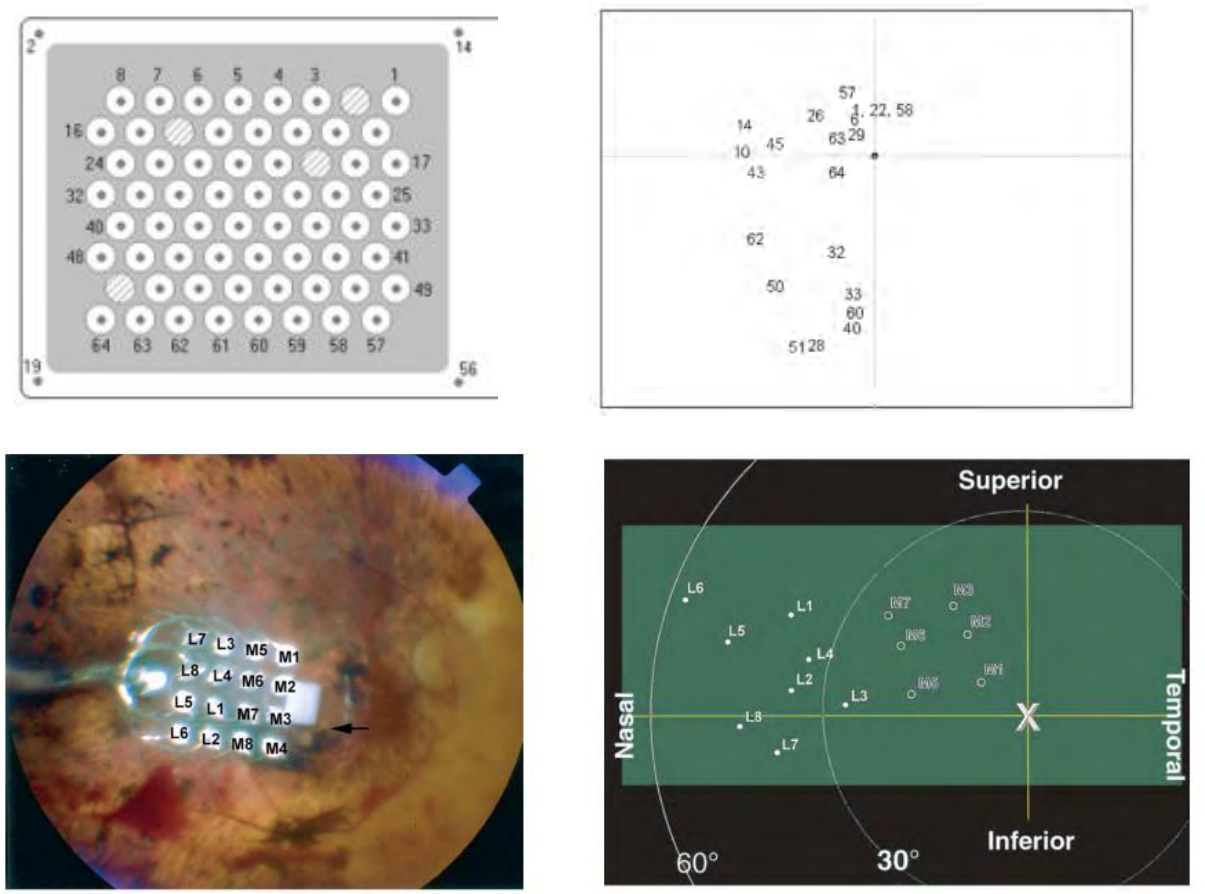


Figure 3-3 Exemples de carte de phosphènes.
À gauche la disposition des électrodes et à droite, la carte des phosphènes. En haut, l'exemple pour un implant cortical de surface, et en bas pour un implant épirétinien [da Cruz et al. 2013; Dobelle 2000; Humayun et al. 2003].

Seuil temporel de fusion des phosphènes

Si l'on applique une seule stimulation électrique à un individu au niveau de la rétine, du nerf optique ou du cortex visuel, celle-ci va induire un phosphène qui va disparaître rapidement. Pour que la personne perçoive ce phosphène de façon continue, et sans effet de scintillement, il est nécessaire de stimuler le tissu nerveux très régulièrement. La fréquence pour obtenir la rémanence des phosphènes est différente suivant le type d'implant : très variable pour les implants corticaux de surface, autour de 40 Hz pour les implants intra-corticaux [Schmidt et al. 1996], entre 8 et 10 Hz pour les implants du nerf optique [Delbeke et al. 2002], et très variable également pour les implants rétiniens (Chen et al. avait rapporté entre 40 et 50 Hz [Chen et al. 2009], mais Fornos et al. n'ont pas observé de scintillements à 20 Hz [Pérez Fornos et al. 2012]). À noter que les seuils sont très différents selon les sujets.

Effet d'adaptation

Bien que l'effet d'adaptation des récepteurs neuronaux ne soit pas forcément la seule explication, on observe classiquement qu'une stimulation de même intensité, répétée dans le temps, induit une perception de plus en plus ténue (le phosphène parait de moins en moins lumineux pour finalement disparaître). C'est ce qui est rapporté par 8 sujets sur 9 dans une étude récente sur des individus équipés de l'Argus II [Pérez Fornos et al. 2012].

ARCHITECTURE DU SIMULATEUR

Les premiers simulateurs de vision prothétique ont été créés au début des années 1990 [K. Cha, K W Horch, et al. 1992; K. Cha, K. W. Horch, et al. 1992; Kichul Cha et al. 1992]. Les phosphènes étaient alors simulés en recouvrant un petit écran avec différents films pré-troués (le nombre de trous correspondant au nombre de phosphènes). Depuis, les technologies ont largement évoluées et aujourd'hui les neuroprothèses visuelles sont simulées de façon beaucoup plus dynamique et réaliste.

L'architecture générale choisie pour notre simulateur est illustrée par la Figure 3-4. Elle est constituée de briques matérielles et logicielles qui permettent de prendre en compte l'ensemble des caractéristiques techniques et visuelles décrites précédemment. Cette solution est conforme aux recommandations émises par Chen et al. [Chen et al. 2009] : l'implant simulé est totalement configurable, l'apparence des phosphènes est modulable, et l'expérience pour l'utilisateur est aussi proche que possible de la réalité (affichage dynamique et temps réel avec possibilité d'asservir les phosphènes à la position du regard).

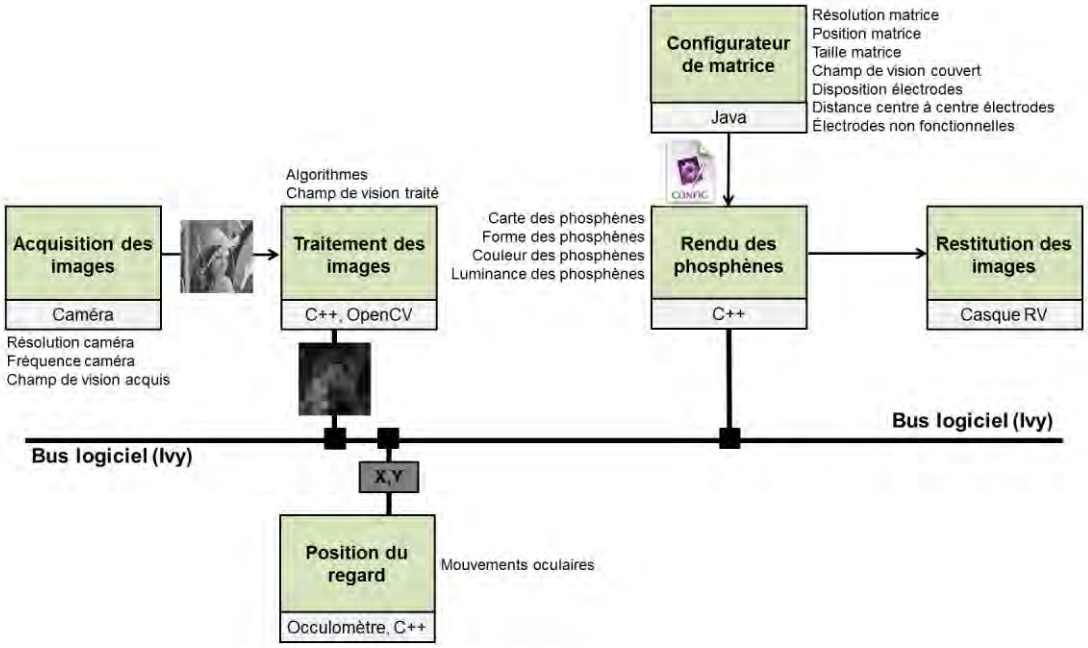


Figure 3-4 Architecture générale du simulateur de vision prothétique.
Le simulateur est découpé en différents modules matériels/logiciels : un module d'acquisition, de traitement, de rendu et de restitution. Le configurateur de matrice est une brique supplémentaire qui permet de spécifier les implants à simuler. Le module de position du regard enregistre les mouvements oculaires pour implémenter l'asservissement des phosphènes à la position des yeux. Les différents composants logiciels communiquent en asynchrone par l'intermédiaire du bus logiciel Ivy.

Acquisition des images

Tout comme dans une neuroprothèse visuelle, nous captons les images à l'aide d'une caméra externe. Celles que nous avons utilisées au cours de nos différentes expériences sont décrites dans le Tableau 3-3.

Tableau 3-3 Caractéristiques des deux caméras utilisées.

Référence	Bumblebee II (Point Grey, BC, États-Unis)	Vuzix 1200AR (Vuzix Cor., CA, États-Unis)
Résolution	2 caméras, 320 x 240 pixels chacune	2 caméras, 640 x 480 pixels chacune
Fréquence	48 Hz	30 Hz
Champ de vision	100°	50°

Traitement des images

Le traitement des images est effectué par un logiciel écrit en C++. Le traitement « scoreboard », qui servira de condition de contrôle dans toutes nos expériences, est développé à l'aide de la librairie OpenCV (Version 2.4.7). Afin d'obtenir la luminance des phosphènes à afficher, l'image capturée est tout d'abord transformée en niveau de gris (fonction cvtColor : G = 0.299*R + 0.587*V + 0.114*B), puis sous-échantillonnée (fonction cvResize avec interpolation bicubique) à la résolution de la matrice de l'implant simulé. Comme la grande majorité des SPV développés, nous prenons le parti de traiter et de restituer la totalité du champ de vision acquis par la caméra. Ce choix nous parait plus fonctionnel pour la personne implantée qui peut accéder à un « champ visuel » large sans avoir à effectuer des mouvements de tête. Pour restreindre ce champ de vision, ou simuler un zoom, il suffirait très simplement de recadrer l'image en entrée. Pour pouvoir mettre en œuvre des expérimentations basées sur la localisation d'objets, des algorithmes spécifiques de vision par ordinateur ont été utilisés. Ils seront détaillés dans chacune des sections les concernant. Quel que soit le traitement effectué, les informations résultantes sont envoyées sur un bus logiciel [Buisson et al. 2002]. Ce bus, appelé Ivy, permet le prototypage rapide de communications asynchrones (des messages sous forme de texte) entre composants logiciels hétérogènes appartenant à un même réseau.

Position du regard

Un implant rétinien ou cortical stimule toujours la même zone du champ visuel. Lors d'une simulation avec une personne voyante, il faut s'assurer qu'une stimulation à un endroit donné du champ visuel soit toujours perçue à cet endroit, indépendamment des mouvements oculaires que peut produire le sujet. C'est la raison pour laquelle il est préférable, pour que la simulation soit la plus réaliste possible, d'asservir l'affichage des phosphènes à la direction du regard. Cet asservissement nécessite l'utilisation d'un oculomètre mais les difficultés de mise place de ce dispositif attaché dans le casque ne nous ont pas permis de l'employer systématiquement. Lorsqu'il est utilisé, un logiciel écrit en C++ envoie toutes les 20 ms la position du regard (sur l'écran, coordonnées en X et en Y) sur le bus logiciel.

Configurateur de matrice

Ce composant logiciel écrit en Java génère un fichier de configuration pseudo-aléatoire contenant les coordonnées et les tailles de chaque phosphène suivant des contraintes fournies en paramètres :

la résolution de l'écran sur lequel sera simulée la vision prothétique (l'écran du casque de réalité virtuelle),

la résolution de la matrice (nombre et espacement des électrodes),

le champ de vision couvert par la matrice,

la disposition des électrodes sur cette matrice,

la distance centre à centre entre électrodes.

Il est également possible de préciser le type d'implant : si la simulation concerne un implant cortical, le configurateur prend en compte l'effet de magnification corticale à savoir l'augmentation de la taille des phosphènes en fonction de leur excentricité [Srivastava et al. 2007].

D'autres paramètres sont modifiables dans le fichier de configuration :

le pourcentage d'électrodes défaillantes,

le bruit (décalage en pixels) à appliquer à la position horizontale et verticale des phosphènes pour simuler l'écart observé entre l'emplacement théorique et l'emplacement réel du phosphène dans le champ de vision,

la prise en compte ou non des phénomènes d'adaptation temporelle : possibilité de simuler l'extinction d'un phosphène après une trop longue stimulation,

l'utilisation ou non de l'oculomètre : possibilité d'asservir les phosphènes à la position du regard.

Cette configuration est le point d'entrée pour le rendu des phosphènes.

Restitution des images

Les images produites par le module de rendu des phosphènes sont restituées par l'intermédiaire d'un casque de réalité virtuelle pour une plus grande immersion. Au cours de nos expériences, nous en avons utilisés deux différents. Leurs caractéristiques sont décrites dans le Tableau 3-4.

Tableau 3-4 Caractéristiques des deux casques de réalité virtuelle utilisés.

Référence	NVisor SX-60 (NVIS Inc., VA, États-Unis)	Vuzix 1200AR (Vuzix Cor., CA, États-Unis)
Résolution	1280 x 1024 pixels par écran	1074 x 768 pixels par écran
Fréquence	60 Hz	60 Hz
Champ de vision	44 x 33°	28 x 21°

Rendu des phosphènes

Cette brique logicielle, développée en C++, est le cœur de la simulation de vision prothétique. La matrice à simuler est choisie par l'intermédiaire d'un fichier issu du configurateur de matrice. Lorsque ce choix est fait, ce composant génère à la volée une carte de phosphènes prenant en compte tous les paramètres du fichier de configuration.

Les phosphènes sont générés à l'aide de la librairie OpenGL. À l'image de ce qui est préconisé, ils sont de forme ronde et leur bordure est lissée à l'aide d'un filtre Gaussien. Dans nos expériences, le nombre de niveau de luminance des phosphènes est toujours inférieur à 8.

Ce composant met à jour l'affichage des phosphènes toutes les 20 ms (ils peuvent aussi bien être éteints que rester allumés à chaque cycle). Pour déduire la luminance et l'emplacement des phosphènes à afficher, le composant fusionne deux informations provenant du bus logiciel : le niveau de gris de chaque pixel de l'image,

et si l'oculomètre est utilisé, les coordonnées du regard. Quelques exemples d'implants simulés sont illustrés dans la Figure 3-5.

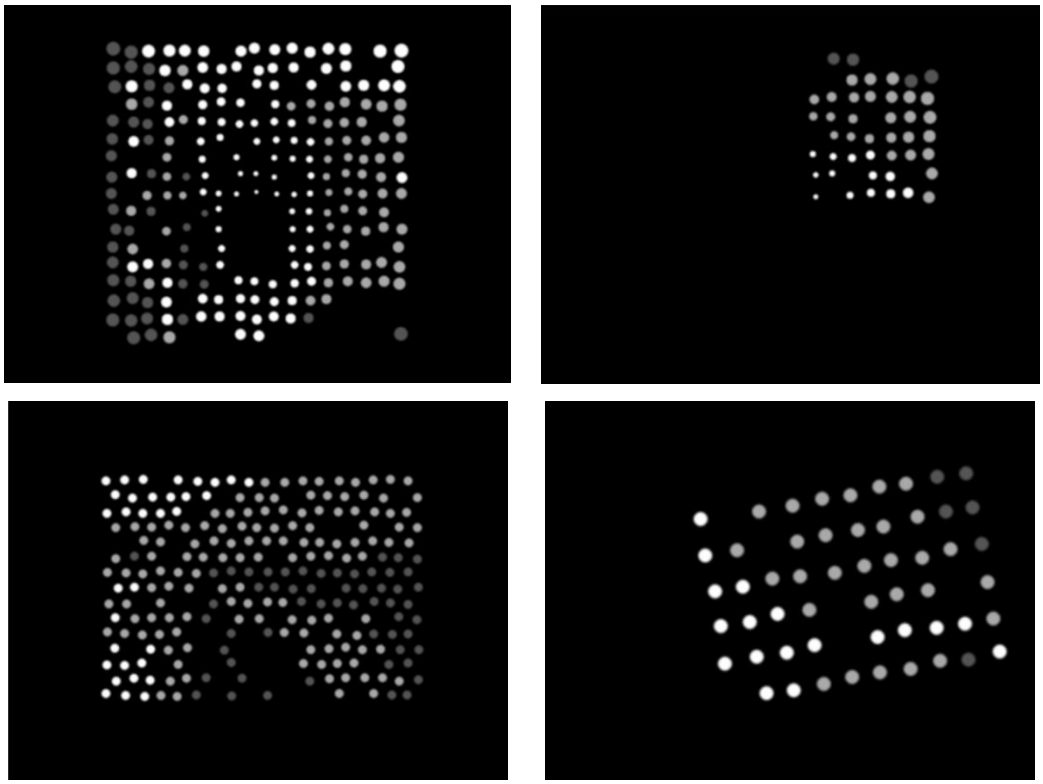


Figure 3-5 Exemples d'implants simulés.
Quatre exemples de simulation d'implants : en haut à gauche, un implant cortical composé de deux matrices de 144 électrodes chacune, implantées dans les deux hémisphères du cortex visuel ; en haut à droite, un implant cortical composé d'une matrice de 50 électrodes implantées dans l'hémisphère gauche ; en bas à gauche, un implant rétinien 15 x 18 électrodes ; en bas à droite, un implant rétinien excentré de la fovéa et légèrement incliné comportant 60 électrodes.

CONCLUSION

Nous avons présenté dans ce chapitre l'architecture générale de notre simulateur de vision prothétique. Cette architecture reproduit les principales caractéristiques des neuroprothèses visuelles équipées d'une caméra externe. Elle est aussi conforme aux différentes recommandations émises pour rendre la simulation la plus réaliste possible. Ce simulateur nous a permis d'évaluer l'approche par localisation de points d'intérêt dans les trois expérimentations qui suivent : la localisation d'objets, de visages et de textes. Pour chacune d'entre elles, la configuration exacte du simulateur sera indiquée. Les matériels et/ou logiciels spécifiques seront détaillés (traitement d'image spécifique, outils de contrôle pour l'expérimentateur…).

CHAPITRE 4

VISION PROTHETIQUE - LOCALISER ET ATTEINDRE UN OBJET

INTRODUCTION

Notre système visuel, nous permet de localiser, discriminer et reconnaître des formes dans des environnements extrêmement complexes et dans des conditions de luminosité très variées. Ces fonctions sont essentielles à notre quotidien. Les neuroprothèses visuelles ont pour objectif de restaurer partiellement la vision des personnes atteintes de cécité, afin d'améliorer leur autonomie, notamment dans des tâches où la vision est prépondérante, comme la détection, la reconnaissance, et la saisie d'objets. Ces tâches ont été testées en condition réelle ou en simulation et les premiers résultats sont encourageants mais restent limités.

Chez des sujets implantés, nous retrouvons quelques données portant sur la localisation et la reconnaissance d'objet. En 2006, Duret et al. publient leurs résultats avec une patiente ayant reçu un implant pour stimuler son nerf optique [Duret et al. 2006]. Sa neuroprothèse lui permet de distinguer 109 phosphènes dans un champ de vision de 14 x 41°. La caméra couvre horizontalement 108° de champ de vision. Ses tâches consistent par exemple à localiser une tasse blanche (9 positions possibles), la discriminer parmi 5 autres objets, et enfin la saisir [Duret et al. 2006]. Après un long entraînement, la personne implantée réalise correctement ces 3 tâches. Ses performances restent cependant très faibles (20 secondes pour localiser, 40 secondes pour discriminer) malgré la simplicité de l'environnement. Jens Naumann, équipé d'une neuroprothèse corticale (72 électrodes dans chaque hémisphère du cortex visuel), relate dans son livre sa capacité à discerner des formes et même certains objets [Naumann 2012]. Enfin, très récemment, des résultats sur des tâches de reconnaissance de formes et d'objets sont également rapportés pour les deux implants rétiniens les plus aboutis à ce jour. Avec l'Alpha-IMS (1500 photodiodes en sous-rétinien couvrant 11 x 11° de champ visuel), 3 sujets sont capables de localiser des objets blancs posés sur une nappe noire. L'un d'entre eux identifie aussi de larges formes géométriques blanches [Stingl et al. 2013]. La localisation d'objets fortement contrastés est aussi possible avec l'Argus II (60 électrodes en épirétinien pour un champ de vision d'environ 13°x17°) [Humayun et al. 2012].

Devant la difficulté et les limitations inhérentes aux expériences impliquant les rares patients implantés jusqu'à présent, de nombreuses équipes de recherche se sont tournées vers des expériences de simulation de neuroprothèse visuelle avec des sujets voyants. Plusieurs de ces travaux de simulation portent spécifiquement sur l'identification et la reconnaissance d'objets. Ils montrent que plusieurs centaines de phosphènes distincts sont nécessaires pour y parvenir. Dagnelie et al. rapportent les premiers résultats, en 2001, dans une expérience en environnement virtuel [Dagnelie et al. 2001]. Les 4 sujets devaient transférer (attraper et poser) des objets virtuels avec une vision pixélisée, mais la taille de la matrice utilisée n'est pas indiquée dans cette étude. En 2003, Hayes et al. simulent trois matrices de tailles différentes (4x4, 6x10 et 16x16) [Hayes et al. 2003]. Ils demandent à 8 sujets de décrire et d'identifier quatre objets blancs (une assiette, une tasse, une cuillère et un stylo) posés sur un fond noir. La seule information fournie aux sujets est que ces objets sont des objets courants. La description devient possible à partir de la matrice 6x10, en revanche l'identification nécessite la plus grande matrice (256 électrodes). Dagnelie et al. étudient en 2006 l'identification et le déplacement d'objets en simulant un implant de 6x10 électrodes [Dagnelie, Walter, et al. 2006]. La première tâche consiste à compter les cases blanches d'un échiquier modifié pour l'expérience (1 à 16 cases blanches sur un total de 64 cases). Dans la seconde tâche, les sujets doivent bouger des pièces noires de l'échiquier sur des cases blanches. Tous les sujets sont capables d'accomplir les deux tâches avec très peu d'erreurs après une période d'apprentissage. En 2008, Fornos et al. investiguent l'impact de la résolution de l'implant et du champ de vision de la caméra dans une tâche de reconnaissance et de déplacement de formes [Pérez Fornos et al. 2008]. Le champ visuel des sujets est restreint à 10°x7°. Ils évaluent 5 tailles de matrices (de 124 à 17920 électrodes) et 3 champs pour la caméra (8,2°x5,8°, 16,5°x11,6°, et 33,0°x23,1°). Leurs résultats indiquent qu'un minimum de 500 phosphènes est nécessaire pour effectuer cette tâche. Zhao et al. ont également effectué une expérience de reconnaissance d'objets [Y. Zhao et al. 2010]. Leurs sujets doivent identifier 20 objets du quotidien avec 6 résolutions de matrices différentes (de 8x8 jusqu'à 64x64 électrodes). Sans surprise, plus la résolution est grande, meilleur est le taux de réussite. Pour atteindre une performance supérieure à 60% d'identification correctes, la résolution nécessaire est comprise entre 256 (16x16) et 576 (24x24) phosphènes. Ces résultats sont confirmés par d'autres études plus récentes [Hu et al. 2013; Lu et al. 2013; Li et al.

2011]. L'ensemble des études publiées sur la reconnaissance d'objets semble montrer que si le champ de vision de la caméra est suffisamment large, un minimum de 500 phosphènes distincts est nécessaire pour réaliser cette tâche. Cependant, tous ces tests ont été réalisés dans des conditions quasi-idéales (fort contraste, peu d'objets…), et il est fort probable que bien plus de phosphènes soient requis pour localiser et reconnaître des objets dans des situations plus naturelles.

Aujourd'hui, et au vu de l'ensemble de ces résultats, les neuroprothèses actuelles, d'une résolution encore limitée, ne permettent pas de localiser et de reconnaître la majorité des objets de notre quotidien dans un environnement naturel. Pour les neuroprothèses équipées d'une caméra externe, il est cependant possible de traiter l'image capturée en utilisant des algorithmes de vision par machine pour modifier le rendu de la scène visuelle affichée via la neuroprothèse. Dans, les deux expériences qui suivent, nous proposons d'aider les sujets à trouver un objet précis, en couplant à notre simulateur de vision prothétique un algorithme de reconnaissance de forme. Pour ce faire notre rendu visuel est basé sur l'approche par localisation de points d'intérêt décrite dans le Chapitre 2 : nous extrayons, en temps réel, la position de l'objet dans l'image, et cette position est restituée sous la forme d'un unique phosphène, de telle sorte que le sujet perçoive sans ambigüité où se situe l'objet d'intérêt dans son champ de vision. La première experience a pour objectif de valider cette approche, la seconde permet de la comparer à l'approche classiquement implémentée dans les systèmes actuels : l'approche « scoreboard » telle que définie dans le Chapitre 2.

EXPERIENCE 1

Cette première expérience[14] a pour objectif de démontrer que l'approche par localisation de points d'intérêt est fonctionnelle dans une situation réaliste, et qu'elle est compatible avec l'utilisation d'un implant cortical basse résolution. Le rendu visuel testé dans cette expérience est très simple : on affiche dans le champ de vision du sujet un unique phosphène indiquant la position de l'objet recherché. Pour tester et valider cette approche, nous avons choisi de simuler quatre configurations de phosphènes différentes : une configuration centrale optimale qui nous sert de condition de contrôle (100 phosphènes couvrant tout le champ de vision Figure 4-1A), une configuration centrale minimale pour simuler un implant basse résolution (9 phosphènes proche de la fovéa Figure 4-1C), et deux configurations excentrées (100 et 9 phosphènes latéraux Figure 4-1B et D). Dans les deux configurations contenant 9 phosphènes, ces derniers sont arrangés sous la forme d'une croix afin d'utiliser cette disposition particulière pour indiquer simplement au sujet une direction à suivre (gauche/droite/haut/bas). Les deux configurations latérales correspondent à une implantation d'électrodes plus simple (un seul hémisphère du cortex visuel), et nous permettent de mesurer l'effet de la position des électrodes sur la performance.

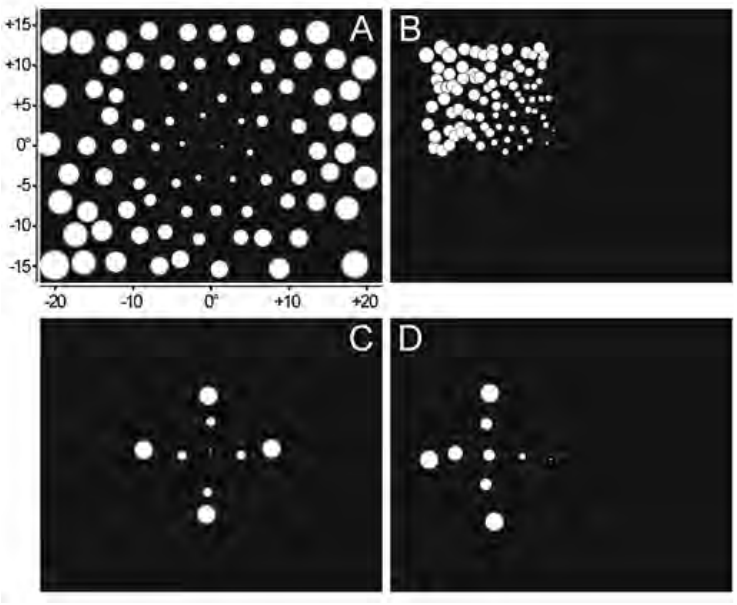


Figure 4-1 Vision prothétique avec les quatre matrices utilisées.
Ici, tous les phosphènes possibles sont affichés simultanément. A. 100Central B. 100Lateral C. 9Central D. 9Lateral.

[14] Un papier pour le journal Artificial Organs est cours de soumission.

Hypothèses de travail

Dans cette première expérience nous avons posé les hypothèses suivantes:

Hypothèse n°1 : il est possible de localiser et d'atteindre un objet avec un rendu visuel basé sur l'approche par localisation de points d'intérêt, en simulant un implant cortical de très basse résolution (quelques électrodes seulement).

Hypothèse n°2 : une vision prothétique excentrée peut engendrer une diminution de performance comparée à une vision centrale (en cortical une vision centrale correspond à une implantation de deux matrices d'électrodes, une dans chaque hémisphère du cortex visuel).

Protocole expérimental

Quatorze sujets voyants (6 femmes et 8 hommes ; entre 22 et 33 ans) ont participé à cette expérience. Deux d'entre eux avaient une connaissance préalable du simulateur de vision prothétique.

Les sujets étaient équipés d'un simulateur de vision prothétique et d'écouteurs audio au travers desquels les instructions leurs étaient indiquées. Ils étaient assis sur une chaise pivotante, face à une table en quart de cercle dont le centre est évidé (Figure 4-2). Sept objets de la vie courante (une souris tactile, un téléphone sans fil, une carte bleue, une tasse, une petite bouteille de lait, une télécommande et un pot de pâte à tartiner) étaient disposés de façon pseudo-aléatoire sur cette table. La tâche consistait à localiser et toucher, le plus rapidement et le plus précisément possible, un objet en particulier.

Figure 4-2 Dispositif.
Le sujet était équipé du simulateur de vision prothétique. Il était face à une table sur laquelle étaient disposés sept objets. Sa tâche consistait à localiser et toucher un des objets.

L'expérience comprenait deux phases : une phase d'entraînement, d'une durée de 5 minutes, où le sujet se familiarisait avec le dispositif et apprenait à réaliser la tâche, et une phase d'expérimentation pendant laquelle nous mesurions les performances obtenues, c'est-à-dire le pourcentage d'essais réussis et le temps pour réaliser cette tâche. La phase d'expérimentation était divisée en 5 blocs. Chaque bloc correspondait à l'utilisation d'un implant cortical particulier (une configuration d'électrodes donnée). Un bloc était lui-même divisé en 6 séries de 7 essais pour un total de 42 essais. Chaque sujet réalisait donc un total de 210 essais (5 blocs x 42 essais) pour une durée totale d'environ 2H30. Un questionnaire sur la tâche, les différents blocs et l'expérimentation dans son ensemble clôturaient l'expérimentation.

Sur la table, nous avions préalablement défini 11 positions sur lesquelles les objets étaient placés. Le sujet n'avait pas connaissance de ces positions. La grande majorité des sujets ont indiqué à ce propos dans le questionnaire post-expérience, que selon eux, les objets étaient positionnés de façon totalement aléatoire sur la table. Avant le démarrage d'une série, nous mettions en place une configuration d'objets particulière : nous disposions les objets sur 7 des 11 positions de la table. Chacune de ces configurations avait été créée aléatoirement. Pendant cette mise en place, le sujet écoutait de la musique pour éviter qu'il ne perçoive le moindre indice sonore pouvant l'aider à localiser un ou plusieurs objets.

Chaque essai démarrait par l'énonciation, dans les écouteurs, de l'objet à toucher. Le sujet devait alors localiser et toucher cet objet à l'aide de sa vision prothétique simulée. Pour y parvenir, nous restituions dans le casque de réalité virtuelle une

information extrêmement simpliste : soit le sujet ne voyait pas de phosphène parce que l'objet cible n'était pas détecté dans le champ de vue de la caméra, soit le sujet voyait un unique phosphène indiquant la position de l'objet cible dans son champ de vision (Figure 4-3).

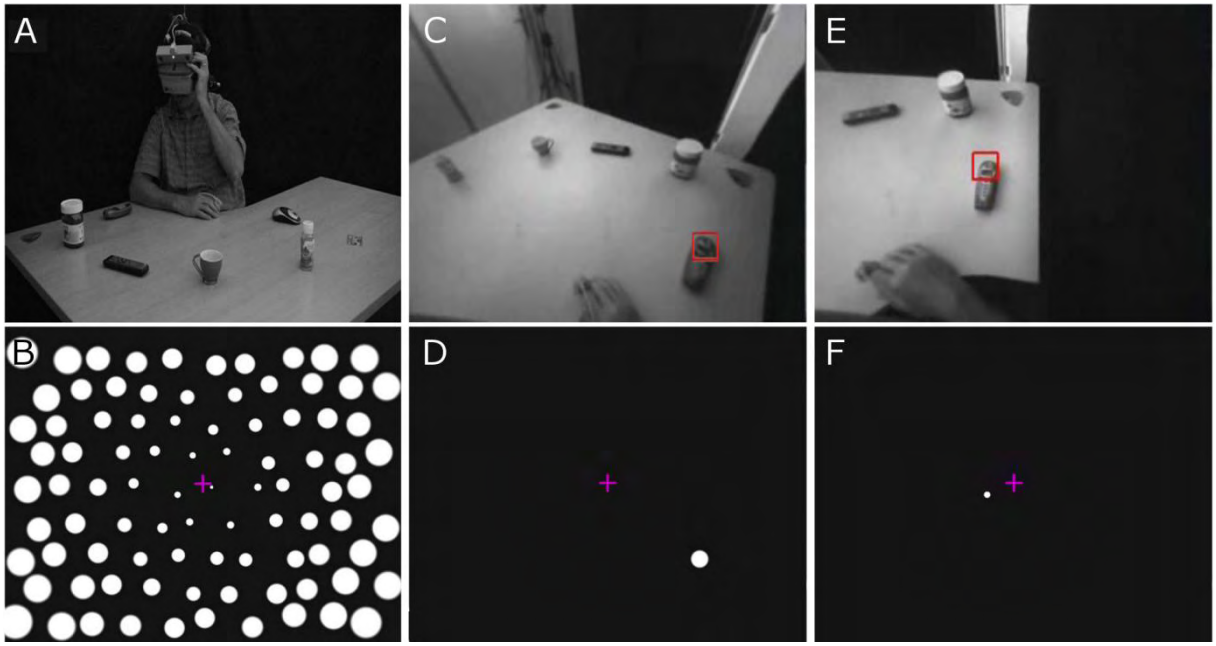


Figure 4-3 Tâche de localisation d'objet.
A. Le sujet parcours l'environnement à la recherche du téléphone. B. Vue d'ensemble des phosphènes qu'il est possible d'activer avec l'exemple d'un implant cortical à 100 électrodes. C et E. Vue caméra : un algorithme de vision artificielle repère le téléphone dans l'image de la caméra. D et F. L'unique phosphène qui s'allume dans le casque de réalité virtuelle indique la position du téléphone dans le champ de vision du sujet.

Nous considérions l'essai comme correct si le premier objet touché était bien celui qui avait été énoncé. Dans l'éventualité où le sujet touchait accidentellement, avec son poignet ou son coude, un autre objet se trouvant sur son chemin, celui-ci était immédiatement replacé à une des 4 positons inoccupée parmi les 11, et l'essai se poursuivait normalement. Dans tous les autre cas, l'essai était considéré comme incorrect. À la fin de l'essai, le sujet replaçait ses mains devant lui avant de démarrer l'essai suivant. Lorsque le sujet avait terminé 7 essais, une nouvelle configuration d'objet était mise en place pour la série suivante. Après les 6 séries d'un bloc, nous changions l'implant simulé, et passions au bloc suivant (c'est-à-dire, à un nouvel implant simulé).

Simulateur

Implants simulés

Pendant l'expérience, chacun des blocs correspondait à la simulation d'un implant cortical spécifique (Figure 4-1). Un premier implant, 100Central, servait de référence. Il correspondait à 100 électrodes régulièrement disposées dans les deux hémisphères du cortex visuel primaire. C'était une configuration optimale, qui était comparée à un second implant, 100Lateral, qui contenait le même nombre d'électrodes à la différence qu'elles étaient réparties sur un seul hémisphère. Les troisième et quatrième neuroprothèses corticales modélisées, 9Central et 9Lateral, étaient à l'image des deux premières mais avec une résolution réduite à 9 électrodes.

Pour l'ensemble des implants, les phosphènes simulés étaient blancs et d'une taille comprise entre 0,4° et 3,4° pour prendre en compte le facteur de magnification corticale. La position des phosphènes était bruitée, car dans la réalité, elle ne correspond jamais parfaitement à la position théorique sur la grille d'électrode. De plus, sur les deux implants contenant 100 électrodes, 10% d'entre elles étaient rendues défaillantes et ne produisaient pas de phosphènes, ce qui correspond aux observations sur les premiers implants chez l'homme. En revanche pour les deux petites matrices, les 9 électrodes étaient toutes fonctionnelles car nous avons estimés qu'elles pouvaient être contrôlées individuellement au moment de l'implantation, et replacées en cas de dysfonctionnement.

Pour tous les sujets, les blocs 1 et 5 étaient réalisés avec l'implant de référence, 100Central, pour vérifier un éventuel effet d'apprentissage. L'ordre de présentation des trois autres matrices était choisi de manière aléatoire pour les blocs 2, 3 et 4, toutes les combinaisons possibles étant pseudo-contrebalancées sur l'ensemble des sujets.

Architecture

L'architecture générale développée pour cette expérience est résumée dans la Figure 4-4. La reconnaissance d'objet s'est faite au travers de SNVision, le moteur de reconnaissance de formes bio-inspiré développé par la SpikeNet Technology (les détails ont été fournis au chapitre 2). Nous avions modélisé les sept objets à différents endroits sur la table. Chaque objet requérait 15 à 35 modèles pour être reconnu à n'importe quelle position. De plus, pour limiter les éventuelles fausses détections, l'utilisation d'une caméra binoculaire permettait de calculer la distance des cibles détectées et de ne pas tenir compte des reconnaissances ayant lieu à plus de 80 cm du sujet (profondeur de la table). Au cours de l'expérience, lorsqu'un essai démarrait, les modèles de l'objet à localiser étaient chargés dynamiquement dans SNVision. Si l'objet était reconnu dans une image capturée par la caméra, les coordonnées 3D du point correspondant au centre de cette détection étaient récupérées. Puis, le phosphène se trouvant être le plus proche de ce point dans le champ de vision du sujet était allumé.

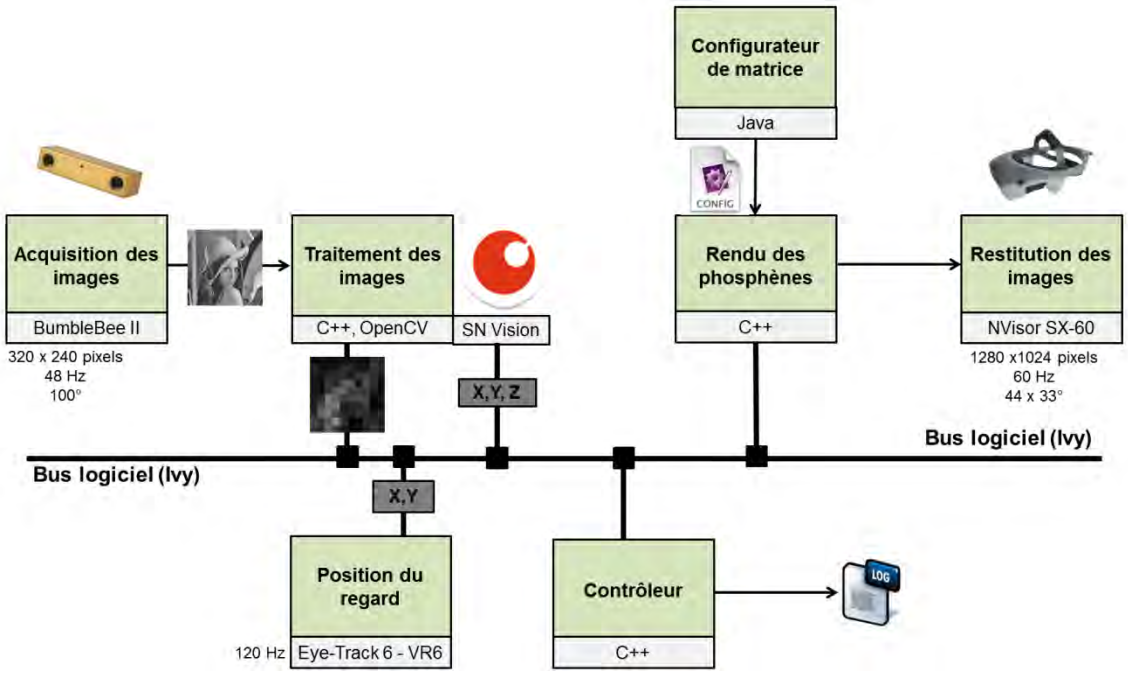


Figure 4-4 Architecture du simulateur pour l'expérience de localisation d'objets.
L'acquisition des images se faisait par l'intermédiaire d'une caméra stéréoscopique. Le logiciel SNVision, développé par la société SpikeNet Technology, était utilisé pour localiser des objets. La position de l'objet détecté était restituée dans un casque de réalité virtuelle à l'aide de phosphènes. Le simulateur intégrait également un eye-tracker pour la gestion de la position du regard.

Comme l'expérience était menée avec des sujets voyants et non avec des patients implantés, il a fallu procéder à quelques adaptations dans le fonctionnement du dispositif. Il n'est pas possible pour une personne implantée (cortical ou rétinien), d'explorer librement son rendu phosphénique car les phosphènes suivent les mouvements oculaires. Dans le simulateur en revanche, les sujets peuvent faire cette exploration, ce qui n'est pas réaliste. Pour annuler, ou tout au moins minimiser ce problème, nous demandions aux sujets de maintenir les yeux fixés sur une petite croix violette placée au centre de l'écran. L'exploration de l'environnement se faisait alors par des mouvements de tête. Nous contrôlions les mouvements oculaires par l'intermédiaire d'un oculomètre (Eye-Track 6 - VR6; Applied Science Laboratories, MA, États-Unis) cadencé à 120 Hz et placé dans le casque de réalité virtuelle. Cependant, cette installation n'était pas assez fiable car l'oculomètre avait tendance à bouger par rapport aux yeux, ce qui obligeait à réétalonner très régulièrement. Finalement ce contrôle minutieux ne sera effectué que pour deux sujets. Pour les 12 autres, nous vérifions manuellement par l'intermédiaire d'un écran de contrôle qu'ils fixaient bien la croix de fixation du regard. Dans le cas contraire, nous leur rappelions la consigne.

Un logiciel développé en C++ permettait de piloter l'expérience et de récupérer l'ensemble des données de chacun des sujets. Il était lancé sur un premier ordinateur (Intel Core 2 Duo, 2,26 GHz). Un second ordinateur (Intel Core 2 Quad CPU, 2,26GHz) était dédié au traitement d'image (capture et reconnaissance d'objet). Enfin, un troisième ordinateur (Intel Pentium Duo, 2,99GHz) calculait le rendu des phosphènes et était connecté à l'oculomètre. Les différents logiciels communiquaient par messages asynchrones envoyés sur un bus logiciel commun [Buisson et al. 2002].

Résultats

Les résultats que nous présentons concernent deux mesures : la précision (le pourcentage d'essai correct) et le temps de mouvement (le temps entre l'annonce de l'objet cible et son toucher par le sujet). L'analyse des données a été réalisée avec Matlab (Matworks, Natick, MA, États-Unis) et les tests statistiques avec R (R Foundation, États-Unis). Nous avons utilisés des tests non-paramétriques, nos données n'étant pas normales, sans moyen évident de les transformer. La

comparaison entre deux groupes ou conditions a été effectuée grâce à un test de Wilcoxon. Au-delà de deux conditions nous avons utilisé l'Anova de Friedman. Pour l'ensemble de nos tests, le seuil de significativité a été fixé à 0,05.

Si l'on considère l'ensemble des 14 sujets, le temps médian pour toucher le bon objet dans les cinq conditions est d'environ 20 secondes. Nous avons enlevés de notre analyse les essais dont le temps de mouvement est au-delà de trois fois l'écart-type moyen soit 72,2 secondes (1,4% de la totalité des essais). Une analyse des vidéos montre que ces essais étaient liés en majorité à des problèmes techniques (câblage gênant, oculomètre, inconfort du sujet, etc.).

Le rapport entre précision et temps de mouvement par sujet est représenté dans la Figure 4-5. Les résultats du sujet 12 contrastent avec ceux des 13 autres sujets. Sa précision est plus de trois écart-types en dessous de la moyenne de tous les sujets, et son temps de mouvement plus de 3 écart-types plus long que la moyenne de tous les sujets. Du fait de ses valeurs extrêmes, le sujet a été exclu de l'analyse qui suit. Sans ce sujet, la précision globale est de 85,4% (+-5,9) correct et le temps médian de mouvement de 18,1 s (± 3,8).

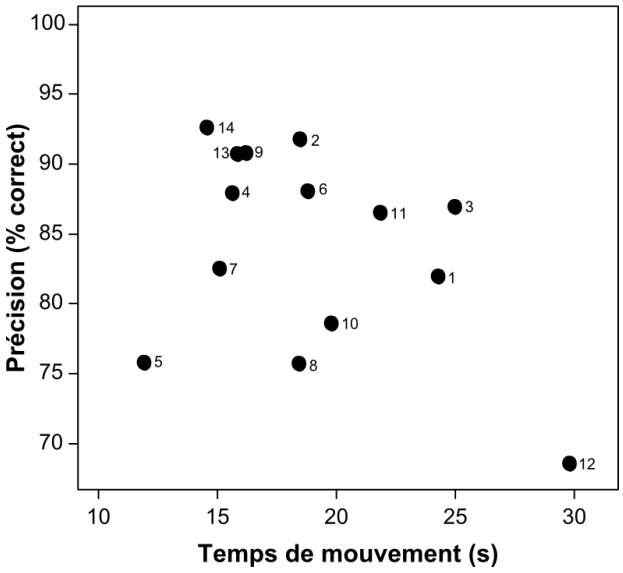


Figure 4-5 Précision et temps de mouvement. Précision (%) en fonction du temps de mouvement médian (s) pour les 14 sujets. En bas à droite, les données du sujet 12 apparaissent clairement comme extrêmes.

L'implant 100Central était utilisé pendant le premier (100Central1) et le dernier bloc (100Central2) pour observer un éventuel effet d'apprentissage. La Figure 4-6 illustre l'évolution de la performance entre ces deux blocs. Des tests de Wilcoxon sur la précision ($z=-2{,}31$, $p<0{,}018$) et le temps de mouvement ($z=1{,}72$, $p=0{,}048$) montrent des différences significatives : les sujets sont plus précis et plus rapides à la fin de l'expérience. Bien qu'il soit significatif, cet effet est limité. Après plus de deux heures d'expérience, les sujets sont plus rapides en moyenne d'une demi-seconde, et augmentent leur précision de 6%. Pour réduire ce léger effet d'apprentissage, la condition 100Central2 correspondant au bloc n°5 est utilisée par la suite comme référence des résultats de l'implant 100Central.

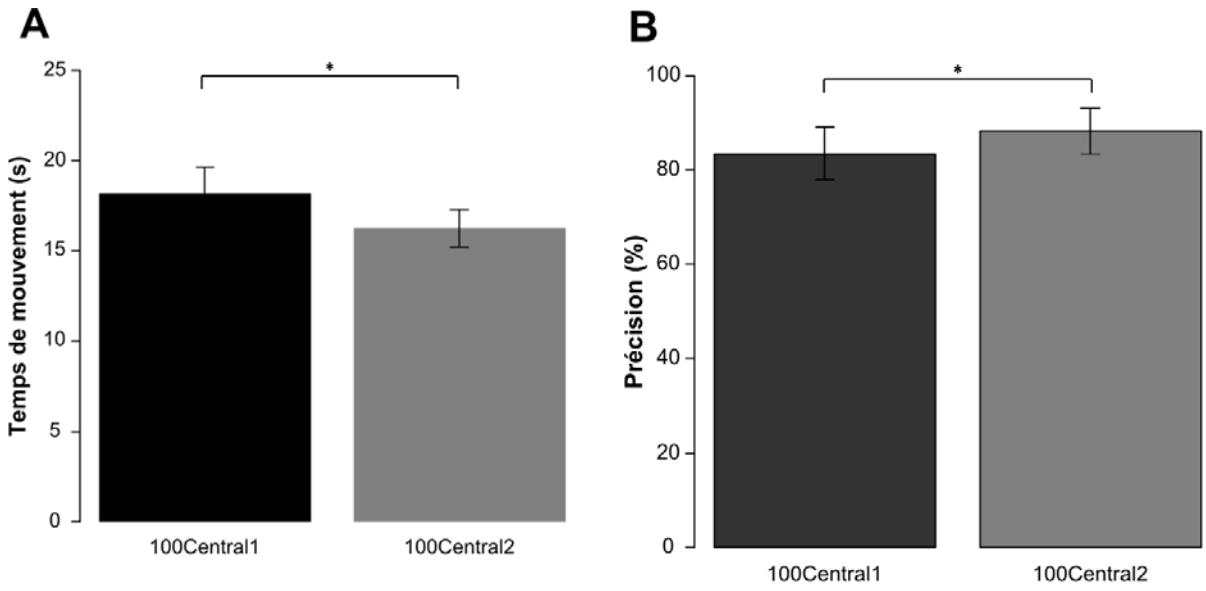


Figure 4-6 Moyenne des temps de mouvement et de la précision.
Moyenne des temps de mouvement (A) et de la précision (B) pour les conditions 100Central1 et 100Central2 (* $p<0.05$).

La Figure 4-7 représente la performance des sujets obtenus avec les 4 matrices différentes. Un test de Friedman révèle que le type d'implant utilisé a un effet significatif sur le temps de mouvement ($\chi^2=7{,}98$, $df=3$, $p=0{,}046$) et non sur la précision ($\chi^2=7{,}14$, $df=3$, $p=0{,}067$). Les tests post-hoc sur le temps de mouvement n'indiquent, en revanche, aucune différence significative entre chaque type d'implant.

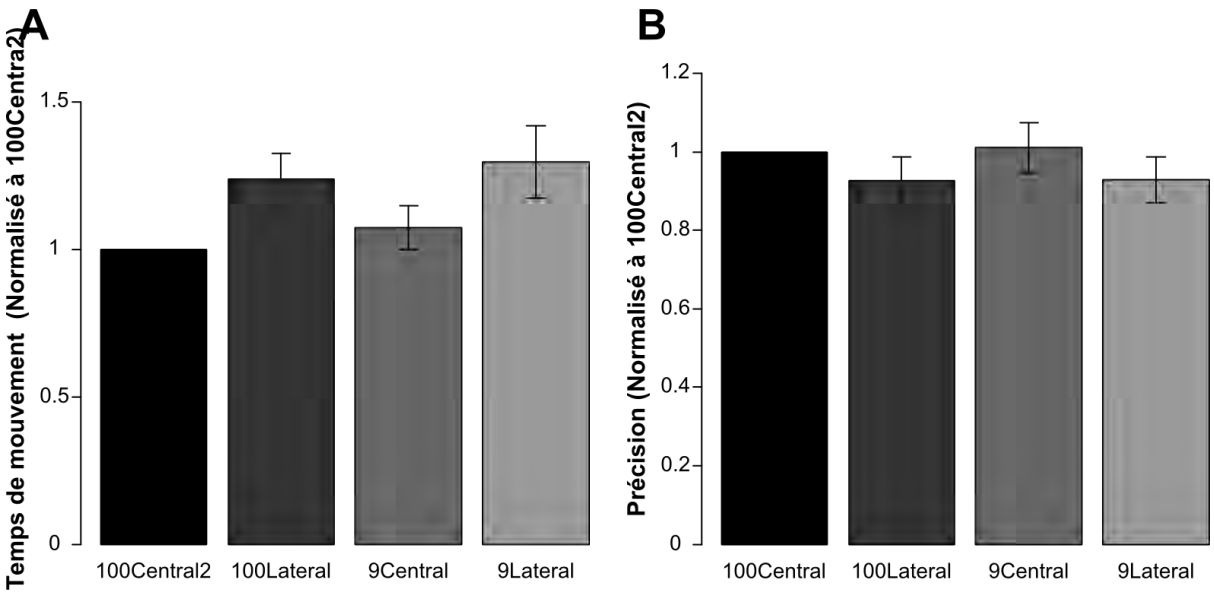


Figure 4-7 Moyenne des temps de mouvement (A) et de la précision (B).
Résultat pour les 4 matrices utilisées, normalisées à la performance obtenue en 100Central2.

Pour savoir si le rendu visuel basé sur l'approche par localisation de points d'intérêt est adapté pour cette tâche, même avec très peu d'électrodes (H1), il convient de comparer les performances obtenues pour un implant de 9 électrodes avec celles obtenues pour un implant de 100 électrodes. Cette comparaison (9Central et 9Lateral vs. 100Central2 et 100Lateral) ne présente pas de différences significatives, que ce soit en temps de mouvement ($z=-1,12$, $p=0,26$) ou en précision ($z=-0,47$, $p=0,69$).

Concernant l'impact de la position de l'implant sur les performances (H2), nous confrontons les résultats pour les deux matrices centrales (100Central2 et 9Central) avec ceux obtenus pour les deux matrices latérales (100Lateral et 9Lateral). Un test de Wilcoxon révèle une différence significative sur le temps de mouvement ($z=2,75$, $p<0,001$) et sur la précision ($z=-4,17$, $p<0,001$). Ces résultats indiquent que la performance est meilleure lorsque les phosphènes sont présentés en vision centrale plutôt qu'en périphérie.

Si nous regardons de plus près les performances par objet, un test de Friedman révèle une différence significative autant pour la précision ($\chi^2= 17,81$, $df = 6$, $p<0,001$) que pour le temps de mouvement ($\chi^2= 34,32$, $df = 6$, $p<0,01$). Les tests post-hoc indiquent que la carte de crédit est plus longue à atteindre que tous les autres objets. La précision est la meilleure avec le pot de pâte à tartiner, et la moins

bonne pour la carte de crédit et la petite bouteille de lait. Ce résultat s'explique notamment par le fait que les objets les plus petits sont plus difficiles à atteindre.

Discussion préliminaire

Nous discutons dans cette section très brièvement des premières conclusions issues de ces résultats. Comme l'expérience suivante se situe dans un contexte expérimental semblable, une discussion plus exhaustive, commune à ces deux expériences, est proposée à la fin de ce chapitre.

Validation ou rejet des hypothèses

Les résultats de cette première expérience montrent qu'un rendu visuel très simple (un unique phosphène), basé sur l'approche par localisation de points d'intérêt, et limité à l'affichage d'un unique phosphène, permet de localiser et d'atteindre des objets posés sur une table. En simulant un implant cortical très basse résolution (constitué de seulement neuf électrodes), tous les sujets arrivent à effectuer correctement cette tâche grâce au rendu visuel adapté (validation de l'hypothèse n°1). Pour cette tâche en particulier, l'utilisation d'un algorithme de détection d'objets rend possible l'utilisation de matrices de très petites tailles.

La performance n'est visiblement pas dépendante du nombre d'électrodes puisque les résultats obtenus en simulant un implant cortical comprenant neuf électrodes sont sensiblement identiques à ceux obtenus avec dix fois plus d'électrodes. En revanche, la position de l'implant dans le cortex a un effet sur la performance des sujets : ils sont moins précis et moins rapides avec une vision phosphénique latérale qu'avec une vision centrale (validation de l'hypothèse n°2). Comme ce qui a été rapporté dans des travaux sur la lecture, on peut faire la supposition que la précision en vision excentrée s'améliorerait avec un peu plus d'entraînement [Sommerhalder et al. 2003; Sommerhalder et al. 2004].

Cette première expérience valide l'approche par localisation de points d'intérêt. L'étape suivante consiste à la confronter aux stratégies de restitution classiquement utilisées dans les neuroprothèses visuelles, à savoir les rendus basés sur l'approche scoreboard. C'est ce que nous proposons de tester dans l'expérience qui suit.

EXPERIENCE 2

Dans cette expérience[15], nous restons dans un contexte de la localisation d'objets du quotidien pour confronter cette fois-ci l'approche par localisation de points d'intérêt à l'approche scoreboard classiquement implémentée dans les neuroprothèses visuelles. Les implants rétiniens étant à ce jour les systèmes les plus aboutis, nous choisissons cette fois-ci (et pour toutes nos expériences futures) de simuler des implants épirétiniens utilisés actuellement dans les essais cliniques ou disponibles dans les années futures selon les informations fournies par les industriels. Trois résolutions d'implant sont simulées : 6x10 électrodes, 15x18 électrodes et 32x38 électrodes. La première correspond à la génération actuelle d'implants épirétiniens, la seconde à la prochaine génération (trois à cinq ans), et la troisième est imaginée comme étant possible à l'échelle de dix ou quinze ans.

Hypothèses de travail

Nous posons les hypothèses suivantes :

Hypothèse n°3 : l'approche par localisation de points d'intérêt permet d'atteindre un objet avec un implant rétinien basse résolution (ici, 6x10 électrodes). En d'autres termes nous pensons confirmer la validité de l'hypothèse n°1 de la première expérience.

Hypothèse n°4 : contrairement à l'approche par localisation de points d'intérêt, l'approche scoreboard nécessite un nombre très important de phosphènes distincts pour détecter et atteindre des objets.

Protocole expérimental

Douze sujets voyants (4 femmes et 8 hommes ; entre 23 et 46 ans) ont participé à cette expérience. La moitié d'entre eux était familiarisée avec le simulateur de vision prothétique (groupe expert) et l'autre n'avait jamais utilisé ce dispositif (groupe novice).

[15] Un papier pour le journal « Journal of Neural Engineering » est en finalisation d'écriture. Une partie des résultats a été publiée dans [Denis et al. 2012].

La tâche était similaire à celle de la première expérience. Les sujets étaient assis sur une chaise pivotante face à une table semi circulaire. Ils devaient localiser et toucher huit objets de la vie courante parmi dix (Figure 4-8). Les deux autres objets faisaient office de distracteurs. Ces objets étaient disposés sur la table parmi 14 positions prédéfinies. Les sujets n'avaient ni connaissance des objets, ni des positions potentielles. L'objectif était toujours de localiser et toucher un objet le plus rapidement et le plus efficacement possible avec une vision prothétique simulée. Nous avions fixé un temps maximum de réponse de 60 secondes par essai.

Chaque sujet devait effectuer la tâche avec trois implants de résolutions différentes dans l'approche scoreboard, et avec l'implant de plus faible résolution dans l'approche par localisation. L'ordre de ces quatre conditions était contrebalancé parmi les 12 sujets afin de limiter les éventuels effets d'apprentissage. Chaque condition était divisée en 3 séries de 8 essais pour un total de 24 essais par condition et de 96 essais au total (24 essais x 4 conditions). À la fin de chaque essai, le simulateur s'éteignait, et nous changions de position l'objet que le sujet devait trouver et celui qu'il allait devoir trouver dans l'essai suivant.

Avant chaque nouvelle condition, le sujet avait dix minutes pour observer, avec sa vision prothétique simulée, l'ensemble des dix objets. Le reste du protocole était identique à celui décrit dans l'expérience précédente. À la fin de l'expérience, les sujets remplissaient un questionnaire (Annexe 1).

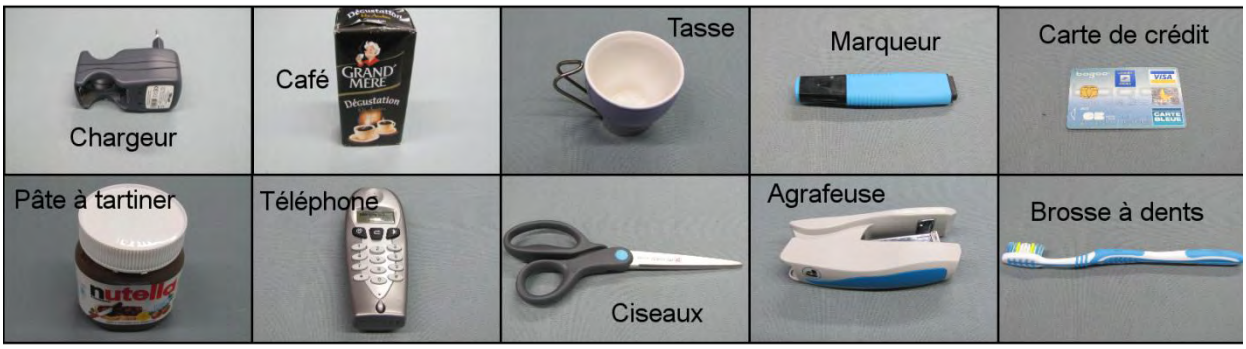


Figure 4-8 Objets utilisés dans l'expérience de localisation.
Dix objets de la vie courante utilisés dans la deuxième expérience de localisation d'objets. Deux d'entre eux (la carte de crédit et la brosse à dents) sont des distracteurs (ils ne sont jamais cibles).

Simulateur

Implants simulés

Pour cette expérience, trois implants rétiniens étaient simulés (Tableau 4-1). Ils contenaient 6x10 (A1), 15x18 (A2) et 32x38 (A3) électrodes. Le premier implant correspond à l'actuelle génération d'implants épirétiniens. Les deux autres n'existent pas encore mais l'implant 15x18 est en projet depuis plusieurs années chez Second Sight et le dernier envisageable à plus long terme. Le champ de vision restitué était respectivement de 11x11°, 13x13° et 26x26°. Les électrodes étaient disposées de façon rectangulaire dans la première matrice (exemple de l'Argus II) et de façon hexagonale dans les deux matrices de résolution plus importante (conforme à ce qui est envisagé pour les futurs implants car cette disposition permet d'augmenter la densité d'électrodes par unité de surface). La taille des phosphènes variait entre 1,0° (A1) et 0,7° (A2 et A3). 1° correspond en moyenne à la taille des phosphènes induits par des électrodes de diamètre de 200µm (les électrodes de l'Argus II par exemple) Nous avons considéré que pour les générations futures (A2 et A3), les électrodes seraient un peu plus fines (0,7° correspondrait à des électrodes de 150µm diamètre). Les phosphènes étaient espacées de respectivement de 1,2° (A1) et 0,8° (A2 et A3). 8 niveaux de luminance étaient disponibles pour chacun des phosphènes. Nous avons simulé également 10% d'électrodes non fonctionnelles (moyenne aujourd'hui rapportée qui tend quand même à diminuer).

Tableau 4-1 Les trois implants rétiniens simulés pour cette expérience.

	Nombre elec.	Disposition elec.	Forme phosph.	Taille phosph.	Distance phosph.	Niveaux de gris	Champ de vision	Dropout
A1	6 x 10	Rectangulaire	rond	1,0°	1.2°	8	11° x 11°	10%
A2	15 x 18	Hexagonal	rond	0,7°	0.8°	8	13° x 13°	10%
A3	32 x 38	Hexagonal	rond	0,7°	0.8°	8	26° x 26°	10%

Enfin, nous avons souhaité modéliser dans cette expérience la dynamique temporelle des phosphènes : lors d'une stimulation électrique continue, les sujets rapportent une diminution progressive de la luminance des phosphènes. Ce phénomène n'est pas clairement expliqué mais pourrait provenir d'un effet d'adaptation ou encore d'une fatigue neuronale. Pour représenter cette dynamique,

les phosphènes étaient pilotés individuellement et ils s'éteignaient pendant environ 100 ms lorsqu'ils n'avaient pas été rafraichis depuis plus de 2 secondes. Depuis que cette expérience a été conçue et réalisée, Fornos et al. ont montré que la dynamique temporelle des phosphènes est probablement plus complexe et qu'elle varie énormément selon les sujets [Pérez Fornos et al. 2012].

La **Erreur ! Référence non valide pour un signet.** illustre la vision prothétique simulée correspondant à ces trois résolutions pour une approche scoreboard.

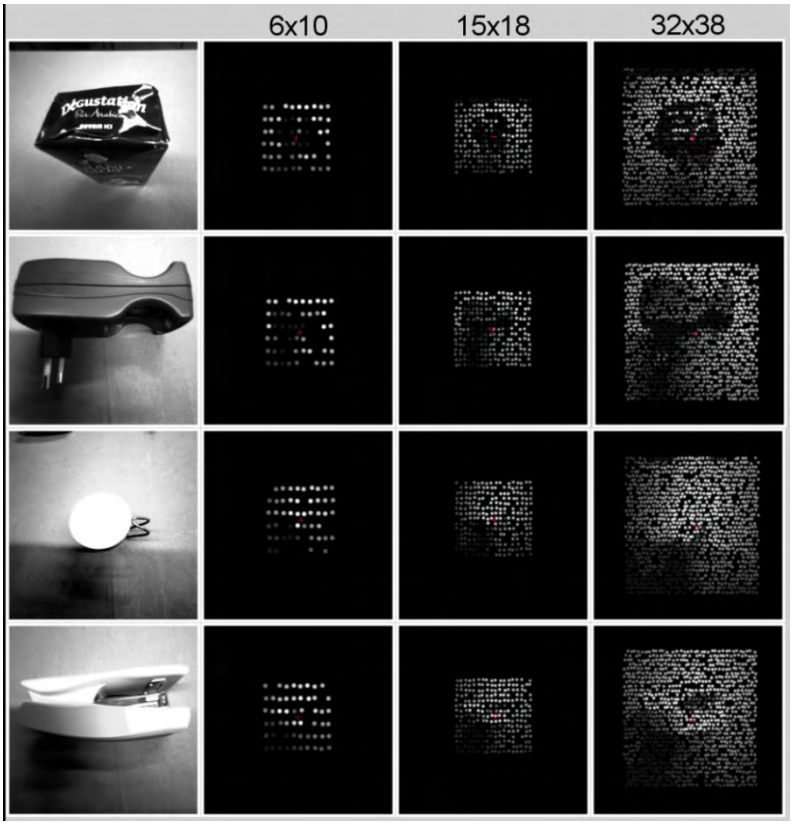


Figure 4-9 Rendu basé sur l'approche scoreboard.
Vision prothétique d'un rendu basé sur l'approche scoreboard de quatre objets avec les trois résolutions de matrices, pour le rendu visuel implémentant l'approche scoreboard.
.

Architecture

Nous avons utilisé quasiment la même architecture que dans l'expérience précédente (Figure 4-10). Le rendu basé sur l'approche scoreboard a été implémentée en C++ à l'aide de la librairie OpenCV. Chaque image capturée par la caméra était convertie en niveau de gris, puis redimensionnée à la résolution de la matrice utilisée (6x10, 15x18 ou 32x38). Cette image réduite était envoyée au simulateur qui transformait et affichait chaque pixel sous la forme d'un phosphène. L'approche par localisation reposait comme dans l'expérience précédente sur le logiciel de reconnaissance de forme SNVision. Nous avons modélisé les 8 objets de telle sorte qu'ils soient reconnus aux différentes positions sur la table. Le nombre de modèles par objet variait entre 21 (téléphone) et 73 (marqueur) : plus l'objet avait de points saillants moins il nécessitait de modèles. Suivant la taille du modèle, les reconnaissances s'effectuaient à une vitesse comprise entre 21ms pour les plus petits et 32ms pour les plus grands.

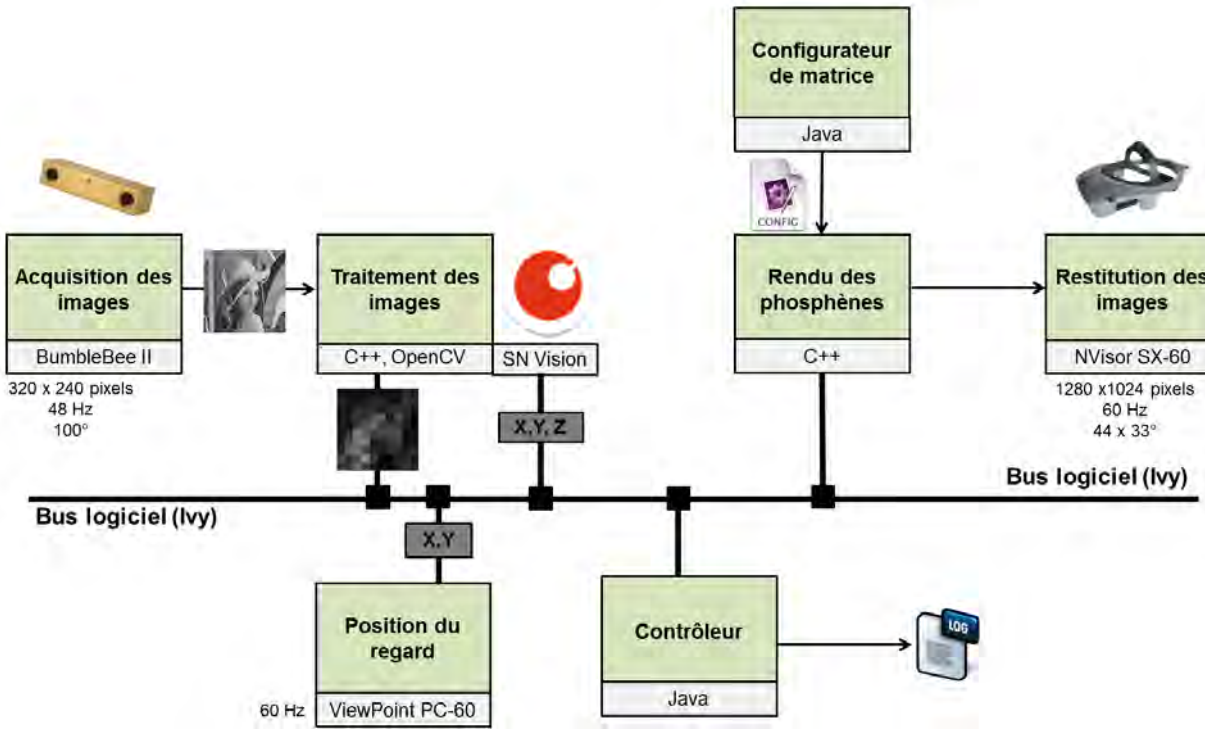


Figure 4-10 Architecture du système.
Une seule différence notable par rapport à l'architecture précédente est à signaler : les coordonnées du regard permettent d'asservir la position de l'image dans le casque. Cette version du simulateur inclut également une implémentation d'un rendu basé sur l'approche scoreboard.

Pour simuler le déplacement des phosphènes lors de mouvements oculaires, nous avons utilisé un oculomètre installé dans le casque de réalité virtuelle, sous l'écran de l'œil droit (ViewPoint PC-60, Arrington Research, AZ, États-Unis). Celui-ci nous permettait de récupérer la position du regard en temps réel (60Hz) et d'asservir la position des phosphènes à la position du regard. Une petite croix de fixation était située au centre de l'écran pour aider les sujets à maintenir le regard droit devant eux. Si le regard déviait de ce point de fixation, tous les phosphènes suivaient le mouvement des yeux.

Un logiciel développé en Java nous a permis de piloter l'expérience et d'enregistrer pour chaque sujet le résultat et le temps de mouvement de tous leurs essais.

Résultats

Les résultats concernent les mêmes mesures que l'expérience 1 à savoir le taux de réussite et le temps pour toucher le bon objet. L'ensemble des statistiques a été effectué avec R (R fondation, États-Unis). Nos données ne suivant pas la loi de distribution normale, nous avons utilisé des tests non paramétriques : Wilcoxon pour la comparaison de deux conditions et l'Anova de Friedman au-delà de deux groupes. Le seuil de significativité a été fixé à 0,05. Dans les paragraphes suivant, SC1 (respectivement SC2 et SC3) correspond au rendu visuel de l'approche scoreboard avec la matrice A1 (respectivement A2 et A3), et LOC fait référence au rendu visuel de l'approche par localisation de points d'intérêt avec la matrice A1.

Si nous prenons en compte tous les sujets, le temps de mouvement moyen pour toucher le bon objet (Figure 4-11 gauche) est de 22,8 secondes avec SC1 (± 4,8), 26,1 s avec SC2 (± 3,6), 22,4 s avec SC3 (± 3,3) et 17,4s avec LOC (± 3,0). Ce temps moyen pour toucher les objets ne présente pas de différences significatives entre les quatre conditions (χ^2=7,3, df=3, p=0,06).

Le taux de réussite avec SC1 (15,6 ±3,9%, Figure 4-11 droite) est légèrement au-dessus du niveau chance qui est à 10% (une chance sur 10 de saisir par hasard le bon objet parmi 10). La précision augmente à 37,5% (± 9,6) avec SC2. Elle atteint plus de 70% correct pour les conditions SC3 et LOC (SC3 : 71,9 ±10,8% ; LOC : 80,9 ±9,4%). Nous observons une différence significative pour cette précision sur

l'ensemble des quatre conditions (χ^2=32,2, df=3, p < 0,001). Les tests post-hoc révèlent une différence significative entre SC1/SC3 (p<0,001), SC1/LOC (p<0,001) et SC2/LOC (p<0,01).

Nous avons également analysé les résultats du groupe novice et du groupe expert pour savoir si le niveau d'expertise à un effet sur la performance. En considérant toutes les conditions scoreboard (SC1, SC2 et SC3), le temps moyen pour toucher l'objet recherché est de 26,7s (± 3,6) pour les novices, contre 20,0s (± 4,9) pour les experts. Avec l'approche par localisation (LOC), ce temps moyen est de 19,5s (± 8,3) pour les débutants et de 16,2s (± 5,9s) pour les experts. Un test de Wilcoxon révèle une différence significative pour le temps de mouvement entre les experts et les novices dans l'approche scoreboard (Z=3,6, p<0,001). Le même test dans l'approche par localisation ne s'avère en revanche pas significatif (Z=1,2, p=0,26).

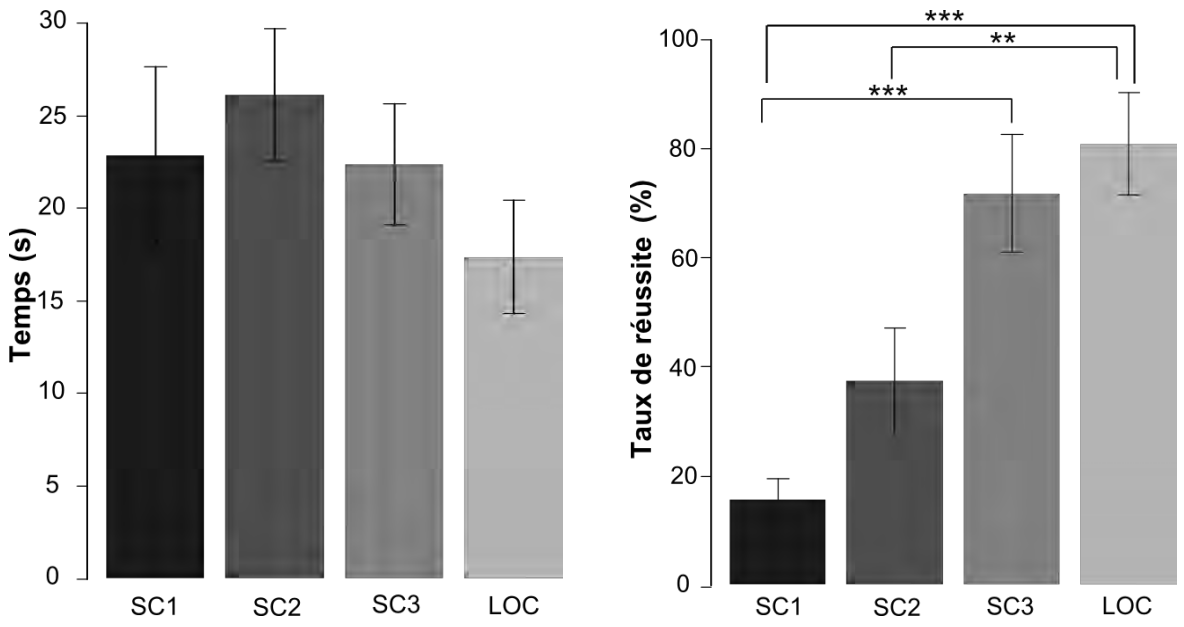


Figure 4-11 Moyenne des temps de mouvement et des taux de réussite.
Moyenne du temps nécessaire pour toucher un objet (gauche) et taux de réussite (droite) sur les 4 conditions de restitution. SC1, SC2 et SC3 : matrice de 6x10, 15x18 et 32x38 en approche scoreboard. LOC : matrice de 6x10 en approche par localisation. **p<0,01, ***p<0,001.

Avec l'approche scoreboard, le taux de réussite pour les trois conditions confondues chez les novices est de 37,0% (± 11,9%), et de 46,3% (±12,5%) chez les experts. Avec l'approche par localisation, il atteint 73,6% (± 22,2) chez les novices et 88,1% (± 15,7%) chez les experts. Comme pour le temps de mouvement, l'expertise a un

effet significatif sur le taux de réussite en scoreboard ($Z=-2,7$, $p<0,05$) et non en localisation ($Z=-2,6$, $p=0,052$).

Si nous regardons de plus près le taux de réussite par objet dans la condition SC1, seul le café (50%) est au-dessus du niveau chance (20%). Avec la condition SC2, le café (80%), le chargeur (58%), la tasse (44%), les ciseaux (38%) et le téléphone (36%) ont un taux de réussite au-dessus du niveau chance. Le taux de réussite pour chaque objet oscille entre 60% (pâte à tartiner) et 100% (café) pour la condition SC3, et il est au-dessus de 75% pour tous les objets dans la condition LOC. Cette dernière est la seule condition pour laquelle les objets ne présentent pas de différences significatives pour le taux de réussite ($\chi^2=8,4$, $df=7$, $p=0,3$).

DISCUSSION GENERALE

Validation ou rejet des hypothèses

Avec un rendu visuel basé sur l'approche scoreboard, les études menées en simulation de vision prothétique, montrent qu'il faut entre 250 et 600 phosphènes distincts pour pouvoir reconnaître des objets. Mais pour l'ensemble de ces travaux, l'environnement de test est extrêmement contrôlé : les objets ou les formes sont noirs ou blancs, relativement grands, et disposés sur un fond très contrasté [Y. Zhao et al. 2010; Pérez Fornos et al. 2008; Hayes et al. 2003]. Notre expérience 2 laisse supposer que plus de 1000 phosphènes sont nécessaires pour atteindre un objet dans des conditions plus naturelles (validation hypothèse n°4). En effet seule la matrice 32x38 (1216 électrodes) permet d'obtenir des taux de réussite satisfaisants (plus de 70% correct).

Nos résultats indiquent que le rendu visuel basé sur l'approche par localisation de points d'intérêt permet d'atteindre de petits objets du quotidien avec une résolution de matrice très limitée : que l'on simule un implant cortical de 9 ou 100 électrodes, ou un implant rétinien de 60 électrodes, les performances sont très satisfaisantes. Plus de 80% de réussite en moyenne, avec un temps moyen pour atteindre les objets compris entre 15 et 20 secondes (validation hypothèse n°1 et n°2). Grâce à cette reconnaissance de haut niveau effectuée dans l'approche par localisation, la performance pour réaliser une tâche de saisie d'objet devient pratiquement

indépendante du nombre de phosphènes, et donc du nombre d'électrodes qu'il est nécessaire d'implanter.

Effet de la position de l'implant

La position des phosphènes dans le champ de vision joue également un rôle mineur dans la performance avec l'approche par localisation. L'expérience 1 nous a permis d'évaluer la performance lorsque l'implant cortical n'est pas positionné sur la zone correspondant à la vision centrale (matrices 9Lateral et 100Lateral). Ceci est réaliste, étant donné qu'il est beaucoup plus simple d'implanter une matrice dans un seul des deux hémisphères du cortex visuel, et ce dans la partie périphérique du champ visuel puisque la partie centrale est difficile d'accès, à l'intérieur de la scissure calcarine (9Central, 100Central). Pour ces deux implants périphériques, les performances pour réaliser la tâche sont inférieures aux performances avec les implants centraux (validation hypothèse n°3), mais elles restent cependant tout à fait acceptables (baisse de 7,1% du taux de réussite et augmentation de 2,7s du temps de mouvement). Ce résultat peut s'expliquer par le fait que l'information à traiter est simple et non ambigüe : l'objet à trouver se traduit par l'apparition d'un unique phosphène. Nous pensons qu'avec un peu plus d'entraînement, les sujets amélioreraient rapidement leur performance. Cette hypothèse pourrait d'ailleurs facilement être testée.

Effet de l'apprentissage

L'approche par localisation de points d'intérêt ne nécessite qu'un faible apprentissage : après deux heures de pratique dans l'expérience 1, le taux de réussite augmente de 5,7% et le temps nécessaire à localiser et toucher un objet diminue de 0,7s (condition 100Central1 vs 100Central2). Bien que cette augmentation soit significative, la performance n'augmente que très légèrement en valeur absolue et se trouve dès le début de l'expérience à un niveau élevé. Plusieurs éléments peuvent expliquer la progression : les sujets apprennent à filtrer les fausses détections en ignorant de manière de plus en plus efficace les phosphènes les moins stables. Ils apprennent également au cours de l'expérience à mieux évaluer la distance qui les sépare des objets à toucher comme l'indiquent certaines réponses au questionnaire post-expérience. Avec l'approche scoreboard utilisée pour les

implants actuels de faible résolution, un apprentissage important est nécessaire pour commencer à appréhender l'environnement [Chen et al. 2005; Hayes et al. 2003; Duret et al. 2006]. L'approche par localisation, en revanche, ne nécessite à priori que très peu de temps de pratique avant d'atteindre la performance maximale.

Vision par ordinateur

Le simulateur de vison prothétique est couplé à un algorithme de reconnaissance de forme temps réel pour implémenter l'approche par localisation. Cet algorithme se concentre sur les caractéristiques saillantes des objets : plus ils sont discriminables, c'est-à-dire d'une forme particulière et fortement texturés, plus ils sont faciles à reconnaître. La carte bleue (expérience 1), par exemple, est plus difficile à identifier par le système, ce qui explique la performance modeste obtenue avec cet objet (en plus de sa petite taille) et le fait qu'il soit considéré par les sujets comme le plus difficile à atteindre. La réussite dans la tâche est donc fortement corrélée à la qualité et à la robustesse de l'algorithme utilisé.

Dans ces deux expériences (et de façon générale dans toutes nos expériences), le but n'était en aucun cas d'évaluer l'efficacité de l'algorithme. Pour les deux expériences, le nombre de fausses détections était lui aussi très faible, car pour chaque image, nous calculions une carte de profondeur et nous filtrions toutes les détections à plus d'un mètre du sujet (distance maximale des objets cibles). Ce même genre de filtre pourrait être implémenté en situation réelle lorsqu'une personne cherche un objet dans son espace péripersonnel. Il est important de noter que les fausses détections apparaissaient moins stables (phosphène moins persistant) que les détections valables. De façon intéressante, nous avons observé que les sujets apprenaient rapidement à ignorer ces fausses détections.

Pour que les performances soient les meilleures possibles, nous avons créé pour les deux expériences des modèles d'objets adaptés à notre environnement expérimental. Pour une application de cette approche à des conditions réelles, il est primordial d'avoir à disposition un nombre important d'objets reconnaissables. Plusieurs solutions sont envisageables pour créer ces modèles. La première solution serait de créer ces modèles de façon supervisée. Dans une maison, on peut se limiter à la modélisation des objets les plus importants de la vie quotidienne. C'est le

parti pris de plusieurs solutions récentes [Schauerte & Martinez 2012; Sudol et al. 2010]. Une approche alternative serait de tagger les objets qui nous intéressent [Jafri, Ali & Arabnia 2013]. Ces deux solutions (modèles supervisés ou tags) nécessitent évidemment l'aide d'une personne voyante pour créer les modèles ou installer les tags. On peut envisager dans un futur proche un service distant permettant d'analyser en temps réel une image sur le même principe que Vizwiz [Bigham et al. 2010], pour fournir en retour les coordonnées de l'objet recherché. Le projet Google Goggles permet déjà une reconnaissance automatique d'objets et de lieux provenant de photographies diverses et variées. Enfin, une solution intéressante pourrait venir des algorithmes capables de construire à la volée des modèles pour apprendre automatiquement de nouveaux objets [Lillywhite et al. 2013].

Scénario d'usage

Dans nos deux expériences, l'utilisabilité de la neuroprothèse est évaluée dans un contexte précis : l'utilisateur cherche un objet en particulier dans son environnement. Dans notre cas, la localisation de l'objet est facile car elle est liée à l'apparition d'un unique phosphène. Cependant dans le cas où le système est capable de reconnaître en parallèle un ensemble d'objets, la vue de l'utilisateur deviendrait vite surchargée et inutilisable si le nombre d'objets détectés était important. Une solution à ce problème serait de pouvoir interagir directement avec la neuroprothèse et lui indiquer l'objet ou les objets recherchés. Le dispositif pourrait, par exemple, être couplé à un logiciel de reconnaissance vocale pour charger à la volée les modèles d'objets que l'utilisateur indiquerait par oral. Cette interaction pourrait être également tactile ou encore gestuelle en fonction du contexte. On peut par exemple s'imaginer qu'en présence d'autres personnes, certains utilisateurs préféraient interagir avec leur smartphone de manière plus discrète qu'à la voix.

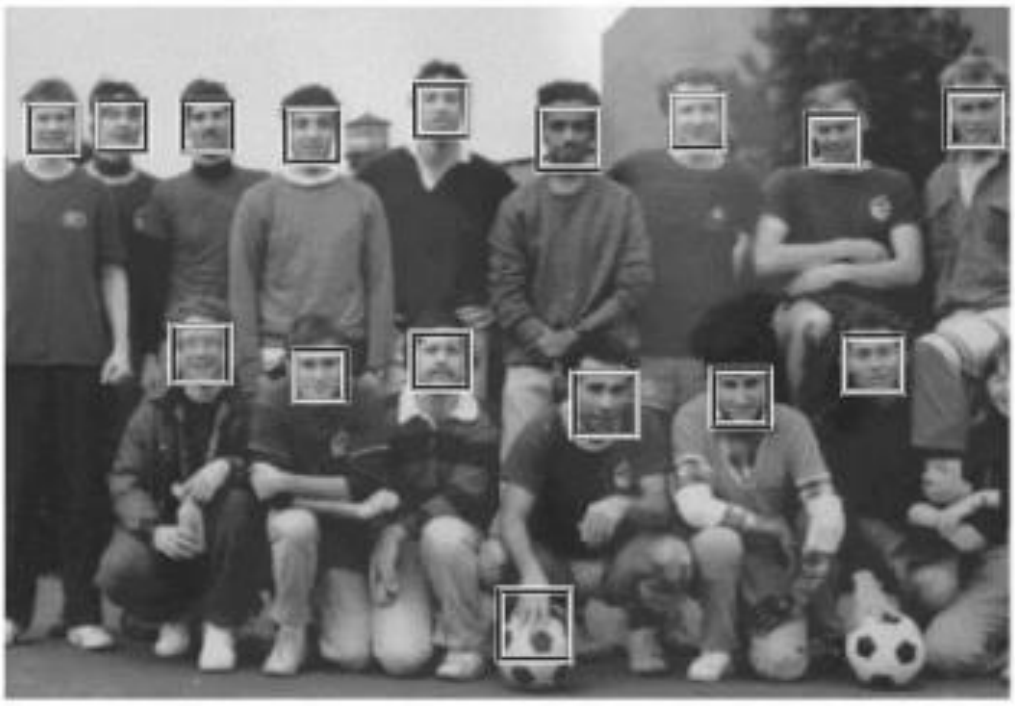
Nous pensons également qu'un système utilisable doit présenter un rendu visuel adapté au contexte de la tâche. Nous suggérons qu'un rendu visuel spécifique est nécessaire pour localiser des objets, ou plus généralement des points d'intérêt, mais qu'un autre rendu serait peut-être plus approprié dans un contexte de navigation par exemple. Ces différents modes de restitution impliquent la possibilité de passer d'un rendu à l'autre et donc d'interagir là aussi avec la neuroprothèse.

Conclusion

Ces deux premières expériences valident l'intérêt d'une approche par localisation, dans un contexte simplifié où l'utilisateur cherche un objet spécifique dans son environnement proche. Nous avons montré que dans cette situation, l'utilisation d'un algorithme de détection d'objets permet de s'appuyer sur un rendu visuel minimaliste qui permet une meilleure utilisabilité des neuroprothèses visuelles. L'approche par localisation de points d'intérêt pourrait être utile dans d'autres contextes. Par exemple, il est certainement intéressant pour les non-voyants de pouvoir localiser les personnes qui les entourent, notamment quand ils arrivent dans de nouveaux bâtiments, ou qu'ils naviguent dans des couloirs. Dans certaines situations, d'autres rendus visuels, toujours basés sur l'utilisation d'algorithmes de vision, pourraient s'avérer plus pertinents qu'un rendu ne comprenant qu'un seul phosphène. Une idée simple serait de fournir la position des points d'intérêt en plus du contexte général restitué par une approche scoreboard. Pour ces différentes raisons, nous proposons au travers de l'expérience suivante de tester une autre forme de rendu basé sur une approche qualifiée de « scoreboard augmenté », dans un contexte de localisation de visages.

CHAPITRE 5

VISION PROTHETIQUE - LOCALISER UN VISAGE



(Source Viola & Jones, 2001)

INTRODUCTION

La détection et l'identification de visages sont des fonctions visuelles complexes que nous utilisons tous les jours. Nous sommes capables de savoir très rapidement si des visages ou des personnes sont dans notre champ de vision et à quel emplacement. Une fois cette détection effectuée, des mécanismes plus fins sont mis en jeu pour procéder à l'identification des visages en mobilisant nos connaissances et nos expériences passées afin de « mettre des noms » sur ces visages. Ces deux fonctions sont essentielles à notre vie sociale et font partie de celles qu'un non-voyant souhaiterait pouvoir à nouveau effectuer avec une neuroprothèse visuelle. Avec les rendus actuels basés sur une approche scoreboard, aucune neuroprothèse visuelle ne permet de détecter et encore moins d'identifier une personne. Des études réalisées en simulation de vision prothétique ont tenté de déterminer les paramètres nécessaires à la réalisation de ces deux tâches.

Pour évaluer la détection de visages avec une vision prothétique simulée, les tâches consistent à indiquer la présence ou l'absence de visage dans une image. Dans leur étude de 2010, Guo et al. étudient l'effet du nombre de phosphènes et de leur organisation spatiale sur les performances de catégorisation d'objets [Guo, Wang, et al. 2010]. Les sujets doivent indiquer la catégorie (visage, voiture, théière et cuillères) à laquelle se rapportent vingt images (4 catégories, 5 exemplaires par catégorie). Les images sont présentées à différentes résolutions (de 8x8, 10x10, 16x16 et 32x32 phosphènes). Trois types de distorsions géométriques sont appliqués au rendu des différentes matrices de phosphènes : une distorsion en barillet, une rotation ou une translation. Dans une configuration sans distorsion, les résultats montrent que 100 phosphènes sont requis pour que la performance atteigne près de 80% de réponses correctes. Ce résultat est à tempérer pour deux raisons. Premièrement seules quatre catégories avec 5 exemplaires par catégorie ont été utilisées dans cette expérience, ce qui facilite grandement la tâche en réduisant les confusions possibles entre catégories. Deuxièmement, les visages présentés aux sujets sont des prises de vues très rapprochées (le visage occupe toute l'image), ce qui ne correspond pas à la majorité des situations réelles, où les personnes qui nous entourent sont à différentes distances et rarement en très gros plan et bien centrées. Une expérience similaire, où les sujets doivent reconnaître des visages parmi d'autres objets

(instruments de musique, fleurs et bâtiments), indique qu'une résolution de 14x16 phosphènes permet de réaliser la tâche efficacement (plus de 80% de réponses correctes) [Guo, Qin, et al. 2010]. Mais là encore les visages sont zoomées au maximum pour occuper complètement le champ de vue couvert par la matrice d'électrodes. À noter que pour ces deux études, les images sont traitées en combinant un seuillage et une extraction de contours, à l'image de ce qui avait été proposé par Buffoni et al. [Buffoni et al. 2005].

L'identification de visage nécessite un nombre encore plus important de phosphènes puisqu'il est nécessaire dans ce cas d'accéder à des détails du visage. Dans l'expérience de Dagnelie et al., les sujets doivent reconnaître un visage (large de 12° d'angle visuel), parmi quatre autres visages tirés au hasard d'un ensemble contenant soixante photographies de visages [Dagnelie et al. 2001]. Les performances sont mesurées en faisant varier le nombre, la taille, la distance bord à bord, les niveaux de luminance des phosphènes, le champ de vision restitué et le pourcentage d'électrodes défaillantes. Les résultats indiquent que l'identification devient possible quand plus de 256 phosphènes sont présents dans un champ de vision d'au moins 7x7° d'angle visuel, que moins de 50% des électrodes sont défaillantes et que la dynamique est d'au moins 4 niveaux de luminance. Ce protocole sera repris deux années plus tard par Thomson et al. [Thompson et al. 2003]. Ils augmentent la taille des phosphènes de 0,38 à 0,53°, et celle du champ de vision restitué (9,6°x9,6°). Comme précédemment, ils arrivent à la conclusion qu'au moins 256 phosphènes sont requis pour réaliser la tâche d'identification de visage. En 2002, Chang et al. comparent différents algorithmes de traitement d'image pour faciliter l'identification de visages familiers [Chang et al. 2012]. Ces algorithmes combinent des filtres d'amélioration de contraste et d'extraction de contours. Dans leur expérience, les phosphènes mesurent 0,6° d'angle visuel et possèdent dix niveaux de luminance. Des matrices de 8x8, 12x12 et 16x16 électrodes sont simulées. Les auteurs montrent que leurs traitements spécifiques permettent d'augmenter significativement la performance pour la résolution 12x12 (plus de 70% de réponse correcte contre 50% sans traitements). Ce résultat est intéressant, mais la vision prothétique simulée dans cette expérience est très « optimiste » : elle ne prend pas en compte les éventuelles défaillances d'électrodes, la carte des phosphènes est parfaitement organisée et le nombre de niveaux de luminance disponible est élevé. De plus,

comme dans la plupart des expérimentations, les visages sont présentés centrés, occupant toute la surface de l'implant.

Toutes les expériences décrites précédemment se concentrent sur une détection ou une identification de visages remplissant la majeure partie de la vue caméra. Mais un nombre bien plus important de phosphènes est nécessaire pour localiser et identifier les visages de personnes avec une taille apparente plus réduite dans le champ de vision. Pour remédier à ce problème, deux équipes se sont intéressées très récemment à l'utilisation d'algorithmes de détection et de suivi de visages en simulation de vision prothétique [He et al. 2012; Wang, Wu, et al. 2014]. Dans ces deux travaux, lorsqu'un visage est détecté, il est automatiquement zoomé pour remplir la majeure partie du champ visuel afin de permettre au sujet d'éventuellement l'identifier. Avec cette approche, l'identification de personnes se tenant jusqu'à cinq mètres est possible, mais elle repose tout de même sur une résolution de matrice élevée (entre 500 et 1000 phosphènes).

Une nouvelle approche : l'approche scoreboard augmenté

Toutes ces données indiquent que la détection et l'identification de visages sont réalisables en vision prothétique, quand plusieurs centaines de phosphènes distincts sont disponibles. Les implants actuels ont une résolution inférieure à cent électrodes, et il faudra de nombreuses années avant de disposer d'implants permettant de percevoir des centaines ou des milliers de phosphènes. Dans ce chapitre, nous proposons d'introduire et d'évaluer un nouveau rendu adapté aux implants de faible résolution. Nous pensons que le premier rendu que nous avons évalué dans les deux premières expériences n'est pas forcément pertinent en situation de mobilité [Vergnieux et al. 2012]. En effet, en choisissant de restituer à l'utilisateur uniquement la position des objets d'intérêts, nous lui supprimons le contexte de la scène auquel il aurait eu accès par un rendu scoreboard. Pour cette raison, nous proposons une nouvelle approche, **l'approche scoreboard augmenté**, qui combine le scoreboard et la localisation de points d'intérêt. L'idée est de conserver le maximum d'informations que peut fournir l'approche scoreboard, tout en augmentant ce rendu visuel grâce une information localisée et de haut niveau qui soit pertinente pour l'utilisateur (dans cet exemple, les visages). La localisation et la distance des visages sont extraites des images en temps réel, puis restituées sous forme de phosphènes

blancs clignotants (la fréquence du clignotement étant fonction de la distance des visages). Bien que le rendu choisi ne permette pas d'identifier directement des visages (il le pourrait en s'appuyant, par exemple, sur une synthèse vocale qui indique le nom de la personne détectée), nous montrons que cette approche est potentiellement utilisable dès à présent pour les personnes actuellement équipées d'implants.

EXPERIENCE

L'objectif premier de ce travail[16] est d'évaluer l'approche scoreboard augmenté, en simulant un implant dont les spécifications sont conformes à celles des implants épirétiniens actuels. Pour cela, nous reprenons à l'identique les caractéristiques de l'implant basse résolution utilisé dans notre seconde expérience (6x10 électrodes). Avec cette résolution, en s'appuyant sur les résultats présentés dans l'introduction, nous affirmons qu'une approche scoreboard seule ne permet pas de localiser des visages dans la plupart des situations quotidiennes. Nous comparons également les résultats obtenus avec l'approche scoreboard augmenté avec ceux réalisés en approche par localisation de points d'intérêt pour ce même implant, afin de mesurer l'impact de notre nouvelle approche sur la performance des sujets.

Enfin, pour les deux rendus (scoreboard augmenté et localisation par points d'intérêt), nous étudions la possibilité de fournir à l'utilisateur la distance des visages en s'appuyant sur la fréquence de clignotement des phosphènes. En différenciant deux fréquences de clignotement (plus rapide si plus proche), le sujet peut déterminer la personne la plus proche de lui. À noter que cette idée, lors de la mise en place de l'expérience, était plus exploratoire : elle supposait une grande maîtrise des fréquences de stimulation, ce qui a été fortement mis en doute par des travaux récents parus après cette expérience [Pérez Fornos et al. 2012]. Une telle approche serait cependant possible avec une partie des sujets implantés et cette variation de clignotement pourraient aussi être représentée avec d'autres dimensions des phosphènes, comme leur luminance.

[16] Ces travaux ont été publiés dans [Denis et al. 2013].

Hypothèses de travail

Dans cette expérience, nous posons deux hypothèses :

Hypothèse n°1 : avec un implant basse résolution, l'approche scoreboard augmenté permet de localiser efficacement un ou plusieurs visages dans son environnement proche.

Hypothèse n°2 : cette performance n'a pas ou peu d'impact comparée à celle obtenue avec un rendu basé sur l'approche par localisation de points d'intérêt.

Protocole expérimental

Quatre hommes âgés de 23 à 27 ans ont participé à cette expérience. Tous ont déjà utilisés un simulateur de vision prothétique.

Au début de l'expérience, les sujets entraient dans la pièce avec les yeux bandés pour ne pas percevoir d'éléments du dispositif qui pourraient éventuellement les aider dans leurs réponses. Ils étaient ensuite équipés d'un casque de réalité virtuelle, et de petits marqueurs permettant de suivre la position de leurs mains et de leurs épaules, grâce à un système de capture de mouvement. Puis, l'expérimentateur positionnait les sujets à un endroit précis dans la salle. Leur tâche consistait à localiser les visages face à eux, sachant qu'à chaque essai pouvait se trouver devant eux zéro, un ou deux visages. Avant chaque essai, entre zéro et deux personnes se positionnaient, face au sujet, à des emplacements préalablement définis comme indiqué sur la Figure 5-1.

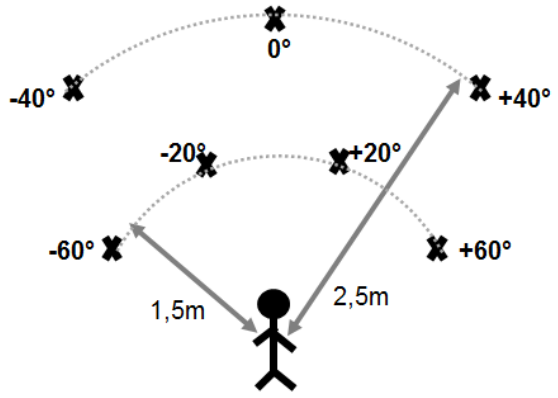


Figure 5-1 Disposition de l'environnement.
Le sujet devait indiquer le nombre de visage(s) se trouvant face à lui. À chaque essai, 0, 1 ou 2 personnes se plaçaient à l'une des sept positions prédéfinies.

Chaque sujet réalisait cette tâche avec un implant simulé de 6x10 électrodes, dans deux conditions de rendu différentes. Le premier rendu était basé sur l'approche par localisation de points d'intérêt présentée en fin de chapitre 2: le ou les phosphènes affichés indiquaient la position du ou des visages dans le champ de vision du sujet. Les phosphènes étaient blancs et clignotaient à une fréquence dépendant de la distance des visages. Plus celle-ci était rapide, plus la personne était proche. Le second rendu combinait le scoreboard et la localisation de points d'intérêt : les informations sur les visages localisés venaient augmenter la vue scoreboard classique par le biais de phosphènes blancs clignotants. Cette approche est désignée sous le terme de **scoreboard augmenté** (Figure 5-2). Les sujets réalisaient 30 essais par condition (une condition correspondait à un rendu). L'ordre de passage de ces deux conditions était contrebalancé entre sujets pour limiter d'éventuels effets d'apprentissage.

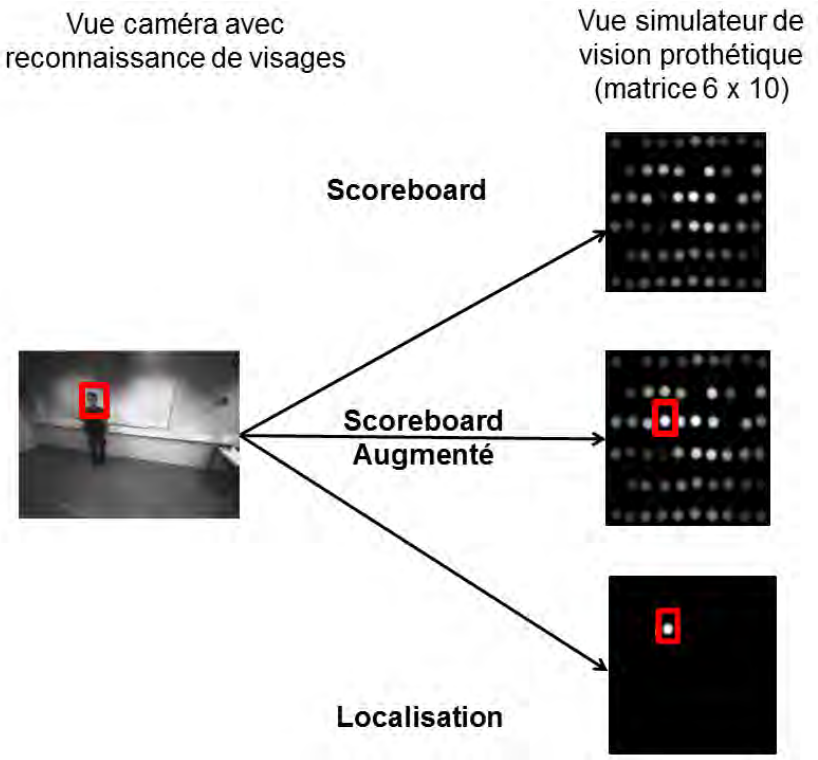


Figure 5-2 Scoreboard vs Scoreboard Augmenté vs Localisation.
Dans l'approche scoreboard augmenté, lorsqu'un visage est détecté et localisé (image de gauche), un phosphène blanc clignotant indique sa position dans le champ de vision du sujet, en surimpression de la vue scoreboard. Dans cette expérience les deux conditions utilisées sont la localisation et le scoreboard augmenté.

À chaque essai, le sujet parcourait l'environnement en tournant la tête pour localiser les visages éventuellement présents et indiquer leur nombre (aucun, un ou deux). Pour ce faire, il pointait du doigt le ou les visages qu'il avait localisé, ou bien ne bougeait pas s'il n'en avait localisé aucun. Lorsqu'il pensait avoir localisé deux visages, il indiquait dans un second temps quel était celui qui se trouvait le plus proche de lui. L'expérimentateur attendait la confirmation du sujet pour enregistrer sa réponse. Lorsque l'essai était terminé, la simulation était éteinte, et la configuration suivante était mise en place. Après 30 essais, la seconde condition de rendu était chargée et le même protocole était exécuté à nouveau.

Simulateur

Implant simulé

Pour ces deux conditions, nous avons simulé un seul implant rétinien basse résolution dont les caractéristiques sont identiques à celui simulé dans la deuxième expérience. Celui-ci était composé d'une matrice de 6x10 électrodes disposées de façon rectangulaire. Sur ces 60 électrodes, 10% (sélectionnées aléatoirement pour chaque sujet) étaient définies comme défaillantes et restaient éteintes pendant l'expérience. Le champ de vision restitué correspondait à une fenêtre de 11x11° d'angle visuel. Les phosphènes générés avaient une taille de 1°, étaient séparés de 1,2° bord à bord et pouvaient afficher huit niveaux de luminance.

Architecture

La Figure 5-3 présente l'architecture du simulateur utilisé pour cette expérience. Comme pour l'expérience de localisation d'objets, l'acquisition des images s'est faite par la caméra stéréoscopique BumbleBee II et la restitution dans le casque de réalité virtuelle NVisor SX-60. Dans cette expérience, la position des phosphènes n'a pas été asservie à la position des yeux et l'oculomètre n'a pas été utilisé. L'asservissement dynamique aurait été incompatible avec une station debout du sujet, utilisée pour pointer librement vers les visages cibles. La détection des visages a été effectuée par l'algorithme de reconnaissance de forme SNVision (SpikeNet Technology), déjà introduit dans l'expérience de localisation d'objet. Ici aussi son utilisation s'avérait pertinente car l'outil est capable de détecter très efficacement des visages (grande sensibilité pratiquement sans fausse détection pendant toute

l'expérience). À nouveau, notre objectif n'était en aucun cas de mesurer la performance des algorithmes de vision utilisé. La fiabilité des algorithmes de détection de visages n'est plus à mettre en doute, et d'autres solutions tout aussi pertinentes auraient pu être utilisées, comme l'utilisation des algorithmes implémentés dans la librairie open source OpenCV.

Un système de capture de mouvement à 12 caméras (OptiTrack, Natural Point, USA), récupérait la position des épaules et des mains du sujet. Ces informations étaient enregistrées en continu et permettaient de déduire (1) le nombre de mains que le sujet avait levé, et (2) la direction pointée, pour calculer la précision du pointage en mesurant l'angle (en degrés) entre la position correcte du visage et la position pointée par le sujet.

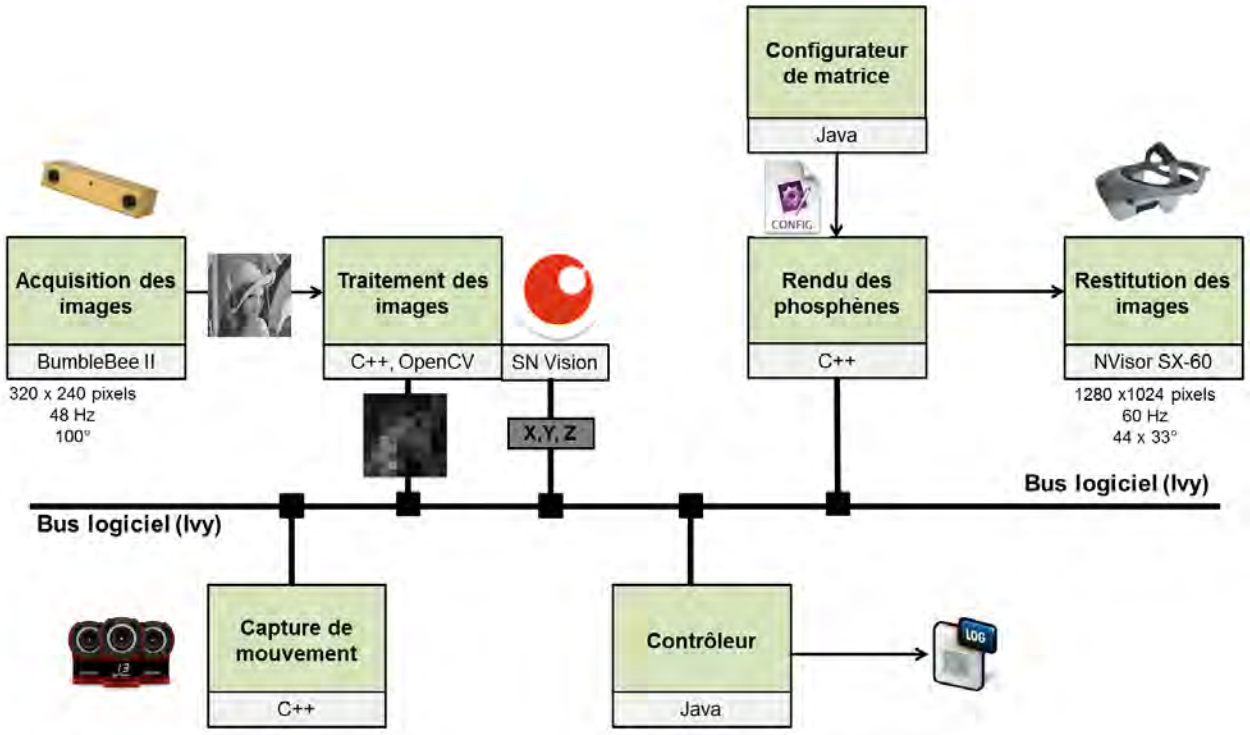


Figure 5-3 Simulateur utilisé dans le cadre de l'expérience de localisation de visages.
La caméra stéréoscopique et la librairie SN Vision permet au module de traitement des images d'envoyer sur le bus Ivy, en temps réel, les positions x,y,z des visages reconnus. Les mouvements du sujet sont suivis par un système de capture de mouvement contenant 12 caméras.

L'expérience était contrôlée à l'aide d'un logiciel développé en Java qui récupérait et enregistrait les réponses des sujets. Un premier ordinateur (Intel Core i7, Windows 7 64 bits) exécutait le module de traitement d'images et le logiciel de contrôle. Un second ordinateur (Intel Core i7, Windows 7 64 bits) était en charge du rendu des phosphènes et de la capture de mouvements. Les différents modules communiquaient par messages asynchrones envoyés sur le bus logiciel Ivy.

Résultats

Quatre paramètres ont été analysés : le taux de réussite (pourcentage de réponses correctes), le temps de réponse (temps en secondes des réponses correctes), la précision de pointage (différence en degré entre le pointage effectué par le sujet et la position réelle du ou des visages) et la précision de distance (pourcentage de jugement de distance correct). L'analyse des données et les tests statistiques ont été réalisés avec R (R Foundation, USA). Les jeux de données étant très limités, et nos données ne suivant pas la loi de distribution normale, nous avons réalisés des tests non-paramétriques. Les comparaisons entre deux groupes ou conditions sont évaluées par un test de Wilcoxon. Au-delà de deux conditions, nous utilisons les Anova de Friedman. Pour l'ensemble de nos tests, le seuil de significativité était fixé à 0,05. Dans la suite de ce paragraphe, les rendus basés sur la localisation de points d'intérêt et sur le scoreboard augmenté seront respectivement nommés LOC et SCB+LOC.

Dans la condition SCB+LOC, tous les sujets étaient capables de compter les visages : le taux de réussite individuel était toujours supérieur à 85% de réponses correctes. Tous sujets confondus, le taux de réussite moyen était de 95,1% (± 8,3) et le temps de réponse de 19,0s (± 4,7). La Figure 5-4 illustre leur performance.

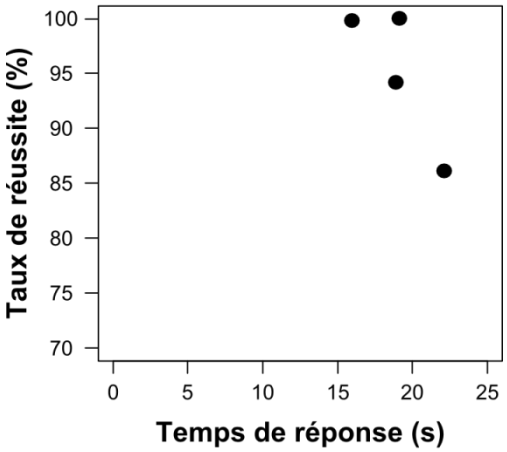


Figure 5-4 Taux de réussite en fonction du temps de réponse pour la condition scoreboard augmenté pour 4 sujets.

La Figure 5-5 montre le taux de réussite (à gauche) et le temps de réponse (à droite), par nombre de visages, pour la condition SCB+LOC. Quel que soit le nombre de visages à indiquer (aucun, un ou deux), le taux de réussite moyen était au-dessus de 90% en SCB+LOC (niveau chance : 33%). En revanche, plus le nombre de visages à trouver était important, plus le temps de réponse s'allongeait (de 16,2s (± 4,6) pour aucun visage, à 22,2s (± 5,0) pour deux visages). Lorsque les sujets ne détectaient aucun visage, ils n'avaient pas besoin de s'appliquer à pointer les cibles, et il était donc logique qu'ils soient plus rapides à indiquer leur décision. Bien qu'il y ait une différence notable, le nombre de visage à trouver n'avait pas d'effet significatif sur le temps de réponse (χ^2=4,5, df=2, p = 0,11). L'effet aurait pu se révéler significatif avec un échantillon plus important.

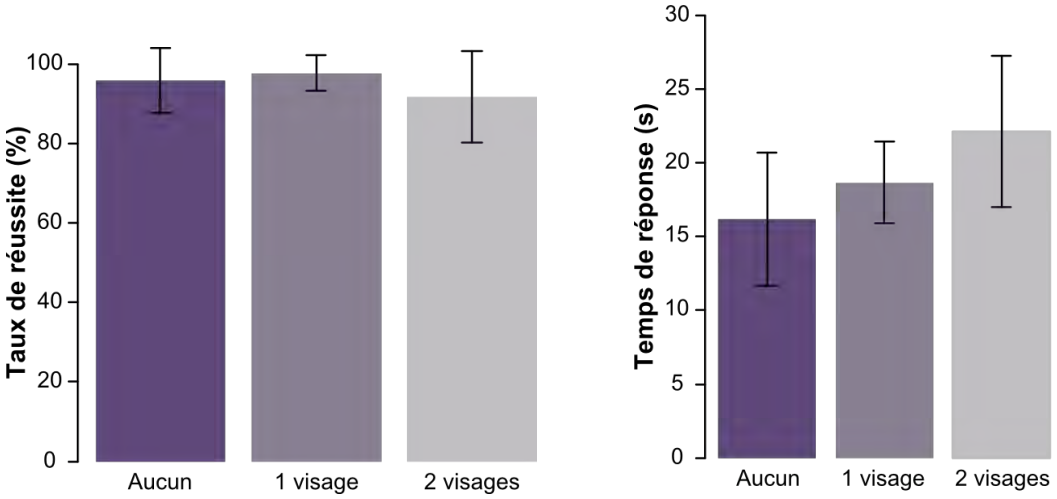


Figure 5-5 Taux de réussite et temps de réponse.
À gauche, le taux de réussite par nombre de visages avec le rendu scoreboard augmenté. À droite, le temps de réponse par nombre de visages pour le même rendu.

Dans la condition SCB+LOC, les sujets pointaient vers les visages avec une erreur moyenne de 13,1° (± 3,9). Cette erreur était sensiblement identique, que les sujets pointent vers un ou deux visages à la fois (elle est respectivement de 13,8° (± 2,7) et 12,2° (± 4,9), voir Figure 5-6).

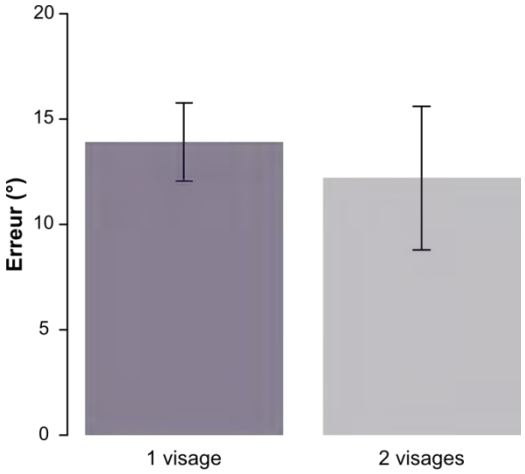


Figure 5-6 Précision de pointage.
Précision de pointage en fonction du nombre de visages à trouver pour la condition scoreboard augmenté (SCB+LOC).

La Figure 5-7 présente le taux de réussite (à gauche) et la précision de pointage (à droite) obtenus en LOC et en SCB-LOC. Pour ces deux variables, les résultats obtenus dans la condition LOC n'étaient pas significativement différents de ceux observés en SCB+LOC (taux de réussite : $Z=-1,6$, $p=0,56$, précision de pointage : $Z=-0,56$, $p=0,60$). En revanche, les sujets étaient plus rapides dans leur choix avec la condition LOC ($Z=-3$, $p<0,001$) : ils mettaient en moyenne 11,4s (± 3,3) à donner la bonne réponse dans cette condition contre 19,0s (± 4,7) en condition SCB+LOC.

Lorsque deux visages étaient présents, les sujets étaient également capables d'indiquer celui qui est le plus proche, quel que soit le rendu utilisé. La précision pour juger de la distance était de 81,2% (± 14,2) en LOC, et de 77,0% (± 8,0) en SCB+LOC. Aucun effet significatif n'a été observé entre les deux conditions pour cette mesure de précision ($Z=-0,73$, $p=0,58$).

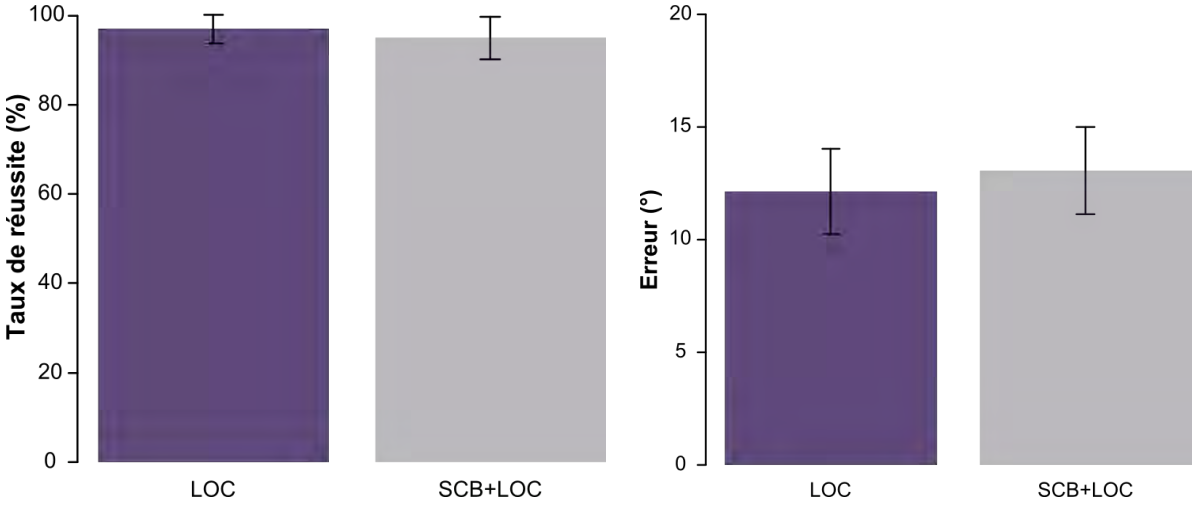


Figure 5-7 Taux de réussite et précision.
À gauche, le taux de réussite pour les conditions LOC et SCB+LOC. À droite, la précision de pointage pour ces mêmes conditions.

DISCUSSION

Validation ou rejet des hypothèses

Dans cette expérience, nous évaluons un rendu basé sur une nouvelle approche : l'approche scoreboard augmenté, qui consiste à surimposer au rendu scoreboard « classique » (pixellisation de l'image par redimensionnement à la taille de la matrice simulée) des informations de localisation de points d'intérêt. Contrairement à l'approche par localisation de points d'intérêt seule, le scoreboard augmenté fournit un contexte sur l'environnement, tout en donnant aux sujets les moyens de situer des objets dans leurs environnements. Comme lors de notre première expérience, les informations de localisation d'éléments de haut niveau (ici les visages) sont extraites à l'aide d'un algorithme de vision par ordinateur. Nous simulons dans cette expérience la perception restaurée par une neuroprothèse visuelle épirétinienne et montrons que malgré la faible résolution de l'implant, cette approche permet de localiser des visages se trouvant aux alentours du sujet (validation de l'hypothèse n°1).

Nous comparons dans cette expérience le scoreboard augmenté à l'approche par localisation de points d'intérêt. Nous observons que la performance est sensiblement identique dans les deux approches, mais que les temps de réponse sont plus rapides pour cette condition : l'absence du « fond » scoreboard simplifie l'information à traiter par le sujet (rejet de l'hypothèse n°2).

Les différents rendus visuels

L'approche par localisation de points d'intérêt peut s'avérer utile lorsque le sujet souhaite détecter rapidement la position d'objets spécifiques, par exemple des visages, lorsqu'il entre dans une pièce. En revanche, dans des situations de navigation, nous suggérons que le rendu scoreboard augmenté soit plus pertinent, car il restitue l'environnement dans sa globalité, tout en indiquant la position des objets d'intérêt. Ainsi, lorsqu'une personne se déplace dans un couloir, le scoreboard augmenté renvoie une information -même peu précise- sur son environnement, à laquelle peut être ajoutée une information sur la position des personnes se trouvant sur son chemin. De façon plus générale, nous soulevons ici l'intérêt majeur d'imaginer et de développer des rendus adaptés à des contextes et à des tâches spécifiques. Nous suggérons que les systèmes de demain intègrent différents rendus, et permettent aux utilisateurs de passer facilement de l'un à l'autre, à leur convenance. Cette idée sera développée plus en détail dans la dernière partie de la thèse lorsque j'aborderai les perspectives de ce travail.

Vision par ordinateur

Avec les neuroprothèses actuelles, il n'est pas envisageable d'identifier un visage : en effet, les études en simulation de vision prothétique montrent que plusieurs centaines de phosphènes distincts sont nécessaires pour effectuer cette tâche. Or, les solutions existantes en génèrent tout au plus quelques dizaines. Avec un rendu scoreboard, nous affirmons qu'il n'est pas non plus possible aujourd'hui de détecter et localiser un visage se trouvant à plus d'un mètre dans la plupart des situations quotidiennes. Cette tâche devient en revanche réalisable, dès l'instant où la position des visages est indiquée aux utilisateurs. C'est ce que nous avons proposé de mettre en place au travers du rendu scoreboard augmenté reposant sur l'utilisation d'un algorithme de détection rapide de visages.

La détection (et à fortiori la localisation) de visages est l'un des domaines les plus actifs en vision par ordinateur. Depuis plusieurs années maintenant, cette tâche est réalisable par un ordinateur en temps réel [Viola & Jones 2001]. Aujourd'hui, les approches les plus performantes utilisent une méthodologie basée sur l'apparence et la forme particulière des visages. Les algorithmes sont rapides et très performants et s'intègrent facilement dans des appareils mobiles (il suffit de se référer à la robustesse de ceux embarqués dans les appareils photos numériques). Dans cette expérience nous avons fait le choix d'utiliser SpikeNet pour effectuer la détection de visages, mais de nombreux algorithmes dédiés à la détection de visages auraient pu être utilisés. Il existe par exemple une implémentation basée sur l'algorithme de Viola & Jones disponible dans la librairie OpenCV.

Les derniers algorithmes de détection de visages sont très performants. Ils ont cependant quelques limites : pour être reconnus, les visages doivent être le plus possible face à la caméra (les algorithmes s'appuient en majorité sur la reconnaissance des yeux et de la bouche), et à une distance raisonnable de celle-ci. Cependant, nous pensons que cette contrainte n'est pas gênante dans les scénarios d'usage car il n'est pas forcément utile pour un utilisateur de localiser un visage très éloigné.

Scénario d'usage

L'utilisation de ce rendu visuel spécifique, dédié à la localisation de visages, trouve son intérêt dans des environnements où le nombre de personnes n'est pas trop important. En effet, dans une rue ou une manifestation publique quelconque, ce rendu risquerait d'être inutilisable en raison du trop grand nombre de visages détectés. Une solution à ce problème serait de limiter la détection à un ou deux mètres, ou de restreindre le nombre de visage détectés (par exemple restituer la position des trois ou quatre visages les plus proches). En revanche, ce rendu pourrait être utile à l'utilisateur pour localiser les personnes qui l'entourent dans certaines situations quotidiennes : l'arrivée dans un nouveau bâtiment ou une salle de réunion, la navigation dans un couloir…

Un autre scénario plus ambitieux consisterait à obtenir, en plus de sa position, l'identité d'une personne. Ceci pourrait être rendu possible par l'utilisation

d'algorithmes temps réel d'identification de visages. L'utilisateur « sélectionnerait » un visage reconnu, et le système lui renverrait l'identité de la personne par exemple par retour audio. Aujourd'hui, le logiciel DeepFace de Facebook combine la modélisation 3D et des réseaux de neurones artificiels [Taigman et al. 2014] pour établir une correspondance d'identité entre deux images avec une précision proche de celle obtenue par des humains (taux de réussite de plus de 97%). Bien que cette précision soit impressionnante, il n'est pas envisageable à court terme que ce traitement soit temps réel du fait de sa complexité algorithmique.

Conclusion

Dans cette expérience, nous avons montré que l'utilisation d'un algorithme de détection de visages permettait de construire un rendu visuel autorisant un utilisateur à localiser les personnes dans son environnement proche, et ce malgré la résolution très faible de l'implant simulé (60 électrodes). En ce sens, l'utilisation de cette catégorie d'algorithmes de vision améliore l'utilisabilité des neuroprothèses visuelles. Le rendu proposé est différent de celui évalué lors de nos deux premières expériences : il ne se limite plus aux informations de localisation, mais combine celles-ci avec les informations de contexte restaurées par une approche scoreboard. Nous ne pouvons pas affirmer que ce rendu scoreboard augmenté est plus pertinent, mais nous pouvons supposer qu'en situation de mobilité, l'accès au contexte soit important pour l'utilisateur.

Si nous tenons compte des résultats de nos trois premières expériences, nous voyons que l'utilisation d'algorithmes de vision par ordinateur peut avoir un réel impact sur la faisabilité de tâches quotidiennes, et que suivant le contexte de la tâche (et l'utilisateur très certainement), différents rendus visuels peuvent être fonctionnels.

Une autre problématique essentielle pour les non-voyants est l'accès à l'information écrite. C'est pourquoi dans notre dernière expérience, nous nous sommes intéressés à l'apport d'un algorithme de détection de texte dans la conception de rendus visuels favorisant l'accès à l'information écrite pour les personnes équipées de neuroprothèses visuelles.

CHAPITRE 6

VISION PROTHETIQUE - LOCALISER UN TEXTE

INTRODUCTION

Dans notre quotidien, l'information écrite est omniprésente. Celle-ci a pour rôle de nous renseigner, nous guider, nous avertir ; au sens large, transmettre de l'information. Dans une ville, on retrouve cette information écrite sous différentes formes : des panneaux signalétiques, des panneaux d'information, des enseignes de magasin, des noms des rues, des numéros des bus, etc. Dans des lieux comme des aéroports, des gares ou des stations de métros, les panneaux nous permettent d'obtenir rapidement l'information nécessaire pour naviguer vers différents lieux (un quai de gare, un terminal d'embarquement, des toilettes, etc.). Dans des bâtiments, l'information écrite peut nous aider à trouver les ascenseurs, les escaliers, un service ou un bureau particulier… Toutes ces informations écrites à des emplacements très variés et sur des supports très divers sont pour leur grande majorité totalement inaccessibles pour les non-voyants.

Parmi l'ensemble des projets de développement de neuroprothèses visuelles, seuls deux projets de rétines artificielles ont rapporté chez quelques sujets des améliorations fonctionnelles plus ou moins importantes dans des tâches de lecture [da Cruz et al. 2013; Zrenner et al. 2011]. Dans les deux cas, les lettres ou mots reconnus par les patients sont larges (entre 2° et 8° d'angle visuel), et fortement contrastés (caractères blancs sur fond noir). Les temps de lecture sont très longs (parfois plusieurs dizaines de secondes pour une seule lettre).

La lecture est la tâche qui a été la plus étudiée en simulation de neuroprothèses visuelles [Barry & Dagnelie 2011]. Les résultats indiquent que cette tâche est fortement dépendante du nombre de phosphènes simulés : un minimum de 3 à 5 phosphènes par caractère est nécessaire pour reconnaître un mot. Les premiers travaux ont été rapportés par Cha et al. au début des années 90 [Kichul Cha et al. 1992]. Dans cette expérience, les sujets doivent lire des blocs de texte avec une vision prothétique simulée qui est restituée dans un champ visuel de 1,7°. Le nombre de phosphènes varie entre 100 et 1024, et les sujets exécutent la tâche dans deux conditions expérimentales différentes. Dans la première condition, les blocs de texte défilent automatiquement ; dans la seconde, il est nécessaire de les scanner par des mouvements de tête. Lorsque le défilement est automatique, et que la matrice contient 625 phosphènes ou plus, les sujets lisent correctement un texte dont les

caractères mesurent 0,4° d'angle visuel à un rythme confortable de 200 mots par minute. Un nombre plus restreint de phosphènes diminue la vitesse de lecture. Avec des résolutions spatiales optimales (625 et 1024 phosphènes), lorsque ce sont les sujets qui scannent le texte par des mouvements de tête, le rythme de lecture est sensiblement plus faible (120 mots par minutes). Tout comme leur prédécesseur, Dagnelie et al. confirment en 2001 que les performances de lecture sont bien sûr dépendantes du nombre de phosphènes [Dagnelie et al. 2001]. Deux conclusions supplémentaires sont présentées dans leurs travaux : (1) il est encore possible de lire 50 mots par minute, même si l'on simule 30% d'électrodes défaillantes dans une matrice en contenant 625 ; (2) le nombre de niveaux de luminance n'a pas d'impact sur la performance. En 2003 puis 2004, Sommerhalder et al. ont également conduit des expérimentations sur la lecture [Sommerhalder et al. 2003; Sommerhalder et al. 2004]. Contrairement aux travaux précédents, et pour être plus proche de la réalité, les sujets ne peuvent pas scanner le texte avec leurs yeux. Dans la première étude, les personnes doivent lire des mots de quatre lettres d'une taille d'environ 2 à 3° d'angle visuel. Les mots apparaissent à différentes excentricités, dans un champ de vision couvrant 20 x 7° ou 10 x 3,5° d'angle visuel. Quelle que soit la taille de ce champ de vision, lorsque les mots sont situés en vision centrale, environ 300 phosphènes (ici des pixels) sont requis pour lire presque parfaitement (plus de 90% de réponses correctes). En revanche, un nombre beaucoup plus conséquent de phosphènes est nécessaire pour effectuer la tâche dès lors que le mot apparaît à une excentricité supérieure à 10 degrés. Dans cette condition, un entraînement important peut améliorer la performance des sujets. Dans la seconde étude, Sommerhalder et al. s'intéressent à la vitesse de lecture. 572 phosphènes (pixels) sont simulés dans un champ de vision couvrant 10 x 7° d'angle visuel. La vitesse atteint 65 mots par minute lorsque les mots apparaissent en vision centrale, mais tombe à 3 mots par minute si ceux-ci sont situés à une excentricité de 15°. Ici encore, un entraînement conséquent (60 sessions sur deux mois) permet d'améliorer cette performance (jusqu'à 23 mots par minute en moyenne). En 2006, Dagnelie et al. montrent que chaque lettre doit être représentée par au moins trois phosphènes pour que les sujets soient capables de reconnaître aisément des mots (90% de réussite) [Dagnelie, Barnett, et al. 2006]. Pérez Fornos et al. ont rapporté plus récemment des conclusions moins optimistes que l'étude précédente [Pérez Fornos et al. 2011]. Les performances maximales sont obtenues pour des résolutions

spatiales comprises entre 3,6 et 4,5 phosphènes (pixels) par caractère. Dans des conditions optimales (4,5 phosphènes par caractère), et avec 60 phosphènes, la vitesse de lecture atteint 34 mots par minute. Des résultats sensiblement identiques ont été publiés pour la lecture de caractères asiatiques [Chai et al. 2007; Zhao et al. 2011].

L'ensemble de ces résultats indique que les neuroprothèses visuelles actuelles permettent de lire, certes très lentement, de gros caractères présentés dans des conditions idéales. La résolution des matrices à venir (200-300 électrodes) permettra d'augmenter sensiblement la vitesse de lecture, mais toujours pour des lettres de taille importante et fortement contrastées. Cependant, dans les situations de déplacement quotidiens les plus communes (rue, bâtiment …), l'information écrite se trouve le plus souvent à une distance importante, et les caractères ont de ce fait une taille apparente bien plus faible que toutes les expériences citées précédemment. De plus, dans la majorité des cas, le contraste entre le fond et les lettres varie énormément entre les textes et entre les conditions d'illumination. Dans ces situations, nul doute que les personnes équipées d'implants de la génération actuelle et de la suivante (50 à 300 électrodes), ne sont et ne seront pas capables de localiser (et encore moins de lire) les textes présents autour d'eux. Pour ces personnes, il n'est pas envisageable d'avoir à fouiller tout l'environnement à une distance de 50 cm pour trouver cette information textuelle.

Pour pallier le problème de l'accessibilité des textes en condition de déplacement, nous proposons dans cette expérience d'évaluer un rendu visuel du texte adapté à ces situations, en permettant aux personnes implantées de localiser rapidement les blocs de texte se trouvant autour d'eux. Ce rendu est basé sur l'approche scoreboard augmenté introduite dans l'expérience précédente : à l'aide d'un algorithme de reconnaissance de texte, nous mettons en surbrillance (des phosphènes blancs) les zones de l'image captée par la caméra contenant du texte, tout en conservant la structure spatiale de la scène visuelle (scoreboard).

Expérience

Dans cette expérience[17], après la localisation d'objets et de visages, nous nous plaçons dans une troisième situation dans laquelle la résolution des implants actuels est un frein à l'utilisabilité des neuroprothèses visuelles : la localisation de texte dans une scène naturelle. Les résultats exposés dans l'introduction laissent supposer que cette tâche est infaisable avec seulement quelques dizaines de phosphènes, dès lors que le texte est éloigné. Nous proposons l'utilisation d'un algorithme de détection de texte pour construire un rendu visuel scoreboard augmenté (présenté au chapitre précédent) adapté à la réalisation de cette tâche. Bien que construit à partir de la même approche, ce rendu diffère de celui évalué dans l'expérience précédente. L'emplacement du texte n'est pas fourni par un seul phosphène, mais par un ensemble de phosphènes représentant la zone couverte par le texte. En effet, en général, ces blocs sont plus étendus que la zone couverte par un objet ou un visage. Nous pensons encore une fois qu'il n'existe pas de rendu visuel « idéal », et qu'il est intéressant au contraire d'en évaluer plusieurs.

L'objectif de cette expérience est d'évaluer ce nouveau rendu et de le confronter au rendu scoreboard dans une tâche de localisation de texte. Pour cela, nous simulons deux implants épirétiniens. Un implant composé d'une matrice 15x18 électrodes correspondant aux caractéristiques de la génération d'implants à venir (en l'occurrence ici, celles de l'Argus III), et un implant composé d'une matrice de 40x50 électrodes. Nous avons choisi ces résolutions pour les raisons suivantes : (1) nous souhaitions nous placer dans un contexte d'implant futur permettant la navigation, et nous savons que deux à trois cents phosphènes suffisent pour cela [Dagnelie et al. 2007; Wang et al. 2008] ; (2) nous souhaitions avoir une condition de contrôle (l'implant haute résolution) dans laquelle les sujets auraient suffisamment de phosphènes pour commencer à lire directement du texte dans une scène naturelle.

[17] Ces travaux ont été publiés dans [Denis et al. 2014].

Hypothèses de travail

Nous posons les deux hypothèses suivantes :

Hypothèse n°1 : notre nouveau rendu visuel basé sur une approche scoreboard augmenté permet de localiser efficacement des blocs de texte indépendamment de la résolution de l'implant simulé.

Hypothèse n°2 : pour détecter du texte dans une scène naturelle avec un rendu scoreboard, il est nécessaire que cette résolution soit très importante.

Protocole expérimental

Seize sujets (5 femmes et 11 hommes âgés de 23 à 38 ans) ont participé à cette expérience. La majorité d'entre eux était familiarisée avec un simulateur de vision prothétique.

Les sujets étaient assis à 57 cm d'un téléviseur. Ils portaient un casque de réalité virtuelle affichant une vision prothétique, et devaient localiser des blocs de texte dans les images qui s'affichaient sur l'écran devant eux (Figure 6-1A). Toutes les images étaient des photos de scènes naturelles prises dans les rues de Toulouse ou ses environs. Chaque sujet effectuait systématiquement cette tâche dans deux rendus (scoreboard et scoreboard augmenté) et deux résolutions de matrices (15x18 et 40x50) différents, soit un total de quatre conditions. Le rendu basé sur l'approche scoreboard augmenté différait de celui utilisé dans l'expérience de localisation des visages : lorsqu'un bloc de texte était reconnu dans l'image, il était restitué sous la forme d'un ensemble de phosphènes blancs, et pour éviter toute confusion, la luminance des autres phosphènes était diminuée (Figure 6-1B).

Pour chacune des conditions, nous avons utilisé un ensemble de 56 photos (1440 x 1080 pixels couvrant 69 x 52° d'angle visuel) contenant au plus un bloc de texte. La répartition était comme suit : 8 images sans texte, 12 avec un bloc en haut à gauche, 12 avec un bloc en haut à droite, 12 avec un bloc en bas à droite et 12 avec un bloc en bas à gauche. Trois tailles de police (1°, 2° et 4°) étaient équiprobablement distribuées dans les 48 photos contenant du texte.

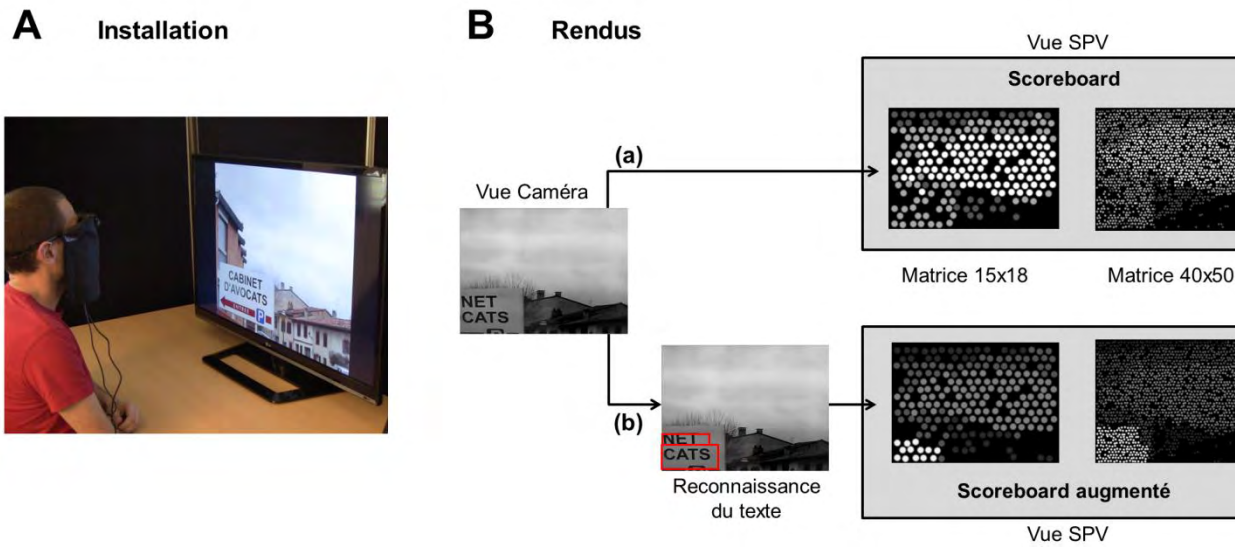


Figure 6-1 Disposition pour l'expérience de localisation de texte.
A - Vue d'ensemble du dispositif pour l'expérience de localisation de texte. Le sujet était assis à 57 cm d'un téléviseur et devait localiser des blocs de texte en vision prothétique simulée. B - Les deux rendus utilisés dans cette expérience : (a) rendu basé sur l'approche scoreboard, l'image est simplement réduite à la résolution de la matrice ; (b) rendu basé sur l'approche scoreboard augmenté, les blocs de texte reconnus apparaissent distinctement avec des phosphènes blancs.

Une condition était composée de 56 essais, correspondant aux 56 images. Les photos étaient affichées dans un ordre aléatoire. Pour chaque essai, la tâche consistait à dire si l'image contenait ou non un bloc de texte, le cas échéant à indiquer dans quel quadrant de l'écran il se situait. Tous les sujets effectuaient un total de 224 essais (56 essais par condition x 4 conditions). L'ordre des conditions était contrebalancé entre les sujets pour compenser un éventuel effet d'apprentissage.

Avant de démarrer l'expérience, nous indiquions aux sujets qu'ils devaient effectuer cette tâche de localisation de texte aussi vite et aussi précisément que possible. Les sujets s'entrainaient sur les quatre conditions pendant une dizaine de minutes sur un ensemble spécifique de photos, puis entamaient l'expérience. Un son très court marquait le début d'un essai. À partir de cet instant, les sujets, sans se rapprocher (un fil tendu horizontalement au niveau de leur menton les empêchait de s'avancer), étaient libres de bouger leur tête pour scanner l'écran du téléviseur et localiser l'éventuel bloc de texte. À la fin de l'essai, les sujets indiquaient leur réponse par oral (0 = pas de bloc, 1-4 = texte situé dans un des 4 cadrans de l'écran). À la fin des quatre conditions de rendu, les sujets remplissaient un questionnaire (Annexe 2). L'expérience durait en moyenne une heure.

Simulateur

Implants simulés

Pour les besoins de l'expérience, nous avons simulé deux matrices d'électrodes : la première avait une résolution de 15x18 électrodes et couvrait 12 x 17° d'angle visuel ; la seconde était constituée de 40x50 électrodes couvrant 19 x 25° d'angle visuel. Dans les deux cas, les électrodes étaient disposées de façon hexagonale pour augmenter la densité du maillage, et nous considérions que 10% d'entre elles étaient défaillantes et n'induisaient pas la perception de phosphènes. Ces derniers avaient une taille de 0,9° pour l'implant de résolution 12x17, et 0,5° pour l'implant 40x50. Pour les deux implants, les phosphènes induits étaient séparés par 0,2° d'angle visuel. Enfin, pour chaque phosphène, quatre niveaux de luminance ont été utilisés.

Architecture

La Figure 6-2 illustre l'architecture du simulateur choisie pour cette expérience. L'acquisition et la restitution des images se sont faites au travers du casque de réalité virtuelle Vuzix 1200AR (Vuzix Cor., CA, États-Unis). Cette expérience ne prenait pas en compte la position du regard pour afficher les phosphènes et n'utilisait donc pas d'oculomètre.

Contrairement à la détection d'objets et de visages, l'algorithme de SpikeNet Technology n'était pas adapté au contexte de la détection de texte. Notre choix s'est porté sur l'implémentation d'un algorithme de localisation de texte basé sur l'extracteur de régions d'intérêt MSER [Matas et al. 2004]. Les temps de calculs de cet extracteur ont été réduits grâce à une version plus performante [Nistér & Stewénius 2008] implémentée dans la librairie OpenCV[18].

[18] À noter que la future version OpenCV 3.0 inclura en natif un algorithme complet de détection de texte s'appuyant en partie sur MSER (http://docs.opencv.org/trunk/modules/objdetect/doc/erfilter.html).

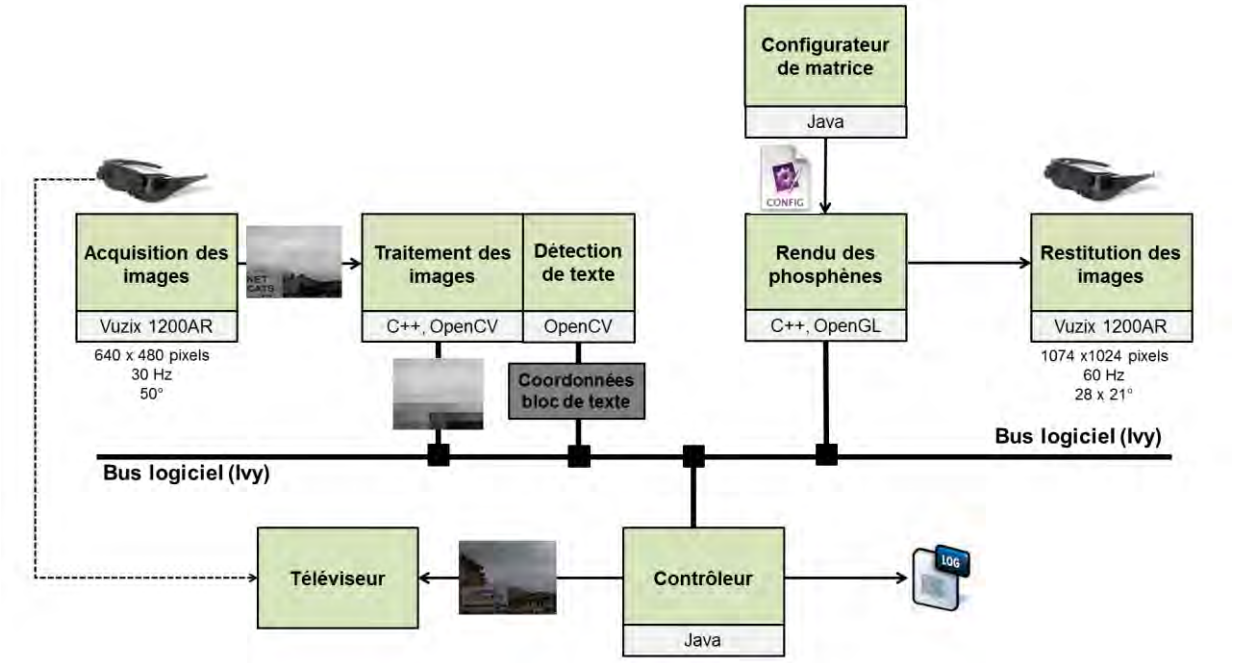


Figure 6-2 Architecture du simulateur pour l'expérience de localisation de texte.
L'acquisition et le traitement des images se font par le casque de réalité virtuelle Vuzix 1200AR. Un composant logiciel localise le texte, et envoie sur le bus Ivy les coordonnées des blocs détectés. Un contrôleur affiche successivement les images sur un téléviseur et consigne toutes les données de l'expérience dans un fichier de log.

MSER est utilisé dans des études récentes, et se montre être un excellent candidat pour la détection de texte en temps réel dans des scènes naturelles [Merino-Gracia et al. 2012]. A partir d'une image en niveaux de gris (640x480 pixels), MSER fournit un ensemble de candidats (de régions) qui peuvent potentiellement contenir un ou plusieurs caractères. Nous filtrons ces candidats pour exclure les régions trop petites, trop larges ou trop allongées. Enfin, nous regroupons horizontalement, puis verticalement, les régions restantes. L'ensemble obtenu correspond aux zones ayant la plus forte probabilité de contenir du texte. Le processus complet pour la localisation de texte dans une image dure moins de 50 ms.

Nous contrôlons l'expérience à l'aide d'un logiciel développé en Java. Celui-ci se charge d'afficher successivement les images sur l'écran du téléviseur, et de consigner dans un fichier de log l'ensemble des données de chacun des sujets. Tous les composants logiciels sont déployés sur un seul ordinateur (Intel dual-core i7, Windows 7 64 bits). Cette configuration totalement portable nous permet d'envisager de prochaines expérimentations en situation de mobilité.

Résultats

Nous avons analysé deux paramètres au cours de cette expérience : la précision, à savoir le pourcentage de réponses correctes, et le temps de réponse (temps en secondes pour fournir une bonne réponse). L'analyse et les tests statistiques ont été effectués avec R (R Foundation, Etats-Unis). La distribution de nos données n'étant pas gaussienne, et le nombre d'observations étant limité, nous avons utilisé des tests non paramétriques (tests de Wilcoxon). Pour corriger le seuil de significativité lors de comparaisons multiples, nous avons utilisé une correction de Bonferroni. Pour des raisons de simplicité, les quatre conditions seront nommées Scb1518, Scb4050, Aug1518 et Aug4050 en référence respectivement aux rendus basés sur le **SC**ore**B**oard, et à ceux basés sur le scoreboard **AUG**menté.

Toutes conditions confondues, le rapport entre la précision et le temps de réponse pour l'ensemble des sujets est illustré dans la Figure 6-3. La précision moyenne est de 70,1% (± 3,7) et le temps de réponse moyen est de 8,6 s (± 1,7).

La précision moyenne tous sujets confondus est au-dessus de 90% dans les conditions Aug1518 et Aug4050 (Figure 6-4 **gauche**). En revanche celle-ci est significativement plus basse dans les conditions Scb4050 et Scb1518 : respectivement 64,0% (± 5,8) et 32,9% (± 7,9) de réponses correctes. Toutes les comparaisons par paires entre les quatre conditions révèlent un effet significatif sur la précision, excepté entre Aug1518 et Aug4050 (Tableau 6-1). Dans chacune des conditions, 8 images parmi les 56 ne contenaient pas de bloc de texte. La précision moyenne sur ces images (considérées comme « pièges », mais les sujets connaissaient l'existence de cette condition) est de 75,2% (± 19,9). Ce résultat confirme que les sujets accomplissaient correctement la tâche, même lorsque le texte était difficile, voire impossible à percevoir.

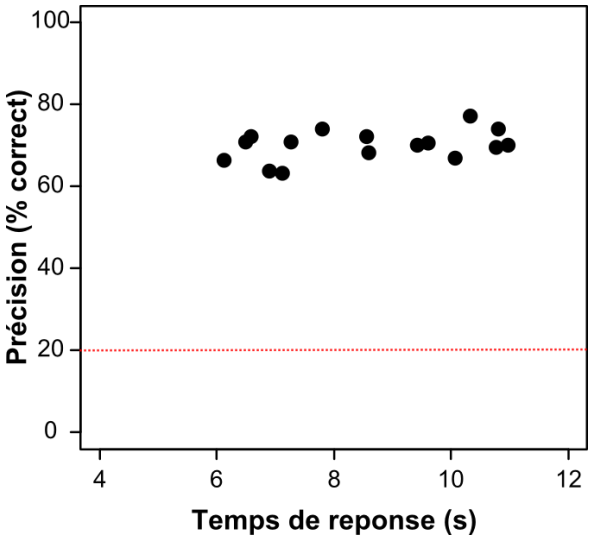


Figure 6-3 Précision vs. temps de réponse.
Précision (pourcentage de réponses correctes) en fonction du temps de réponse (en secondes) pour les 16 sujets.
La ligne en pointillés rouges correspond au niveau chance.

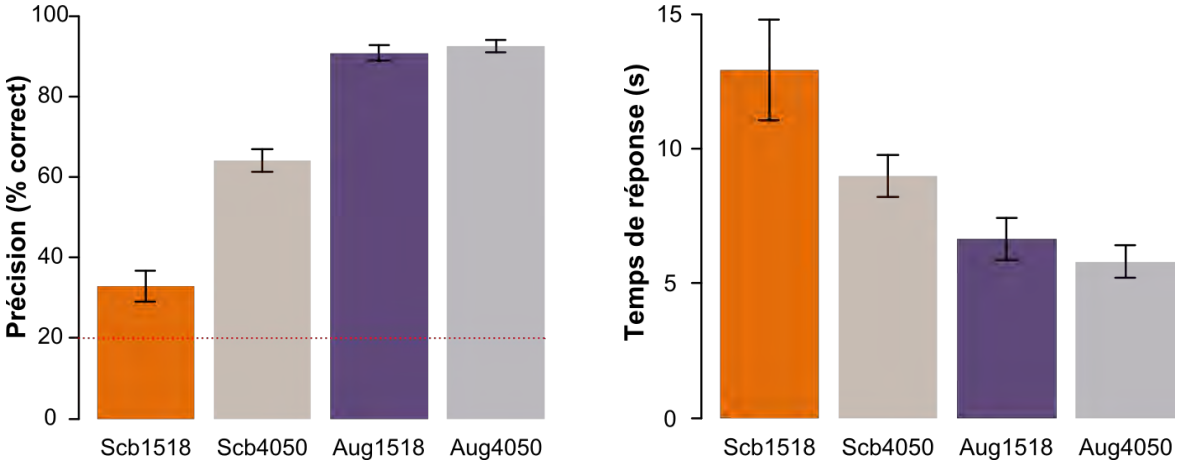


Figure 6-4 Précision et temps de réponse.
À gauche, la précision des sujets par condition. La ligne en pointillés rouges correspond au niveau chance (20% : 5 réponses possibles). À droite, le temps de réponse des sujets pour chacune des quatre conditions.

Tableau 6-1 Comparaisons paire à paire.
Comparaison par paires des différentes conditions pour la précision (à gauche) et pour le temps de réponse (à droite) (** p < 0,01).

Précision	Scb1518	Scb4050	Aug1518	Aug4050
Scb1518		**	**	**
Scb4050	**		**	**
Aug1518	**	**		0.2
Aug4050	**	**	0.2	

Temps de réponse	Scb1518	Scb4050	Aug1518	Aug4050
Scb1518		**	**	**
Scb4050	**		**	**
Aug1518	**	**		**
Aug4050	**	**	**	

Pour l'ensemble des sujets, le temps moyen pour localiser correctement un bloc de texte est de 12,9s (± 3,8) et 8,9s (± 1,6) dans les conditions Scb1518 et Scb4050, et de 6,6s (± 1,6) et 5,8s (± 1,2) dans les conditions Aug1518 et Aug4050 (Figure 6-4 **droite**).

Pour évaluer l'effet de la distance sur la capacité à localiser des blocs de texte, nous avons contrôlé la taille des caractères dans les images. Ainsi, les 48 images contenant du texte se répartissaient en 3 groupes de 12 images avec des lettres de 1°, 2° et 4°). Afficher des caractères à des tailles différentes revient à afficher des caractères à une taille définie, mais perçus à différentes distances. Ceci correspond plus souvent à une situation vécue (texte situé à différentes distances), mais est plus contraignant à manipuler en conditions d'expérience. La Figure 6-5 illustre la précision obtenue par condition, pour ces trois tailles. En condition Scb1518, les sujets sont très proches du niveau chance (23,8% (± 12,9)) pour les lettres de 1° (le niveau chance était à 20% car les sujets avaient 1 chance sur 5 de répondre correctement). Leur performance augmente légèrement pour les tailles plus grandes, mais restent très basses (32,0% pour 4°).

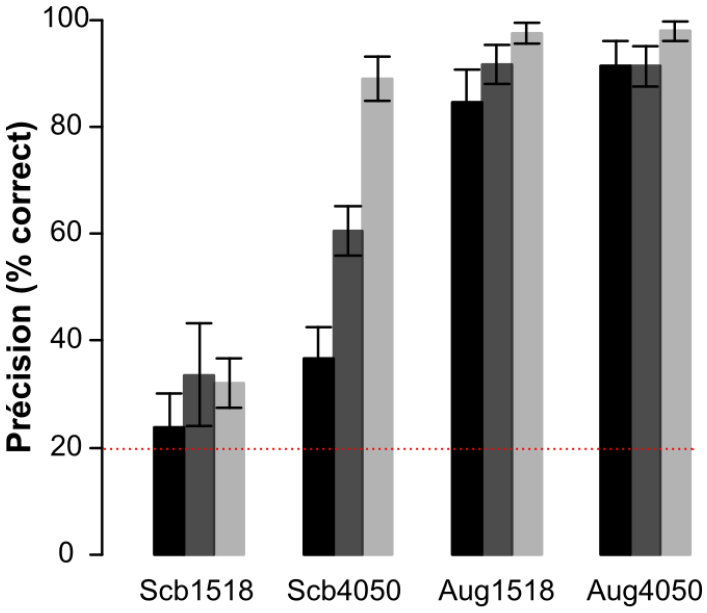


Figure 6-5 Précision par condition et taille de caractère.
Noir : 1°, gris foncé : 2°, gris clair : 4°. La ligne en pointillés rouges correspond au niveau chance.

Dans la condition Scb4050 (huit fois plus de phosphènes simulés), la précision reste faible pour localiser du texte dont les lettres mesurent 1° (36,7%). Lorsque ce texte

est composé de lettres de 4°, la performance se rapproche des scores obtenus avec les conditions Aug1518 et Aug4050 (89,1% vs. 97,6% et 98,0%). Dans ces deux conditions AUG, la précision est indépendante de la taille des lettres. Ce résultat est logique puisque notre algorithme de détection de texte était capable de localiser de façon similaire les trois tailles de caractères.

DISCUSSION

Validation ou rejet des hypothèses

Dans cette expérience de simulation de vision prothétique, nous avons proposé un nouveau rendu visuel basé sur l'approche scoreboard augmenté. Par l'utilisation d'un algorithme temps réel de reconnaissance de texte, nous mettons en évidence, dans le champ de vision des sujets, les zones contenant du texte. Avec l'approche scoreboard, classiquement implémentée dans les neuroprothèses visuelles actuelles, un implant contenant 270 électrodes ne permet pas de détecter dans son champ de vue du texte dont les caractères mesurent 4° ou moins. Cette taille apparente de 4° correspond à des lettres de 4 cm perçues à 50 cm, ou encore à des lettres de 8 cm situées à un mètre. Pour se ramener à une situation concrète, il n'est donc pas possible avec un implant de 270 électrodes de détecter par exemple un panneau de signalétique se situant à plus d'un mètre (les lettres de ces panneaux mesurant en moyenne entre 8 et 10 cm). Avec l'approche scoreboard augmenté et le rendu visuel que nous proposons, nos résultats indiquent qu'il devient possible, avec la même résolution d'implant, de détecter du texte à une distance au moins quatre fois supérieure. De plus, en scoreboard augmenté, la performance obtenue ne dépend pas de la résolution puisque les résultats sont sensiblement équivalents (validation de l'hypothèse n°1). En revanche, en scoreboard, même une configuration optimale contenant 2000 phosphènes ne suffit pas à détecter efficacement des blocs de texte de 1°. Avec cet implant, seuls les caractères les plus gros sont détectés correctement (validation de l'hypothèse n°2).

Vision par ordinateur

La distance de détection pourrait être encore plus grande en utilisant des caméras de résolution supérieure. En effet, dans ces conditions, l'algorithme de reconnaissance de texte serait capable de localiser des lettres encore plus petites. La limite pour

l'algorithme n'est en effet pas la taille apparente des lettres mais leur taille en pixels, qui augmente quand la résolution de la caméra augmente. Cependant, augmenter la résolution des images implique automatiquement un temps de calcul supérieur pour l'algorithme.

Pour détecter le texte en temps réel, nous avons implémenté une version simplifiée de l'algorithme décrit par Merino-Gracia et al. [Merino-Gracia et al. 2012]. Elle repose sur l'utilisation de l'extracteur de régions d'intérêt MSER. Dans cette expérience, comme dans les trois autres, notre objectif n'était pas d'évaluer la performance de cet algorithme, mais plutôt de reposer sur une implémentation réaliste, et d'être conscient des forces et des faiblesses de ce type d'algorithmes. Ceux qui sont temps réel ne sont pas aussi robustes que ceux employés pour la détection de visages. Certaines solutions offrent des performances de détection impressionnantes, mais au détriment du temps de calcul. L'algorithme que nous avons implémenté produisait un assez grand nombre de fausses détections (sur plus de la moitié des images). Cependant, comme dans nos deux expériences de localisation d'objets, nous avons observé que les sujets apprenaient rapidement à les ignorer : les blocs de phosphènes blancs représentant du texte étaient sensiblement plus stables que ceux correspondant à des fausses détections. Par contre dans une situation plus réaliste, on peut faire la supposition que l'utilisateur serait rapidement gêné si trop de fausses détections lui étaient présentées.

Dans notre simulateur, l'algorithme est exécuté sur un ordinateur contenant un microprocesseur beaucoup plus puissant que celui que l'on peut trouver dans les dispositifs mobiles actuels. Cependant, ces derniers évoluent très vite et intègrent des composants de plus en plus puissants. Le processeur mobile le plus évolué à ce jour est développé par Nvidia (CA, États-Unis) : le Tegra K1 est composé de quatre cœurs, cadencés à 2.3GHz, et d'un processeur graphique comportant 192 cœurs CUDA (plus puissant que les processeurs des consoles Xbox 360 et PS3). À très court terme, une piste intéressante serait de programmer le même algorithme pour qu'il s'exécute en parallèle sur les nombreux cœurs d'une telle puce graphique, afin de le rendre temps réel sur un dispositif mobile.

Comportement « visuomoteur »

Avec un implant rétinien composé de 60 électrodes, quelques sujets sont capables, dans des environnements très contrôlés, de reconnaître de grandes lettres, voire de lire très lentement des mots pour certains d'entre eux [da Cruz et al. 2013]. Avec plus de 200 électrodes, cette même tâche deviendra certainement réalisable pour tous les sujets implantés, et dans des délais plus raisonnables. Cependant, comme nous le montrons, la résolution spatiale restera insuffisante pour localiser les zones de texte dans la plupart des situations rencontrées en extérieur. Or, avant de pouvoir lire du texte, il est nécessaire de savoir où il se situe. Avec le rendu visuel que nous proposons, nous rendons accessible certaines informations écrites : les sujets sont capables, avec des implants basse résolution, de connaître l'emplacement de zones de texte plus ou moins éloignées, et de s'approcher de l'une d'entre elles pour pouvoir lire son contenu.

Scénario d'usage

Plutôt que d'avoir à s'approcher physiquement du texte pour lire, nous pourrions également imaginer que les utilisateurs contrôlent directement le zoom de la caméra, pour agrandir ces blocs de texte sans bouger. Avec cette fonctionnalité, les personnes implantées auraient la possibilité de lire du texte éloigné, voire inaccessible (des panneaux en hauteur, le nom des magasins dans une rue, le numéro d'un bus en approche etc.). Mieux encore, l'utilisateur pourrait sélectionner un bloc de texte (par exemple le centrer dans son champ de vision), et demander au système de le lire automatiquement. En effet, les algorithmes de reconnaissance optique de caractères deviennent performants même dans des images naturelles [Bissacco et al. 2013]. Il suffirait ensuite d'utiliser un logiciel de synthèse vocale pour verbaliser le texte détecté.

Malgré tout, cette approche a une limite importante : dans beaucoup de contextes (une rue, une gare, une station de métro…) énormément d'informations textuelles nous entourent et le rendu visuel que nous proposons, spécifique à la localisation de texte risque rapidement d'être inutilisable car beaucoup trop surchargé. Une idée serait de pouvoir facilement les filtrer (par distance ou par taille) pour ne présenter qu'un ensemble restreint de blocs de texte.

Conclusion

Comme pour l'expérience de localisation de visage nous avons démontré l'intérêt de l'approche scoreboard augmenté dans l'accomplissement d'une tâche spécifique. En effet, avec un rendu basé sur cette approche, une quantité très limitée de phosphènes suffit à rendre accessible à l'utilisateur l'emplacement des zones de texte qui l'entourent, afin qu'il s'en approche. Dans la dernière partie, nous dressons le bilan de nos expérimentations, et ouvrons vers les perspectives à court et moyen-terme qui résultent de ces travaux.

DISCUSSION ET PERSPECTIVES

VERROUS INITIAUX

Nous sommes partis du constat que les neuroprothèses visuelles actuelles ne permettent pas de restaurer une vision suffisamment fonctionnelle pour assurer les tâches de la vie quotidienne et gagner en indépendance. Ces systèmes sont principalement limités par la résolution qu'ils offrent : seuls quelques dizaines de phosphènes distincts sont restitués aux personnes implantées. Les générations d'implants à venir dans les cinq à dix prochaines années proposeront des résolutions supérieures mais resteront tout de même très limitées. De façon intéressante, une grande majorité des neuroprothèses visuelles inclut une caméra externe pour capter les images à restaurer. Pour ces systèmes, il est donc possible d'appliquer en temps réel des traitements de vision par machine sur les images acquises par la caméra, afin de générer des motifs de phosphènes plus facilement utilisables par l'utilisateur de l'implant.

SOLUTIONS PROPOSEES

Partant de ce principe, nous avons suggéré l'utilisation d'algorithmes issus de la vision par ordinateur pour le développement de rendus visuels adaptés aux tâches de la vie courante. Notre idée a été d'utiliser ces algorithmes pour extraire une information de haut niveau, utile à la réalisation d'une tâche en particulier, et de restituer cette information par l'intermédiaire de quelques phosphènes. Nous nous sommes particulièrement intéressés à la problématique de la localisation d'objets (au sens large du terme) dans un environnement naturel.

Dans un premier temps, et afin d'évaluer de nouveaux rendus visuels, nous avons implémenté un simulateur de vision prothétique qui nous a permis d'étudier ces questions chez des personnes voyantes, l'accès aux patients implantés étant encore trop restreint. Nous avons investigué deux approches : la première, **l'approche par**

localisation de points d'intérêt, a consisté à extraire des images la position d'objets particuliers, puis de restituer celle-ci par le biais de phosphènes. Dans la seconde approche, **l'approche scoreboard augmenté**, les informations de localisation ont été ajoutées en surimpression du rendu scoreboard (sous-échantillonnage de l'image) classiquement généré dans ces systèmes. La première approche a été évaluée dans un contexte de localisation d'objets du quotidien et de localisation de visages, la seconde dans un contexte de localisation de visages et de localisation de blocs de texte.

RESULTATS

Avec l'approche par localisation de points d'intérêt, tous les sujets sans exception sont parvenus à localiser et atteindre des objets de la vie courante disposés sur une table. Nous avons montré que cette tâche était réalisable avec un implant cortical restituant seulement 9 phosphènes et avec un implant rétinien en restituant 60. Avec une approche scoreboard, vingt fois plus de phosphènes ont été nécessaires pour parvenir à des performances similaires. Dans une autre simulation de l'implant rétinien de 60 électrodes, l'approche par localisation de points d'intérêt a également permis de localiser un ou deux visages situés à une distance comprise entre 1,5 m et 2,5 m, tâche ici aussi non réalisable pour ce nombre d'électrodes avec un rendu scoreboard.

Tous les sujets ont pu effectuer cette même tâche avec l'approche scoreboard augmenté. Cette dernière a été évaluée dans un troisième contexte : la localisation d'un bloc de texte dans une scène naturelle. Cette approche s'est révélée pertinente car les sujets ont réussi à localiser la très grande majorité des blocs de texte, pour des caractères mesurant de 1 à 4° d'angle visuel. Les performances de tous les sujets ont été équivalentes, que l'on simule un implant rétinien restituant 270 ou 2000 phosphènes distincts. En approche scoreboard, l'implant de faible résolution (270 phosphènes) n'a pas permis de localiser le texte quelle que soit sa taille, et seuls les caractères de 4° ont pu être localisés efficacement avec l'implant constitué de 2000 électrodes.

Nous pouvons tirer quelques conclusions supplémentaires de nos expérimentations. Tout d'abord, l'expérience de détection de visage permet de localiser les visages

plus vite qu'avec l'approche scoreboard augmenté. Deuxièmement, nos deux nouvelles approches ne nécessitent pratiquement pas d'entraînement. Les utilisateurs ont très vite été capables d'interpréter l'information restituée pour réaliser les différentes tâches (quelques minutes). De plus la performance des sujets naïfs du système est très vite comparable à celle des sujets experts. Troisièmement, avec ces deux approches, la réussite de la tâche n'est pas dépendante de la résolution de l'implant. Pour preuve, les sujets ont localisé et atteint efficacement des objets aussi bien avec un implant constitué de 9 électrodes qu'avec un implant en comprenant plus de mille. Enfin, en analysant leur comportement, nous avons pu nous rendre compte que les sujets apprenaient à « reconnaître » les fausses détections générées par les algorithmes de vision en filtrant par des mouvements de tête le ou les phosphènes qui n'étaient pas « stables ». En effet, la plupart du temps, un changement rapide de point de vue permet de faire disparaître ces fausses détections.

Au regard de l'ensemble de ces résultats, nous pouvons indiquer que dans ces trois contextes expérimentaux spécifiques, l'utilisation d'algorithmes de vision a clairement amélioré l'utilisabilité des neuroprothèses visuelles. Ces résultats sont intéressants, mais ils sont cependant à relativiser. Ils ont été obtenus en simulation de vision prothétique, avec des personnes voyantes. Bien qu'elles fournissent un moyen simple et peu coûteux pour évaluer l'utilisabilité de nouveaux rendus visuels, ces simulations restent optimistes par rapport à la réalité. Par exemple chez une personne implantée, les phosphènes n'ont pas tous la même apparence [Dobelle & Mladejovsky 1974; Humayun et al. 1996; Delbeke et al. 2003]. Les niveaux de luminance sont variables selon les sujets, tout comme la dynamique temporelle des phosphènes [Pérez Fornos et al. 2012]. Il n'est donc pas raisonnable de conclure que des performances équivalentes auraient été atteintes par des personnes réellement implantées. De plus, dans les trois tâches, nous nous sommes placés dans des environnements relativement contrôlés, favorables à la précision des algorithmes de vision. Dans toutes les expériences, les conditions lumineuses ne variaient pas ou très peu. Dans l'expérience de localisation d'objet, pour simplifier leur modélisation, ces derniers étaient présentés à des positions contrôlées (une modélisation plus fine aurait cependant été possible). Dans l'expérience de localisation de texte, les images utilisées ne contenaient qu'un bloc de texte. En

condition réelle, la présence de plusieurs blocs de texte peut rendre la scène plus complexe, mais l'expertise de l'utilisateur peut entrer en jeu pour résoudre ces situations. Les fausses détections que les sujets avaient à ignorer dans notre expérience pourraient être diminuées à l'avenir en s'appuyant sur des algorithmes de vision plus robustes aux conditions d'éclairage et au changement de point de vue.

LOCALISATION DE POINTS D'INTERET, SCOREBOARD AUGMENTE

Au tout début de la thèse, nous avons introduit une approche très simple, en partant du principe qu'un rendu visuel contenant un unique phosphène suffirait à localiser un point d'intérêt. C'est ce que nous avons évalué au travers de l'approche par localisation de points d'intérêt. Nos résultats, concluants, ont permis de valider que cette approche pouvait améliorer l'utilisabilité des neuroprothèses actuelles qui ne restituent que quelques dizaines de phosphènes.

Bien que difficile à interpréter, un rendu scoreboard fournit néanmoins un contexte. Nous pensons que dans certaines tâches, ce contexte peut être utile à l'utilisateur. Pour cela nous avons développé et évalué l'approche scoreboard augmenté qui combine un rendu scoreboard et une information de localisation. Cette approche a également permis la réalisation de différentes tâches.

Atouts

Les deux approches ont des atouts communs évoqués dans les résultats ci-dessus. En premier lieu, l'un de leurs points forts est que l'interprétation de leurs rendus visuels ne requiert pratiquement pas d'apprentissage. D'autre part, comme les informations de localisation sont « codées » par l'intermédiaire de très peu de phosphènes, la réalisation des différentes tâches est indépendante de la résolution de l'implant, mais également de l'objet d'intérêt. Par exemple dans nos expériences, les sujets ont la même information (un phosphène unique) qu'ils aient à localiser un téléphone, une tasse ou un visage. Un objet visuellement plus complexe à détecter ne nécessite donc pas une résolution d'implant plus grande.

Ces approches permettent aussi de proposer à l'utilisateur différents codes lui permettant de localiser des objets d'intérêt. Dans nos travaux, nous en avons proposé deux : une détection est restituée soit par un phosphène unique (exemple

de la localisation d'objets et de visages), soit par un groupe de phosphènes représentant la dimension de l'objet (exemple des blocs de texte). D'autres codes pourraient être imaginés et mieux encore, conçus directement avec les utilisateurs finaux.

Comparée à l'approche scoreboard augmenté, l'approche par localisation à l'avantage de fournir une information très concise, rapidement interprétable. L'approche scoreboard augmenté quant à elle, ajoute des informations de contexte aux informations de localisation. Ce contexte pourrait apporter une aide supplémentaire à l'utilisateur et l'aider à encore mieux interpréter son environnement.

Limites

Les deux approches proposées sont très efficaces lorsque le sujet cherche un objet précis dans son environnement. Elles trouvent probablement leurs limites lorsque le nombre d'objets à restituer augmente et devient très important (par exemple, pour un objet présent en de nombreux exemplaires ou pour un groupe d'objets). Les rendus visuels générés deviendraient vraisemblablement inutilisables par surcharge d'informations. Pour chaque rendu, des tests supplémentaires permettraient d'estimer plus précisément le nombre de points d'intérêt que l'utilisateur serait capable d'interpréter en parallèle. Dans un même temps, il serait intéressant d'étudier si la localisation de plusieurs objets d'intérêt ne structurerait pas l'espace de l'utilisateur lui permettant de mieux l'appréhender. Une autre limite tient au fait que ces approches reposent sur des algorithmes de vision qui n'ont pas un taux de détection de 100% et un taux d'erreur de 0%. Dans un environnement moins contrôlé, les conditions de luminosité et le point de vue affecteraient certainement la précision des algorithmes et les sujets pourraient manquer des objets non détectés par le système. Les utilisateurs pourraient également avoir à exclure d'avantage de fausses détections. Dans tous les cas, il est probable que les utilisateurs abandonneraient l'usage de ces rendus si les algorithmes de vision artificielle utilisés n'étaient pas assez robustes. Toutes ces questions devraient être posées dans la continuité de cette thèse.

VISION PAR ORDINATEUR ET NEUROPROTHESES VISUELLES

Si nous nous sommes intéressés dès le départ de ce travail à l'usage d'algorithmes provenant de la vision par ordinateur c'est avant tout parce que les neuroprothèses visuelles incluent pour la plupart d'entre elles une caméra et un processeur capable de traiter les images. Pour autant, une contrainte majeure doit être respectée : le traitement embarqué doit être temps réel pour ne pas affecter la boucle « visuomotrice » (par temps réel on entend ici 15 à 20 Hz). En effet, si ce traitement n'était pas assez rapide, un décalage se ferait entre la scène filmée par la caméra et l'information restituée à l'utilisateur ce qui compromettrait la boucle perception-action de ce dernier, et probablement, en conséquence, sa compréhension de la scène.

Aujourd'hui les neuroprothèses visuelles sont fonctionnellement très limitées car leur résolution est faible, et l'approche scoreboard utilisée pour l'ensemble de ces systèmes ne permet pas aux personnes implantées de suffisamment comprendre la scène. De façon simplifiée, avec cette approche, une scène acquise par une caméra est encodée par des millions de pixels en couleur va être restituée par des dizaines de phosphènes en niveaux de gris. La perte d'information est immense. Tout ce qui a un sens dans la scène de départ (le sol, les murs, la position des objets, des personnes etc.) n'est pas transféré dans le rendu phosphénique.

Dans le domaine de la vision par ordinateur, de nombreux algorithmes permettent d'analyser une scène et d'en extraire le sens (à quelle distance est cette personne ? où se situe le verre ? que contient ce texte ? ...). Le domaine est très large, dynamique et aujourd'hui beaucoup de ces algorithmes sont temps réel. Nous nous sommes particulièrement intéressés à ceux qui permettent de connaître l'emplacement d'objets spécifiques. En effet, si nous sommes capables d'extraire très rapidement la position d'un objet, il suffit de restituer un phosphène à l'utilisateur pour lui transmettre cette information. La position d'objets d'intérêt n'est pas la seule information qu'il est possible d'extraire d'une scène en temps réel, et d'autres algorithmes auraient pu jouer ce rôle. Des travaux se sont par exemple intéressés à l'extraction du sol en segmentant l'image [McCarthy et al. 2011], ou à la détection d'obstacles par l'usage de carte de saillance [Parikh et al. 2013]. La vision par ordinateur peut donc jouer un rôle fondamental dans l'extraction et la restauration d'informations essentielles à la réalisation d'une tâche, informations qui aujourd'hui

ne sont pas interprétables avec les implants actuels. En jouant ce rôle, la vision par machine peut indirectement restaurer aux personnes implantées un comportement « visuomoteur ».

L'un des avantages à utiliser la vision par machine est que le coût d'intégration des algorithmes dans les neuroprothèses visuelles se limite pratiquement à leur implémentation. De plus, certaines de ces implémentations sont déjà disponibles « sur étagère » dans des bibliothèques open source très populaires et très bien documentées[19]. D'autre part, ce domaine étant extrêmement dynamique, la précision des algorithmes ne cesse de progresser, et il serait très facile de mettre à jour le système de l'utilisateur lorsque des solutions plus performantes viendraient à être disponibles.

La limite principale à l'usage d'algorithmes de vision par ordinateur est leur temps d'exécution comme évoqué plus haut. D'un autre côté, les capacités de calcul des dispositifs embarqués s'améliorent chaque année de façon significative. De façon plus spécifique, les algorithmes de détection et de reconnaissance d'objets les plus rapides aujourd'hui s'appuient sur un apprentissage supervisé de modèles. Dans une situation réaliste, il n'est pas envisageable de modéliser tous les objets de la vie courante. Cependant, cette limitation pourrait être levée par la constitution de large base de données de modèles. Le projet Google Goggle montre par exemple qu'il est déjà possible de localiser et reconnaître automatiquement une grande variété d'objets dans des scènes naturelles. Les bases de données pourraient également être constituées de façon collaborative afin qu'elles contiennent rapidement un ensemble d'objets utiles pour les non-voyants [Bigham et al. 2010].

NEUROPROTHESES VISUELLES CONTEXTUELLES OU INTERACTIVES ?

Nos résultats suggèrent qu'un ensemble de rendus visuels pourraient être bénéfiques pour les utilisateurs suivant le contexte et la tâche qu'ils ont à réaliser. En situation de mobilité, différents rendus ont été proposés [Vergnieux et al. 2014; Mohammadi et al. 2012; McCarthy et al. 2013; McCarthy & Barnes 2012; Feng & McCarthy 2013]. Pour la localisation d'objets d'intérêts, d'autres rendus seraient

[19] OpenCV (http://opencv.org), VXL (http://vxl.sourceforge.net), CCV (http://libccv.org)

probablement mieux adaptés [Denis et al. 2012; Denis et al. 2013; Denis et al. 2014; Lui et al. 2012].

Dans cette optique, nous pouvons nous demander s'il serait plus judicieux pour l'utilisateur que le système génère automatiquement un rendu visuel dépendant du contexte, ou si au contraire ce choix doit être contrôlé explicitement par la personne implantée. Dans la première solution, l'intérêt pour les personnes serait de continuer à utiliser leur neuroprothèse comme elles le font aujourd'hui. En revanche, il est possible que certaines décisions prises par le système ne soient pas en adéquation avec leur besoin. Avec une neuroprothèse visuelle interactive, les sujets pourraient trouver trop complexe leur utilisation. Cependant, avec un minimum d'apprentissage, il leur serait très vite possible de choisir à la demande un rendu visuel adapté au contexte de la tâche qu'ils souhaitent réaliser.

Au-delà de la multiplicité des rendus visuels, certaines actions nécessiteraient dans tous les cas une interaction avec le système. Il pourrait par exemple être intéressant de pouvoir zoomer dans l'image sans avoir à se déplacer [Boyle 2008; Pérez Fornos et al. 2011; Tsai et al. 2009; Buffoni et al. 2005; Wang, Wu, et al. 2014]. Avec les approches que nous avons proposées, il serait aussi utile de pouvoir « sélectionner » un point d'intérêt pour en connaître son contenu (connaître l'identification d'un visage, le contenu d'un bloc de texte…).

Dans ce cadre, nous proposons en preuve de concept un prototype[20] de système interactif. En plus de la neuroprothèse visuelle, le système est composé d'une application logicielle (Figure Disc-1) déployable sur un dispositif mobile (un smartphone, une smartwatch, des Google Glass…). L'utilisateur interagit directement avec sa neuroprothèse par le biais de cette application, soit par commande vocale, soit plus discrètement par interaction tactile avec le dispositif portable (Figure Disc-2). Ainsi il lui est de possible de changer très facilement de rendu visuel (mode) à la demande. Pour chaque mode, une liste d'actions est disponible. Par exemple, dans le mode localisation de blocs de texte, l'utilisateur peut zoomer, dézoomer ou encore sélectionner un bloc dont il veut lire le contenu à distance. Enfin, chaque action est suivie d'un feedback audio et/ou visuel pour conforter la personne dans

[20] A l'heure où ce manuscrit est écrit, ce prototype est en cours de développement.

son choix. Ce premier prototype démontre qu'avec quelques fonctionnalités simples, l'utilisateur pourrait interagir facilement avec sa neuroprothèse, en contrôlant différents rendus visuels adaptés aux multiples situations de la vie courante.

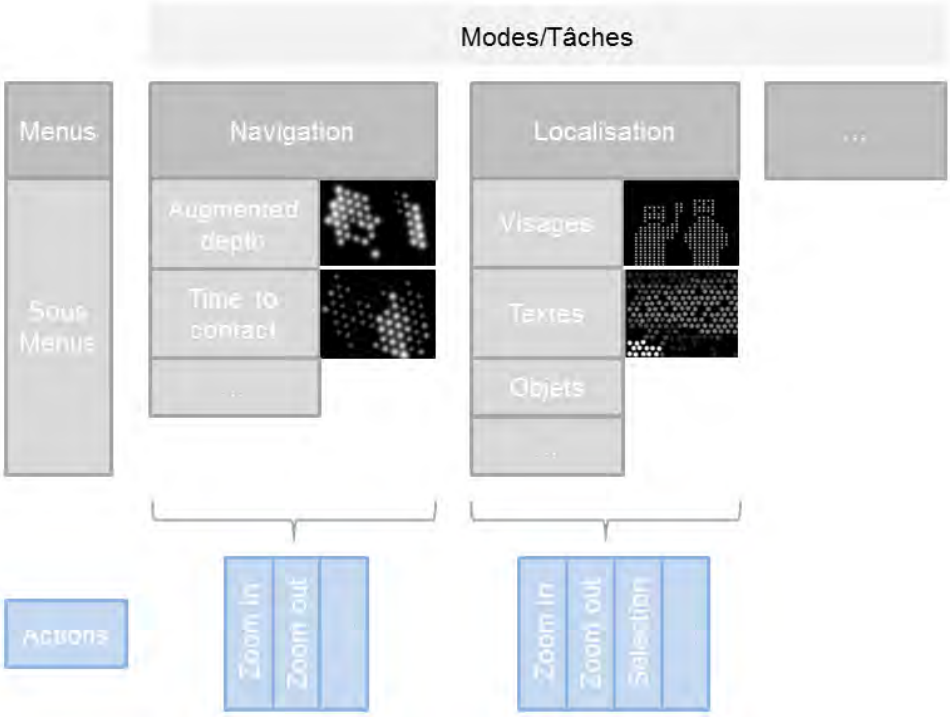


Figure Disc-1 Rendus visuels et actions disponibles dans le logiciel servant à l'interaction avec la neuroprothèse.
Plusieurs modes sont accessibles pour la navigation et la localisation d'objets d'intérêt. Une liste d'action est disponible pour chacun des modes (Source des simulations : Denis, Macé, & Jouffrais, 2014; Lui, Browne, Kleeman, Drummond, & Li, 2012; McCarthy & Barnes, 2012; McCarthy, Feng, & Barnes, 2013).

Action	Interaction		Feedback
	Tactile	Orale	
Sélectionner un rendu	SWIPE + TAP	« mode navigation »	Retour audio des noms de menus et sous-menus + Retour visuel (modification du rendu)
Zoomer / Dézoomer	ZOOM / PINCH	« action zoomer » / « action dézoomer »	Retour audio sur le niveau de zoom + Retour visuel (modification de la scène)
Sélectionner un bloc de texte	centrer le bloc + TAP	centrer le bloc dans le champ de vision + « action sélectionner »	Retour audio sur le contenu du texte

Figure Disc-2 Exemples d'interactions possibles avec la neuroprothèse visuelle.
Chaque action est exécutable par une interaction avec la surface tactile du dispositif portable ou activable par commande vocale.

PERSPECTIVES

Conception de nouveaux rendus visuels

Dans ce travail, nous avons proposé d'utiliser des algorithmes de vision pour concevoir des rendus visuels compréhensibles et fonctionnels. Pour autant, le développement de rendus phosphéniques adaptés ne se limite probablement pas au seul usage de la vision par machine. Aujourd'hui, la plupart d'entre nous embarquons un nombre important de capteurs (GPS, accéléromètre, gyroscope…) intégrés à nos dispositifs mobiles, et les données fournies par ces capteurs pourraient jouer indirectement un rôle dans le développement de nouveaux rendus. Par exemple, la position géographique de l'utilisateur pourrait conduire le système à lui indiquer la station de métro ou de bus la plus proche. Sur le même principe, l'usage d'informations provenant d'objets communicants - l'internet des objets [Atzori et al. 2010] - pourrait s'avérer également pertinent dans la conception de rendus visuels fonctionnels.

Vers une conception participative

À moyen terme, il serait particulièrement intéressant de tester les approches que nous proposons chez des personnes implantées. Ceci pourrait devenir réalisable assez rapidement après l'annonce en France d'un remboursement par la sécurité sociale de l'implant Argus 2 pour un maximum de 36 personnes par an via le nouveau dispositif « Forfait Innovation ». En conséquence, trois hôpitaux (Paris, Strasbourg, Bordeaux) vont prendre en charge les actes de chirurgie et la mise en place du système. Une collaboration avec l'un de ces centres hospitaliers permettrait de valider l'utilité de nos différents rendus visuels auprès de leurs patients. Une autre collaboration pourrait voir le jour avec l'Institut de la Vision à Paris qui développe son propre implant avec la société Pixium Vision, et conçoit de nouveaux algorithmes de traitement d'images capturées par des caméras biomimétiques [Benosman et al. 2012]. En plus d'évaluer en situation réelle l'utilité de différentes approches, l'accès à des personnes implantées permettrait de les impliquer dans la conception et le développement de rendus visuels répondant précisément à leurs besoins.

BIBLIOGRAPHIE

Ahuja, A.K. et al., **2011**. Blind subjects implanted with the Argus II retinal prosthesis are able to improve performance in a spatial-motor task. *The British Journal of Ophthalmology*, 95(4), pp.539–43.

Alahi, A., Ortiz, R. & Vandergheynst, P., **2012**. FREAK: Fast Retina Keypoint. In *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, pp. 510–517.

Asher, A. et al., **2007**. Image processing for a high-resolution optoelectronic retinal prosthesis. *IEEE Transactions on Bio-medical Engineering*, 54(6 Pt 1), pp.993–1004.

Atzori, L., Iera, A. & Morabito, G., **2010**. The Internet of Things: A survey. *Computer Networks*, 54(15), pp.2787–2805.

Auvray, M., **2004**. *Immersion et perception spatiale*.

Auvray, M., Hanneton, S. & O'Regan, J.K., **2007**. Learning to perceive with a visuo-auditory substitution system: localisation and object recognition with "the vOICe." *Perception*, 36(3), pp.416–30.

Bach-y-Rita, P. et al., **1998**. Form perception with a 49-point electrotactile stimulus array on the tongue: a technical note. *Journal of Rehabilitation Research and Development*, 35(4), pp.427–30.

Bach-y-Rita, P., **1983**. Tactile vision substitution: past and future. *The International Journal of Neuroscience*, 19(1-4), pp.29–36.

Bach-y-Rita, P. et al., **1969**. Vision substitution by tactile image projection. *Nature*, 50(5184), pp.83–91.

Bach-y-Rita, P. & W. Kercel, S., **2003**. Sensory substitution and the human–machine interface. *Trends in Cognitive Sciences*, 7(12), pp.541–546.

Bak, M. et al., **1990**. Visual sensations produced by intracortical microstimulation of the human occipital cortex. *Medical & Biological Engineering & Computing*, 28(3), pp.257–259.

Barry, M.P. & Dagnelie, G., **2011**. Simulations of Prosthetic Vision. In *Visual Prosthetics*. pp. 319–341.

Bay, H., Tuytelaars, T. & Gool, L. Van, **2006**. Surf: Speeded up robust features. In *European Conference on Computer Vision*. pp. 404–417.

Bec, J.-M. et al., **2012**. Characteristics of laser stimulation by near infrared pulses of retinal and vestibular primary neurons. *Lasers in Surgery and Medicine*, 44(9), pp.736–45.

Bec, J.-M., **2010**. *Etude de la stimulation laser de neurones pour des applications de prothèses visuelles*.

Benosman, R. et al., **2012**. Asynchronous frameless event-based optical flow. *Neural Networks*, 27, pp.32–7.

Bigham, J.P. et al., **2010**. VizWiz::LocateIt - enabling blind people to locate objects in their environment. In *Computer Vision and Pattern Recognition Workshops*. pp. 65–72.

Bissacco, A. et al., **2013**. PhotoOCR: Reading Text in Uncontrolled Conditions. In *International Conference on Computer Vision*. IEEE, pp. 785–792.

Boyle, J.R., **2008**. Region-of-interest processing for electronic visual prostheses. *Journal of Electronic Imaging*, 17(1), pp.1–12.

Boyle, J.R., Maeder, A. & Boles, W., **2002**. Image enhancement for electronic visual prostheses. *Australasian Physics & Engineering Sciences in Medicine*, 25(2), pp.81–86.

Boyle, J.R., Maeder, A. & Boles, W., **2003**. Scene specific imaging for bionic vision implants. In *International Symposium on Image and Signal Processing and Analysis.* IEEE, pp. 423–427.

Brelén, Må.E. et al., **2005**. Creating a meaningful visual perception in blind volunteers by optic nerve stimulation. *Journal of Neural Engineering*, 2(1), pp.S22–8.

Brindley, G.S. & Lewin, W.S., **1968**. The sensations produced by electrical stimulation of the visual cortex. *The Journal of Physiology*, 196(2), pp.479–93.

Buffoni, L.-X., Coulombe, J. & Sawan, M., **2005**. Image processing strategies dedicated to visual cortical stimulators: a survey. *Artificial Organs*, 29(8), pp.658–64.

Buisson, M. et al., **2002**. Ivy : Un bus logiciel au service du développement de prototypes de systèmes interactifs. In *French-speaking Conference on Human Computer Interaction*. Poitiers, France: ACM, pp. 223–226.

Button, J. & Putnam, T., **1962**. Visual responses to cortical stimulation in the blind. *J Iowa Med Soc*, (1), pp.1–21.

Canny, J., **1986**. A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6), pp.679–714.

Capelle, C. et al., **1998**. A real-time experimental prototype for enhancement of vision rehabilitation using auditory substitution. *IEEE Transactions on Bio-medical Engineering*, 45(10), pp.1279–93.

Caspi, A. et al., **2009**. Feasibility Study of a Retinal Prosthesis. *Archives of Ophthalmology*, 127(4), pp.398–401.

Castagnola, V. et al., **2014**. Parylene-based flexible neural probes with pedot coated surface for brain stimulation and recording. *Biosensors and Bioelectronics*, in press.

Cha, K. et al., **1992**. Reading speed with a pixelized vision system. *Journal of the Optical Society of America A*, 9(5), p.673.

Cha, K., Horch, K.W. & Normann, R.A., **1992**. Mobility performance with a pixelized vision system. *Vision research*, 32(7), pp.1367–72.

Cha, K., Horch, K.W. & Normann, R.A., **1992**. Simulation of a phosphene-based visual field: visual acuity in a pixelized vision system. *Annals of Biomedical Engineering*, 20(4), pp.439–49.

Chader, G.J., Weiland, J.D. & Humayun, M.S., **2009**. Artificial vision: needs, functioning, and testing of a retinal electronic prosthesis. *Progress in Brain Research*, 175, pp.317–32.

Chai, X. et al., **2008**. C-sight visual prostheses for the blind. *Engineering in Medicine and Biology Magazine*, 27(5), pp.20–8.

Chai, X. et al., **2007**. Recognition of pixelized Chinese characters using simulated prosthetic vision. *Artificial Organs*, 31(3), pp.175–82.

Chang, M.H. et al., **2012**. Facial identification in very low-resolution images simulating prosthetic vision. *Journal of Neural Engineering*, 9(4), p.046012.

Chen, H., Tsai, S. & Schroth, G., **2011**. Robust text detection in natural images with edge-enhanced maximally stable extremal regions. In *International Conference on Image Processing*. pp. 2609 – 2612.

Chen, S.C. et al., **2005**. Learning prosthetic vision: a virtual-reality study. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 13(3), pp.249–55.

Chen, S.C. et al., **2009**. Simulating prosthetic vision: I. Visual models of phosphenes. *Vision Research*, 49(12), pp.1493–1506.

Chow, A.Y. et al., **2001**. Implantation of silicon chip microphotodiode arrays into the cat subretinal space. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 9(1), pp.86–95.

Chow, A.Y. et al., **2004**. The artificial silicon retina microchip for the treatment of vision loss from retinitis pigmentosa. *Archives of Ophthalmology*, 122(4), pp.460–9.

Comaniciu, D. & Meer, P., **2002**. Mean shift: a robust approach toward feature space analysis. *Pattern Analysis and Machine Intelligence*, 24(5), pp.603–619.

Coulombe, J., Sawan, M. & Gervais, J.-F., **2007**. A highly flexible system for microstimulation of the visual cortex: design and implementation. *IEEE Transactions on Biomedical Circuits and Systems*, 1(4), pp.258–69.

Da Cruz, L. et al., **2013**. The Argus II epiretinal prosthesis system allows letter and word reading and long-term function in patients with profound vision loss. *The British Journal of Ophthalmology*, 97(5), pp.632–6.

D'Arsonval, A., **1896**. Dispositifs pour la mesure des courants alternatifs de toutes fréquences. *C. R. Soc. Biol.*, 2, pp.450–451.

Dagnelie, G., Barnett, D.G., et al., **2006**. Paragraph text reading using a pixelized prosthetic vision simulator: parameter dependence and task learning in free-viewing conditions. *Investigative Ophthalmology & Visual Science*, 47(3), pp.1241–50.

Dagnelie, G. et al., **2007**. Real and virtual mobility performance in simulated prosthetic vision. *Journal of neural engineering*, 4(1), pp.S92–101.

Dagnelie, G. et al., **2001**. Simulated prosthetic vision: Perceptual and performance measures. In *Vision Science and its Applications*. pp. 43–6.

Dagnelie, G., Walter, M. & Yang, L., **2006**. Playing checkers: detection and eye–hand coordination in simulated prosthetic vision. *Journal of Modern Optics*, 53(9), pp.1325–1342.

Delbeke, J. et al., **2002**. The microsystems based visual prosthesis for optic nerve stimulation. *Artificial Organs*, 26(3), pp.232–4.

Delbeke, J., Oozeer, M. & Veraart, C., **2003**. Position, size and luminosity of phosphenes generated by direct optic nerve stimulation. *Vision Research*, 43(9), pp.1091–1102.

Denis, G. et al., **2013**. Human faces detection and localization with simulated prosthetic vision. In *SIGCHI Conference on Human Factors in Computing Systems*. pp. 61–66.

Denis, G., Macé, M. & Jouffrais, C., **2014**. Simulated Prosthetic Vision: improving text accessibility with retinal prostheses. In *International Conference of the IEEE Engineering in Medicine and Biology Society*. pp. 1719–22.

Denis, G., Macé, M.J.-M. & Jouffrais, C., **2012**. Simulated prosthetic vision: object recognition and localization approach. In *International Conference on Neuroprosthetic Devices*. pp. 40–41.

Djilas, M. et al., **2011**. Three-dimensional electrode arrays for retinal prostheses: modeling, geometry optimization and experimental validation. *Journal of Neural Engineering*, 8(4), p.046020.

Dobelle, W.H., **2000**. Artificial vision for the blind by connecting a television camera to the visual cortex. *American Society for Artificial Internal Organs*, 46(1), pp.3–9.

Dobelle, W.H. & Mladejovsky, M.G., **1974**. Phosphenes produced by electrical stimulation of human occipital cortex, and their application to the development of a prosthesis for the blind. *The Journal of Physiology*, 243(2), pp.553–76.

Dorn, J.D. et al., **2012**. The Detection of Motion by Blind Subjects With the Epiretinal 60-Electrode (Argus II) Retinal Prosthesis. *Archives of Ophthalmology*, pp.1–7.

Dowling, J., **2009**. Current and future prospects for optoelectronic retinal prostheses. *Eye*, 23(10), pp.1999–2005.

Dowling, J.A., Maeder, A. & Boles, W., **2004**. Mobility enhancement and assessment for a visual prosthesis. In A. A. Amini & A. Manduca, eds. *Medical Imaging*. pp. 780–791.

Dramas, F., Thorpe, S.J. & Jouffrais, C., **2010**. Artificial Vision For The Blind: A Bio-Inspired Algorithm For Objects And Obstacles Detection. *International Journal of Image and Graphics*, 10(4), pp.531–544.

Duret, F. et al., **2006**. Object localization, discrimination, and grasping with the optic nerve visual prosthesis. *Restorative Neurology and Neuroscience*, 24(1), pp.31–40.

Eiber, C.D., Lovell, N.H. & Suaning, G.J., **2013**. Attaining higher resolution visual prosthetics: a review of the factors and limitations. *Journal of Neural Engineering*, 10(1), p.011002.

Epshtein, B., Ofek, E. & Wexler, Y., **2010**. Detecting text in natural scenes with stroke width transform. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Ieee, pp. 2963–2970.

Felzenszwalb, P.F. & Huttenlocher, D.P., **2004**. Efficient Graph-Based Image Segmentation. *International Journal of Computer Vision*, 59(2), pp.167–181.

Feng, D. & McCarthy, C., **2013**. Enhancing scene structure in prosthetic vision using iso-disparity contour perturbance maps. In *International Conference of the IEEE Engineering in Medicine and Biology Society*. pp. 5283–6.

Fernández, E. et al., **2005**. Development of a cortical visual neuroprosthesis for the blind: the relevance of neuroplasticity. *Journal of Neural Engineering*, 2(4), pp.R1–12.

Fink, W., Tarbell, M. & Sivaprakasam, M., **2005**. Image Processing and Interface for Retinal Visual Prostheses. In *International Symposium on Circuits and Systems*. IEEE, pp. 2927–2930.

Foerster, O., **1929**. Beitrage zur Pathophysiologie der Sehbahn und der Sehsphare. *J. Psychol. Neurol.*, 39, pp.463–485.

Greenwald, S.H. et al., **2009**. Brightness as a function of current amplitude in human retinal electrical stimulation. *Investigative Ophthalmology & Visual Science*, 50(11), pp.5017–25.

Guo, H., Qin, R., et al., **2010**. Configuration-based processing of phosphene pattern recognition for simulated prosthetic vision. *Artificial Organs*, 34(4), pp.324–30.

Guo, H., Wang, Y., et al., **2010**. Object recognition under distorted prosthetic vision. *Artificial Organs*, 34(10), pp.846–56.

Harris, C. & Stephens, M., **1988**. A combined corner and edge detector. In *Alvey Vision Conference*. pp. 147–151.

Hartley, R. & Zisserman, A., **2003**. *Multiple view geometry in computer vision*, Cambridge University Press.

Hayes, J.S. et al., **2003**. Visually guided performance of simple tasks using simulated prosthetic vision. *Artificial Organs*, 27(11), pp.1016–28.

He, X., Kim, J. & Barnes, N., **2012**. An face-based visual fixation system for prosthetic vision. *International Conference of the IEEE Engineering in Medicine and Biology Society*, 2012, pp.2981–4.

Horne, L. et al., **2012**. Image segmentation for enhancing symbol recognition in prosthetic vision. *International Conference of the IEEE Engineering in Medicine and Biology Society*, 2012, pp.2792–5.

Hornig, R., Zehnder, T. & Velikay-Parel, M., **2008**. The IMI retinal implant system. In *Artificial Sight*. pp. 111–128.

Horowitz, S.L. & Pavlidis, T., **1976**. Picture Segmentation by a Tree Traversal Algorithm. *Journal of the ACM*, 23(2), pp.368–388.

Hu, J. et al., **2013**. Recognition of Similar Objects Using Simulated Prosthetic Vision. *Artificial Organs*, 38(2), pp.159–67.

Humayun, M.S. et al., **2012**. Interim results from the international trial of Second Sight's visual prosthesis. *Ophthalmology*, 119(4), pp.779–88.

Humayun, M.S. et al., **1999**. Pattern electrical stimulation of the human retina. *Vision research*, 39(15), pp.2569–76.

Humayun, M.S. et al., **1996**. Visual perception elicited by electrical stimulation of retina in blind humans. *Archives of Ophthalmology*, 114(1), pp.40–6.

Humayun, M.S. et al., **2003**. Visual perception in a blind subject with a chronic microelectronic retinal prosthesis. *Vision Research*, 43(24), pp.2573–2581.

Iannizzotto, G. et al., **2005**. Badge3D for Visually Impaired. In *Computer Vision and Pattern Recognition*. IEEE, p. 29.

Itti, L. & Koch, C., **2001**. Computational modelling of visual attention. *Nature reviews. Neuroscience*, 2(3), pp.194–203.

Jafri, R., Ali, S.A., Arabnia, H.R., et al., **2013**. Computer vision-based object recognition for the visually impaired in an indoors environment: a survey. *The Visual Computer*.

Jafri, R., Ali, S.A. & Arabnia, H.R., **2013**. Computer Vision-based Object Recognition for the Visually Impaired Using Visual Tags. In *International Conference on Image Processing, Computer Vision, and Pattern Recognition*. pp. 400–406.

Ji, Z., Wang, J. & Su, Y., **2009**. Text detection in video frames using hybrid features. In *International Conference on Machine Learning and Cybernetics*. IEEE, pp. 318–322.

Jouffrais, C., **2011**. *Les nouvelles technologies au service de la cognition spatiale des déficients visuels*. Université of Toulouse.

Jung, C., Liu, Q. & Kim, J., **2008**. A new approach for text segmentation using a stroke filter. *Signal Processing*, 88(7), pp.1907–1916.

Kim, E.T. et al., **2009**. Feasibility of microelectrode array (MEA) based on silicone-polyimide hybrid for retina prosthesis. *Investigative Ophthalmology & Visual Science*, 50(9), pp.4337–41.

Kiral-Kornek, F.I., Savage, C.O. & Grayden, D.B., **2011**. The focus of attention under phosphenated vision through retinal implants. In *International Conference on Intelligent Sensors, Sensor Networks and Information Processing*. IEEE, pp. 113–118.

Klauke, S. et al., **2011**. Stimulation with a wireless intraocular epiretinal implant elicits visual percepts in blind humans. *Investigative Ophthalmology & Visual Science*, 52(1), pp.449–455.

Krause, F., **1924**. Die Sehbahn in Chirurgischer Beziehung und die Faradische Reizung des Sehzentrums. *Journal of Molecular Medicine*, 3(28), pp.1260–1265.

Krause, F. & Schum, H., **1931**. Die epileptischen Erkrankungen. *Neue Deutsche Chirurgie*, 49a.

LeRoy, C., **1755**. Mémoire où l'on rend compte de quelques tentatives que l'on a faites pour guérir plusieurs maladies par l'Électricité. In A. des Sciences, ed. *Histoire de l'Académie Royale des Sciences*. Paris, pp. 60–98.

Li, S. et al., **2011**. Image recognition with a limited number of pixels for visual prostheses design. *Artificial Organs*, 36(3), pp.266–274.

Li, Y., McCarthy, C. & Barnes, N., **2012**. On just noticeable difference for bionic eye. *International Conference of the IEEE Engineering in Medicine and Biology Society*, 2012, pp.2961–4.

Liang, J., Doermann, D. & Li, H., **2005**. Camera-based analysis of text and documents: a survey. *International Journal of Document Analysis and Recognition*, 7(2-3), pp.84–104.

Lichtsteiner, P., Posch, C. & Delbruck, T., **2008**. A 128× 128 120 dB 15 μs Latency Asynchronous Temporal Contrast Vision Sensor. *IEEE Journal of Solid-State Circuits*, 43(2), pp.566–576.

Lieby, P. et al., **2011**. Substituting depth for intensity and real-time phosphene rendering: Visual navigation under low vision conditions. In *International Conference of the IEEE Engineering in Medicine and Biology Society*. pp. 8017–20.

Lillywhite, K. et al., **2013**. A feature construction method for general object recognition. *Pattern Recognition*, 46(12), pp.3300–3314.

Loudin, J.D. et al., **2007**. Optoelectronic retinal prosthesis: system design and performance. *Journal of Neural Engineering*, 4(1), pp.S72–84.

Lowe, D., **1999**. Object recognition from local scale-invariant features. In *IEEE International Conference on Computer Vision*. IEEE, pp. 1150–1157 vol.2.

Löwenstein, K. & Borchardt, M., **1918**. Symptomatologie und elektrische Reizung bei einer Schußverletzung des Hinterhauptlappens. *Deutsche Zeitschrift für Nervenheilkunde*, 58(3-6), pp.264–292.

Lu, Y. et al., **2013**. Recognition of objects in simulated irregular phosphene maps for an epiretinal prosthesis. *Artificial Organs*, 38(2), pp.10–20.

Lui, W.L.D. et al., **2012**. Transformative Reality: improving bionic vision with robotic sensing. *International Conference of the IEEE Engineering in Medicine and Biology Society*, 2012, pp.304–7.

Matas, J. et al., **2004**. Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 22(10), pp.761–767.

McCarthy, C. & Barnes, N., **2012**. Time-to-contact maps for navigation with a low resolution visual prosthesis. *International Conference of the IEEE Engineering in Medicine and Biology Society*, 2012, pp.2780–3.

McCarthy, C., Barnes, N. & Lieby, P., **2011**. Ground surface segmentation for navigation with a low resolution visual prosthesis. In *International Conference of the IEEE Engineering in Medicine and Biology Society*. pp. 4457–60.

McCarthy, C., Feng, D. & Barnes, N., **2013**. Augmenting intensity to enhance scene structure in prosthetic vision. In *International Conference on Multimedia and Expo Workshops*. IEEE, pp. 1–6.

Meffin, H., **2013**. What limits spatial perception with retinal implants? In *International Conference on Image Processing*. pp. 1545 – 1549.

Meijer, P.B., **1992**. An experimental system for auditory image representations. *IEEE Transactions on Bio-medical Engineering*, 39(2), pp.112–21.

Merino-Gracia, C., Lenc, K. & Mirmehdi, M., **2012**. A head-mounted device for recognizing text in natural scenes. In *International Workshop on Camera-Based Document Analysis and Recognition*. pp. 29–41.

Mohammadi, H.M., Ghafar-Zadeh, E. & Sawan, M., **2012**. An image processing approach for blind mobility facilitated through visual intracortical stimulation. *Artificial Organs*, 36(7), pp.616–28.

Morillas, C. et al., **2007**. A neuroengineering suite of computational tools for visual prostheses. *Neurocomputing*, 70(16-18), pp.2817–2827.

Morimoto, T. et al., **2011**. Chronic implantation of newly developed suprachoroidal-transretinal stimulation prosthesis in dogs. *Investigative Ophthalmology & Visual Science*, 52(9), pp.6785–92.

Nanduri, D. et al., **2012**. Frequency and amplitude modulation have different effects on the percepts elicited by retinal stimulation. *Investigative Ophthalmology & Visual Science*, 53(1), pp.205–14.

Naumann, J., **2012**. *Search for Paradise: A Patient's Account of the Artificial Vision Experiment*, Xlibris Corporation.

Nistér, D. & Stewénius, H., **2008**. Linear time maximally stable extremal regions. In *European Conference on Computer Vision*. pp. 183–196.

Normann, R.A. et al., **1999**. A neural interface for a cortical vision prosthesis. *Vision Research*, 39(15), pp.2577–87.

Normann, R.A. et al., **2009**. Toward the development of a cortically based visual neuroprosthesis. *Journal of Neural Engineering*, 6(3), p.035001.

Palanker, D. V et al., **2005**. Design of a high-resolution optoelectronic retinal prosthesis. *Journal of Neural Engineering*, 2(1), pp.S105–20.

Parikh, N. et al., **2013**. Performance of visually guided tasks using simulated prosthetic vision and saliency-based cues. *Journal of Neural Engineering*, 10(2), p.026017.

Parikh, N., Itti, L. & Weiland, J., **2010**. Saliency-based image processing for retinal prostheses. *Journal of Neural Engineering*, 7(1), p.16006.

Pelayo, F. et al., **2003**. Cortical visual neuro-prosthesis for the blind: retina-like software/hardware preprocessor. In *International IEEE EMBS Conference on Neural Engineering*. IEEE, pp. 150–153.

Pelayo, F. et al., **2004**. Translating image sequences into spike patterns for cortical neuro-stimulation. *Neurocomputing*, 58-60, pp.885–892.

Penfield, W. & Jasper, H., **1954**. Epilepsy and the Functional Anatomy of the Human Brain. *JAMA: The Journal of the American Medical Association*, 155(1), p.86.

Penfield, W. & Rasmussen, T., **1950**. *The cerebral cortex of man* MacMillan., New York.

Pérez Fornos, A. et al., **2008**. Simulation of artificial vision: IV. Visual information required to achieve simple pointing and manipulation tasks. *Vision Research*, 48(16), pp.1705–18.

Pérez Fornos, A. et al., **2012**. Temporal properties of visual perception on electrical stimulation of the retina. *Investigative Ophthalmology & Visual Science*, 53(6), pp.2720–31.

Pérez Fornos, A., Sommerhalder, J. & Pelizzone, M., **2011**. Reading with a simulated 60-channel implant. *Frontiers in Neuroscience*, 5, p.57.

Prewitt, J., **1970**. Object enhancement and extraction. In *Picture Processing and Psychopictorics*. pp. 75–150.

Rheede, J.J. Van, Kennard, C. & Hicks, S.L., **2010**. Simulating prosthetic vision: Optimizing the information content of a limited visual display. *Journal of Vision*, 10(14), pp.1–14.

Rizzo III, J.F. et al., **2003a**. Methods and perceptual thresholds for short-term electrical stimulation of human retina with microelectrode arrays. *Investigative Ophthalmology & Visual Science*, 44(12), p.5355.

Rizzo III, J.F. et al., **2003b**. Perceptual efficacy of electrical stimulation of human retina with a microelectrode array during short-term surgical trials. *Investigative Ophthalmology & Visual Science*, 44(12), pp.5362–5369.

Rizzo III, J.F., **2011**. Update on retinal prosthetic research: the Boston Retinal Implant Project. *Journal of Neuro-ophthalmology*, 31(2), pp.160–8.

Roessler, G. et al., **2009**. Implantation and explantation of a wireless epiretinal retina implant device: observations during the EPIRET3 prospective clinical trial. *Investigative Ophthalmology & Visual Science*, 50(6), pp.3003–8.

Rogister, P. et al., **2012**. Asynchronous event-based binocular stereo matching. *IEEE Transactions on Neural Networks and Learning Systems*, 23(2), pp.347–53.

Rosten, E. & Drummond, T., **2006**. Machine Learning for High-Speed Corner Detection. In A. Leonardis, H. Bischof, & A. Pinz, eds. *European Conference on Computer Vision*. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 430–443.

Rublee, E. et al., **2011**. ORB: An efficient alternative to SIFT or SURF. In *International Conference on Computer Vision*. IEEE, pp. 2564–2571.

Sakaguchi, H. et al., **2009**. Artificial vision by direct optic nerve electrode (AV-DONE) implantation in a blind patient with retinitis pigmentosa. *Artificial Organs*, 12(3), pp.206–9.

Sampaio, E., Maris, S. & Bach-y-Rita, P., **2001**. Brain plasticity: “visual” acuity of blind persons via the tongue. *Brain Research*, 908(2), pp.204–207.

Scharstein, D. & Szeliski, R., **2002**. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Visionomputer vision*, 47(1-3), pp.7–42.

Schauerte, B. & Martinez, M., **2012**. An assistive vision system for the blind that helps find lost things. In *International Conference on Computers Helping People with Special Needs*. pp. 566–572.

Schmidt, E.M. et al., **1996**. Feasibility of a visual prosthesis for the blind based on intracortical micro stimulation of the visual cortex. *Brain*, 119(2), pp.507–522.

Second Sight Medical Products, **2013**. *Argus II Retinal Prosthesis System Surgeon Manual*,

Sharma, N., Pal, U. & Blumenstein, M., **2012**. Recent advances in video based document processing: A review. In *International Workshop on Document Analysis Systems*. Ieee, pp. 63–68.

Shi, J. & Malik, J., **2000**. Normalized cuts and image segmentation. *Pattern Analysis and Machine Intelligence*, 22(8), pp.888–905.

Shire, D.B. et al., **2009**. Development and implantation of a minimally invasive wireless subretinal neurostimulator. *IEEE Transactions on Bio-medical Engineering*, 56(10), pp.2502–11.

Shivakumara, P., Huang, W. & Tan, C.L., **2008**. An Efficient Edge Based Technique for Text Detection in Video Frames. In *IAPR International Workshop on Document Analysis Systems*. IEEE, pp. 307–314.

Shivakumara, P., Phan, T.Q. & Tan, C.L., **2011**. A Laplacian approach to multi-oriented text detection in video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(2), pp.412–9.

Shivakumara, P., Phan, T.Q. & Tan, C.L., **2009**. A Robust Wavelet Transform Based Technique for Video Text Detection. In *International Conference on Document Analysis and Recognition*. IEEE, pp. 1285–1289.

Sivic, J. & Zisserman, A., **2009**. Efficient visual search of videos cast as text retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(4), pp.591–606.

Sobel, I., **1970**. *Camera models and machine perception*. Stanford University.

Sommerhalder, J. et al., **2003**. Simulation of artificial vision: I. Eccentric reading of isolated words, and perceptual learning. *Vision Research*, 43(3), pp.269–83.

Sommerhalder, J. et al., **2004**. Simulation of artificial vision: II. Eccentric reading of full-page text and the learning of this task. *Vision Research*, 44(14), pp.1693–706.

Srivastava, N.R. et al., **2007**. Estimating phosphene maps for psychophysical experiments used in testing a cortical visual prosthesis device. In *International IEEE EMBS Conference on Neural Engineering*. pp. 130–133.

Srivastava, N.R. & Troyk, P.R., **2005**. A proposed intracortical visual prosthesis image processing system. In *International Conference of the IEEE Engineering in Medicine and Biology Society*. pp. 5264–7.

Stacey, A., Li, Y. & Barnes, N., **2011**. A salient information processing system for bionic eye with application to obstacle avoidance. In *International Conference of the IEEE Engineering in Medicine and Biology Society*. pp. 5116–9.

Stingl, K. et al., **2013**. Artificial vision with wirelessly powered subretinal electronic implant alpha-IMS. *Proceedings of the Royal Society B: Biological Sciences*, 280(1757), pp.20130077–20130077.

Sudol, J. et al., **2010**. Looktel—A comprehensive platform for computer-aided visual assistance. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, pp. 73–80.

Sun, J. et al., **2011**. Spatiotemporal properties of multipeaked electrically evoked potentials elicited by penetrative optic nerve stimulation in rabbits. *Investigative Ophthalmology & Visual Science*, 52(1), pp.146–54.

Szeliski, R., **2011**. *Computer Vision*, London: Springer London.

Taigman, Y. et al., **2014**. DeepFace: Closing the Gap to Human-Level Performance in Face Verification. In *Conference on Computer Vision and Pattern Recognition*.

Tehovnik, E.J. & Slocum, W.M., **2007**. Phosphene induction by microstimulation of macaque V1. *Brain Research Reviews*, 53(2), pp.337–43.

Thompson, R.W. et al., **2003**. Facial recognition using simulated prosthetic pixelized vision. *Investigative Ophthalmology & Visual Science*, 44(11), pp.5035–5042.

Troyk, P.R. et al., **2003**. A model for intracortical visual prosthesis research. *Artificial Organs*, 27(11), pp.1005–15.

Tsai, D. et al., **2009**. A wearable real-time image processor for a vision prosthesis. *Computer Methods and Programs in Biomedicine*, 95(3), pp.258–69.

Veraart, C. et al., **1998**. Visual sensations produced by optic nerve stimulation using an implanted self-sizing spiral cuff electrode. *Brain Research*, 813(1), pp.181–186.

Vergnieux, V., Macé, M.J.-M. & Jouffrais, C., **2012**. Spatial navigation with a simulated prosthetic vision in a virtual environment. In *Proceedings of the NeuroComp/KEOpS'12 workshop*. pp. 1–6.

Vergnieux, V., Macé, M.J.-M. & Jouffrais, C., **2014**. Wayfinding with Simulated Prosthetic Vision: Performance comparison with regular and structured-enhanced renderings. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. pp. 2585–88.

Villalobos, J. et al., **2013**. A wide-field suprachoroidal retinal prosthesis is stable and well tolerated following chronic implantation. *Investigative Ophthalmology & Visual Science*, 54(5), pp.3751–62.

Vincent, L. & Soille, P., **1991**. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(6), pp.583–598.

Viola, P. & Jones, M.J., **2001**. Rapid object detection using a boosted cascade of simple features. In *Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE Comput. Soc, pp. I–511–I–518.

Viola, P. & Jones, M.J., **2004**. Robust Real-Time Face Detection. *International Journal of Computer Vision*, 57(2), pp.137–154.

Vurro, M. et al., **2006**. Simulation and assessment of bioinspired visual processing system for epi-retinal prostheses. *International Conference of the IEEE Engineering in Medicine and Biology Society*, 1, pp.3278–81.

Wang, J., Wu, X., et al., **2014**. Face recognition in simulated prosthetic vision: face detection-based image processing strategies. *Journal of Neural Engineering*, 11(4), p.046009.

Wang, J., Lu, Y., et al., **2014**. Moving Objects Recognition under Simulated Prosthetic Vision Using Background-subtraction-based Image Processing Strategies. *Information Sciences*, 277, pp.512–524.

Wang, L., Yang, L. & Dagnelie, G., **2008**. Virtual wayfinding using simulated prosthetic vision in gaze-locked viewing. *Optometry and vision science*, 85(11), pp.1057–1063.

Weiland, J.D. et al., **2012**. Smart image processing system for retinal prosthesis. *International Conference of the IEEE Engineering in Medicine and Biology Society*, 2012, pp.300–3.

WHO, **2012**. *Visual Impairment and blindness Fact Sheet N° 282*, World Health Organization.

Wilke, R.G.H. et al., **2011**. Electric crosstalk impairs spatial resolution of multi-electrode arrays in retinal implants. *Journal of Neural Engineering*, 8(4), p.046016.

Yanai, D. et al., **2007**. Visual performance using a retinal prosthesis in three subjects with retinitis pigmentosa. *American Journal of Ophthalmology*, 143(5), pp.820–827.

Ye, Q. et al., **2007**. Text detection and restoration in natural scene images. *Journal of Visual Communication and Image Representation*, 18(6), pp.504–513.

Zhang, C. & Zhang, Z., **2010**. *A survey of recent advances in face detection*,

Zhao, M., Li, S. & Kwok, J., **2010**. Text detection in images using sparse representation with discriminative dictionaries. *Image and Vision Computing*, 28(12), pp.1590–1599.

Zhao, Y. et al., **2011**. Chinese character recognition using simulated phosphene maps. *Investigative Ophthalmology & Visual Science*, 52(6), pp.3404–12.

Zhao, Y. et al., **2010**. Image processing based recognition of images with a limited number of pixels using simulated prosthetic vision. *Information Sciences*, 180(16), pp.2915–2924.

Zhou, D.D., Dorn, J.D. & Greenberg, R.J., **2013**. The Argus® II retinal prosthesis system: An overview. In *International Conference on Multimedia and Expo Workshops*. pp. 1–6.

Zrenner, E. et al., **1999**. Can subretinal microphotodiodes successfully replace degenerated photoreceptors? *Vision Research*, 39(15), pp.2555–2567.

Zrenner, E. et al., **2011**. Subretinal electronic chips allow blind patients to read letters and combine them to words. *Proceedings of the Royal Society B: Biological Sciences*, 278(1711), pp.1489–97.

Zrenner, E. et al., **2009**. Subretinal Microelectrode Arrays Implanted Into Blind Retinitis Pigmentosa Patients Allow Recognition of Letters and Direction of Thin Stripes. In O. Dössel & W. C. Schlegel, eds. *International Federation for Medical and Biological Engineering*. IFMBE Proceedings. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 444–447.

Zrenner, E., **2002**. Will retinal implants restore vision? *Science*, 295(5557), pp.1022–5.

Zrenner, E., Wilke, R. & Zabel, T., **2007**. Psychometric analysis of visual sensations mediated by subretinal microelectrode arrays implanted into blind retinitis pigmentosa patients. *Investigative Ophthalmology & Visual Science*, 48, p.659.

ANNEXE 1 - QUESTIONNAIRE POST EXPERIENCE LOCALISATION D'OBJETS

Questionnaire: Localisation et Scoreboard

* Required

N° sujet *

Question sur le dispositif

Comment qualifieriez-vous le port du casque de réalité virtuelle ? (de zéro à plusieurs réponses possibles) *

- ☐ Douloureux
- ☐ Inconfortable
- ☐ Trop lourd
- ☐ Gênant dans les mouvements de la main
- ☐ Gênant dans les mouvements de la tête
- ☐ Gênant mais supportable
- ☐ Fantastique!
- ☐ Other:

Comment qualifieriez-vous l'asservissement de la matrice à la position des yeux ? (de zéro à plusieurs réponses possibles) *

- ☐ Inutilisable
- ☐ Très perturbant
- ☐ Inconfortable
- ☐ Intéressant
- ☐ Pas gênant
- ☐ Amusant
- ☐ Other:

Notez la difficulté de saisie des objets en approche par localisation *

	Très facile	Facile	Moyenne	Difficile	Très difficile
Agrafeuse	○	○	○	○	○
Café	○	○	○	○	○
Chargeur	○	○	○	○	○
Ciseaux	○	○	○	○	○
Marqueur	○	○	○	○	○
Nutella	○	○	○	○	○
Tasse	○	○	○	○	○
Téléphone	○	○	○	○	○

Remarques:

Notez la difficulté de saisie des objets en approche scoreboard indépendamment de la matrice utilisée *

	Très facile	Facile	Moyenne	Difficile	Très difficile
Agrafeuse	○	○	○	○	○
Café	○	○	○	○	○
Chargeur	○	○	○	○	○
Ciseaux	○	○	○	○	○
Marqueur	○	○	○	○	○
Nutella	○	○	○	○	○
Tasse	○	○	○	○	○
Téléphone	○	○	○	○	○

Remarques:

Notez la difficulté d'utilisation des matrices d'affichages, indépendamment des objets à saisir *

	Très facile	Facile	Moyenne	Difficile	Très difficile
Localisation 6*10	○	○	○	○	○
Scoreboard 6*10	○	○	○	○	○
Scoreboard 15*18	○	○	○	○	○
Scoreboard 32*38	○	○	○	○	○

Remarques:

Selon vous, à combien de positions les objets pouvaient-ils se trouver?

- ○ Un nombre aléatoire
- ○ Un nombre précis, difficile à approximer
- ○ Other:

Remarques sur votre utilisation du dispositif en approche par localisation

Décrivez brièvement votre stratégie pour repérer et saisir un objet

Avez-vous jugé la non-visibilité de la main gênante ? Pourquoi ?

Remarques sur votre utilisation du dispositif en approche scoreboard

Décrivez brièvement votre stratégie pour repérer et saisir un objet

Remarques générales

Avez-vous des remarques, critiques ou propositions à faire sur le dispositif ?

Submit

ANNEXE 2 - QUESTIONNAIRE POST EXPERIENCE LOCALISATION DE TEXTE

Experience de localisation de texte

* Required

N° sujet *

Le dispositif

Comment qualifieriez-vous le port du casque de réalité virtuelle ? (de zéro à plusieurs réponses possibles)

- ☐ Douloureux
- ☐ Inconfortable
- ☐ Trop lourd
- ☐ Gênant
- ☐ Léger
- ☐ Agréable

La tâche, l'approche "Scorboard"

Quelle est la difficulté de la tâche dans l'approche "Scoreboard" avec la petite matrice?

- ○ Très facile
- ○ Facile
- ○ Moyenne
- ○ Difficile
- ○ Très difficile

Remarques

Quelle est la difficulté de la tâche dans l'approche "Scoreboard" avec la grande matrice?

- ○ Très facile
- ○ Facile
- ○ Moyenne
- ○ Difficile
- ○ Très difficile

Remarques

Décrivez brièvement votre stratégie pour localiser le texte dans l'image

La tâche, l'approche "Scoreboard" augmenté

Quelle est la difficulté de la tâche dans l'approche "Scoreboard" augmenté avec la petite matrice?

- Très facile
- Facile
- Moyenne
- Difficile
- Très difficile

Remarques

Quelle est la difficulté de la tâche dans l'approche "Scoreboard" augmenté avec la grande matrice?

- Très facile
- Facile
- Moyenne
- Difficile
- Très difficile

Remarques

Décrivez brièvement votre stratégie pour localiser le texte dans l'image

Remarques générales

Avez-vous des remarques, critiques ou propositions à faire sur le dispositif ?

Submit

FIGURES

Figure 1-1 L'implant cochléaire. 14
Figure 1-2 Anatomie et physiologie de l'œil humain. 16
Figure 1-3 Les différents relais visuels. 17
Figure 1-4 Prototype de neuroprothèse corticale de Brindley et Lewin. 19
Figure 1-5 Organisation rétinotopique théorique de l'aire visuelle V1. 21
Figure 1-6 Neuroprothèse visuelle développée par William Dobelle. 22
Figure 1-7 Matrice UEA (Utah Electrode Array). 25
Figure 1-8 Neuroprothèse intra-corticale imaginée par PolySTIM. 26
Figure 1-9 Neuroprothèse intra-corticale envisagée par le Monash Vision Group. 26
Figure 1-10 Carte des phosphènes évoqués par stimulation du nerf optique. 28
Figure 1-11 Les catégories d'implants rétiniens. 30
Figure 1-12 L'implant épirétinien développé par la société Second Sight. 31
Figure 1-13 L'implant de seconde génération du Boston Retinal Implant Project. 33
Figure 1-14 L'implant sous-rétinien développé par la société Retina Implant AG. 36
Figure 1-15 L'implant sous-rétinien développé à l'université de Stanford. 38
Figure 1-16 Prévision des résolutions d'implants. 44
Figure 2-1 Exemples de filtres. 51
Figure 2-2 Exemples de points et de régions d'intérêt. 52
Figure 2-3 Exemple de segmentation d'image. 54
Figure 2-4 Détection d'objets reposant sur SIFT. 56
Figure 2-5 Principe général du module de traitement de l'image. 60
Figure 2-6 Les traitements d'images effectués dans l'implant de la société IMI Gmbh. 62
Figure 2-7 Schéma du Video Processing Unit (VPU) du système Argus II. 63
Figure 2-8 Transformative Reality. 66
Figure 2-9 Système de fixation automatique de visages. 66
Figure 2-10 L'approche scoreboard. 67
Figure 2-11 L'approche par localisation de points d'intérêt. 69
Figure 3-1 Paramètres d'acquisition, de traitement et de restitution d'image pour l'Argus II. 74
Figure 3-2 L'implant Argus II. 74
Figure 3-3 Exemples de carte de phosphènes. 78
Figure 3-4 Architecture générale du simulateur de vision prothétique. 80
Figure 3-5 Exemples d'implants simulés. 84
Figure 4-1 Vision prothétique avec les quatre matrices utilisées. 89
Figure 4-2 Dispositif. 91
Figure 4-3 Tâche de localisation d'objet. 92
Figure 4-4 Architecture du simulateur pour l'expérience de localisation d'objets. 94
Figure 4-5 Précision et temps de mouvement. 96
Figure 4-6 Moyenne des temps de mouvement et de la précision. 97
Figure 4-7 Moyenne des temps de mouvement (A) et de la précision (B). 98
Figure 4-8 Objets utilisés dans l'expérience de localisation. 101
Figure 4-9 Rendu basé sur l'approche scoreboard. 102
Figure 4-10 Architecture du système. 104
Figure 4-11 Moyenne des temps de mouvement et des taux de réussite. 106
Figure 5-1 Disposition de l'environnement. 118
Figure 5-2 Scoreboard vs Scoreboard Augmenté vs Localisation. 119
Figure 5-3 Simulateur utilisé dans le cadre de l'expérience de localisation de visages. 121

Figure 5-4 Taux de réussite en fonction du temps de réponse pour la condition scoreboard augmenté pour 4 sujets. 123
Figure 5-5 Taux de réussite et temps de réponse. 123
Figure 5-6 Précision de pointage. 124
Figure 5-7 Taux de réussite et précision. 125
Figure 6-1 Disposition pour l'expérience de localisation de texte. 135
Figure 6-2 Architecture du simulateur pour l'expérience de localisation de texte. 137
Figure 6-3 Précision vs. temps de réponse. 139
Figure 6-4 Précision et temps de réponse. 139
Figure 6-5 Précision par condition et taille de caractère. 140
Figure Disc-1 Rendus visuels et actions disponibles dans le logiciel servant à l'interaction avec la neuroprothèse. 153
Figure Disc-2 Exemples d'interactions possibles avec la neuroprothèse visuelle. 153

TABLEAUX

Tableau 2-1 Les traitements d'images effectués dans les neuroprothèses visuelles. 61
Tableau 2-2 Les différents traitements d'images imaginés en simulation.....64
Tableau 3-1 Principales caractéristiques techniques d'une neuroprothèse visuelle. 73
Tableau 3-2 Principales caractéristiques d'un phosphène pour les trois types d'implant. 75
Tableau 3-3 Caractéristiques des deux caméras utilisées. 80
Tableau 3-4 Caractéristiques des deux casques de réalité virtuelle utilisés. 83
Tableau 4-1 Les trois implants rétiniens simulés pour cette expérience.....102
Tableau 6-1 Comparaisons paire à paire. 139