THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par *Université de Toulouse*
Discipline ou spécialité : *Informatique*

Présentée et soutenue par *Florian Dramas*
Le *11/06/2010*

Titre : Localisation d'objets pour les non-voyants :

augmentation sensorielle et neuroprothèse

JURY
Edwige Pissaloux - Rapporteur
José Luis Gonzalez-Mora - Rapporteur
Christophe Jouffrais – Encadrant de thèse
Simon Thorpe – Directeur de thèse
Régine André-Obrecht – Directeur de thèse
Annelies Braffort - Examinatrice

Ecole doctorale : *MITT*
Unité de recherche : *IRIT*
Directeur(s) de Thèse : *Simon Thorpe, Régine André-Obrecht*
Encadrant : Christophe Jouffrais

Résumé :

L'organisation Mondiale de la Santé estime qu'il y a environ 47 millions de personnes aveugles dans le monde. Les difficultés éprouvées par les non-voyants dans leur vie quotidienne peuvent être classées dans quatre domaines principaux : l'accès à l'information écrite, le déplacement, l'orientation et la reconnaissance/localisation d'objets. C'est sur la conception et le développement d'un système de suppléance permettant aux non-voyants de reconnaître et de localiser des objets que porte cette thèse.

La cane blanche et le chien guide répondent en partie aux difficultés des non-voyants. Ces aides sont particulièrement bien acceptées et démocratisées mais elles ne répondent qu'aux seules difficultés liées au déplacement. Les progrès dans le domaine des sciences et technologies de l'information et de la communication ont permis depuis les années 1970 de concevoir de nouvelles aides électroniques pour les non-voyants afin de pallier les difficultés auxquelles le chien et la cane ne répondent pas. L'état de l'art sur les systèmes de suppléance électroniques pour les non-voyants fait apparaître deux catégories d'aides : les systèmes de substitution sensorielle et les systèmes d'augmentation sensorielle. Les premiers capturent une image basse résolution de la scène visuelle en la transformant pour être restituée dans autre modalité sensorielle (tactile ou auditive). Ces systèmes expérimentaux rendent l'interprétation de la scène visuelle très difficile et ne sont donc que rarement utilisés au quotidien. Les systèmes d'augmentation sensorielle sont plus utilisés. Ils augmentent un canal sensoriel (tactile ou auditif) en restituant une faible quantité d'information identifiée comme pertinente pour améliorer l'autonomie des personnes non-voyantes. C'est cette dernière approche qui a été développée dans cette thèse.

Une des fonctions principales du système visuel humain est de localiser des objets dans l'espace. Cette faculté est essentielle pour se déplacer, s'orienter en se représentant l'espace mais aussi pour atteindre et saisir des objets. L'objectif du système de suppléance présenté dans cette thèse est de restaurer cette fonction centrale du système visuel qu'est la reconnaissance et la localisation d'objets. Pour cela, nous nous sommes intéressés à l'évaluation d'un algorithme de vision artificielle rapide et robuste permettant de reconnaître et localiser des objets en trois dimensions. La restitution de cette information de position sur un objet a été envisagée selon deux modes de restitution : auditive dans un système d'augmentation sensorielle et « visuelle » pour la simulation d'une neuroprothèse corticale. Dans ce travail, nous avons étudié les capacités humaines à localiser un son dans l'espace dans le but de concevoir une interface de restitution exploitant au mieux ces capacités. Nous avons dans un premier temps évalué la précision de localisation d'une source sonore dans deux expériences de psychophysique concernant l'espace proche. Un deuxième mode de restitution expérimental a été étudié : par stimulation électrique du système visuel humain. Pour cela, un modèle de neuroprothèse a été développé sur la base d'informations issues des recherches en neurosciences. Il a été évalué dans une étude préliminaire de navigation vers des objets dans un espace virtuel.

La restauration de la reconnaissance et de la localisation d'objets permet de répondre au besoin en mobilité des non-voyants par l'aide au déplacement, l'aide à la navigation et la localisation d'objets. Les perspectives de ces travaux (dont certaines ont été reprises dans le projet ANR NAVIG) permettent d'envisager un outil de suppléance pour l'aide à la navigation pour les non-voyants, en s'appuyant sur un système de géolocalisation pour la navigation, la vision artificielle pour reconnaître des objets dans la scène visuelle et un système de sonification pour restituer ces informations.

Table des matières

Introduction Générale

Selon l'Organisation Mondiale de la Santé, 314 millions de personnes sont atteintes de déficience visuelle dans le monde, 15% d'entre elles sont aveugles. Si à l'échelle mondiale, la majorité des maladies engendrant la perte de la vision pourraient être évitées, il n'en demeure pas moins une augmentation du nombre des déficients visuels dans tous les pays, qu'ils soient industrialisés ou non, du fait de l'accroissement de la durée de vie et de la proportion de non-voyants plus importante parmi les personnes âgées. La cécité de l'enfant est un problème important, et plus particulièrement dans les pays en voie de développement du fait du manque de diagnostic précoce qui permettrait d'éviter de nombreux cas. Ainsi, seulement 13% des personnes déficientes visuelles vivent dans des pays développés, dans lesquels le diagnostique est effectué plus précocement engendrant un traitement plus rapide.

Ces statistiques montrent toute l'importance du diagnostic précoce ainsi que l'impact des recherches actuelles sur son traitement. La France a recensé environ 1,7 millions de personnes atteintes de déficience visuelle, dont 11% de personnes aveugles. Les pouvoirs publics se sont beaucoup intéressés ces dernières années aux différents aspects de la vie d'une personne handicapée visuelle, du diagnostic à l'accessibilité. De nombreux projets ont vu le jour pour faciliter la mobilité des personnes en situation de handicap ainsi que leur accès à l'information pour une meilleure insertion dans la société. Ainsi de nombreux projets de recherche ont émergé ces dernières années, la législation s'est aussi intéressée au problème de l'insertion des personnes handicapées dans les études et le monde du travail ainsi que l'accessibilité des lieux publics (Gold and Simson, 2005). Malgré de nombreux investissements ponctuels dans les agglomérations pour rendre l'environnement accessible, ces installations nécessitent un coût de maintenance très élevé et sont parfois laissées à l'abandon. Une des questions que nous pouvons nous poser est de savoir s'il est préférable de créer des installations supplémentaires dans les agglomérations pour les rendre accessibles ou bien s'il est préférable de créer des outils personnels permettant de mieux appréhender l'environnement sans avoir à le modifier. Les équipements rendant les installations urbaines plus accessibles sont très utiles mais pour ces raisons, nous sommes convaincus de l'utilité de concevoir et de développer des systèmes de suppléance permettant de réduire les handicaps liés à la déficience visuelle, sans modifier

l'environnement. Cette thèse a pour objectif d'augmenter l'autonomie des personnes non-voyantes en restaurant leur faculté à localiser des objets visuels.

Depuis des décennies, les recherches sur le handicap visuel ont fait émerger de nombreux dispositifs pour rendre plus accessible l'information écrite, l'informatique, le déplacement contrôlé d'un endroit à un autre (la navigation) et tout ce qui a trait aux fonctions assurées en temps normal par le système visuel. Certaines aides comme la canne blanche ou le chien d'aveugle sont devenues des standards qui ont constitué de grandes avancées pour l'autonomie des personnes aveugles. Une multitude d'aides électroniques a été développé depuis les années 60, évoluant au gré des progrès de la miniaturisation des composants et des avancées technologiques. Pendant longtemps, les systèmes ont trop souvent été construits à partir de la technologie, que les développeurs adaptaient pour répondre à des besoins. Cette démarche de conception a permis de créer des systèmes de suppléance qui fonctionnent mais qui ne sont pas adaptés aux difficultés quotidiennes des personnes en situation de handicap. Les outils mis à leur disposition ne répondent pas, pour la plupart d'entre eux, à des besoins spécifiques. Les systèmes ne peuvent pas être conçus sans une étude préalable des besoins des utilisateurs. Ces dernières années, la volonté de rassembler des chercheurs de tous horizons dans des projets pluridisciplinaires a permis de décloisonner les recherches dans les diverses disciplines et d'utiliser les ressources de chacun dans un objectif : répondre au besoin des personnes handicapées. L'objectif de cette thèse est centré sur l'hypothèse qu'un système interactif qui restitue très peu d'informations très ciblées concernant des objets utiles uniquement est plus adapté que les systèmes traditionnels avec pas ou peu de filtrage.

1) Les déficiences visuelles, classification

Grâce aux progrès thérapeutiques, le traitement de la cécité progresse. En revanche, le nombre de personnes malvoyantes s'accroit. La cécité absolue est définie par une acuité visuelle en dessous de 1/100ème après correction ou lorsque la largeur du champ visuel est inférieure à 20 degrés, quand la norme est de 180°. L'acuité visuelle est le quotient de deux nombres:

- le numérateur représente la distance à laquelle un objet est vu par le sujet,
- le dénominateur représente la distance à laquelle le même objet (sans en changer ses caractéristiques) est vu par un sujet normal.

Par exemple, une acuité de 6/60ème signifie que l'objet perçu à 60 mètres par une personne ayant une vision normale doit être rapproché à 6 mètres de la personne déficiente visuelle pour être perçu de la même façon. L'OMS a mis à profit la neuvième révision de la classification internationale des maladies pour reclasser les déficiences visuelles selon l'acuité résiduelle et le champ visuel. Elle a ainsi défini cinq catégories de déficiences visuelles numérotées de I à V. Les catégories I et II correspondent à ce qu'il est convenu d'appeler la malvoyance. On parle aussi de basse vision ou encore de vision réduite. Les critères d'évaluation reposent toujours sur une baisse d'acuité visuelle ou sur une diminution du champ visuel :

- Catégorie I : Acuité visuelle binoculaire corrigée inférieure à 3/10ème et supérieure ou égale à 1/10ème avec un champ visuel d'au moins 20°.
- Catégorie II : Acuité visuelle binoculaire corrigée inférieure à 1/10ème et supérieure ou égale à 1/20ème. En pratique, les sujets comptent les doigts de la main à trois mètres.

Les trois catégories suivantes correspondent à la notion de cécité :

Catégorie III : Acuité visuelle binoculaire corrigée inférieure à 1/20ème et supérieure ou égale à 1/50ème. En pratique, le sujet compte les doigts d'une main à un mètre mais ne peut le faire à trois mètres.

Catégorie IV : Acuité visuelle binoculaire corrigée inférieure à 1/50ème mais perception lumineuse préservée. En pratique, le sujet ne compte pas les doigts à un mètre ou champ visuel inférieur à 5°.

Catégorie V : Cécité absolue. Pas de perception lumineuse.

2) Principales causes de déficience visuelle

De nombreuses maladies affectent la vision : la Figure 1 présente la proportion des principales maladies aboutissant à la cécité dans le monde. Parmi les principales causes de cécité, seules trois maladies peuvent être traitées avec une intervention et un traitement simple : la cataracte, une opacification du cristallin, l'onchocercose, une affection parasitaire transmise par la simulie (petit moustique d'Afrique et d'Amérique) et enfin le trachome, une conjonctivite grave causée par une bactérie.

Selon l'Institut National Canadien des Aveugles (INCA), près de la moitié des personnes aveugles dans le monde le sont à la suite d'une cataracte. Cette maladie atteint plus d'une

personne sur cinq à partir de 65 ans, plus d'une sur trois à partir de 75 ans et près de deux sur trois après 85 ans. Elle provoque une baisse graduelle de l'acuité visuelle jusqu'à la cécité si celle-ci n'est pas traitée. Cette maladie affectant le cristallin est aujourd'hui bien prise en charge chirurgicalement. L'opération de la cataracte a connu de grands progrès au cours des dernières années, notamment par le remplacement du cristallin devenu opaque par un cristallin artificiel. Des complications sont possibles mais restent le plus souvent mineures. La deuxième cause de cécité, le glaucome, est une maladie engendrant une surpression intraoculaire qui se traduit par une compression du nerf optique. Il n'existe pas de traitement curatif. L'acuité visuelle perdue en raison d'un glaucome ne peut être retrouvée, l'objectif du traitement est donc de prévenir ou de ralentir les dommages subséquents. Cette maladie est la principale cause de cécité dont la prise en charge est très difficile.

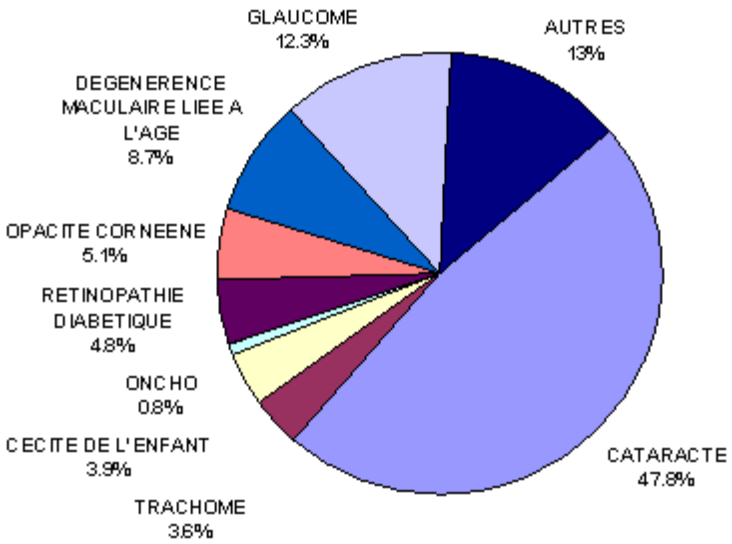


Figure 1: Origines de la cécité à l'échelle mondiale selon une étude de l'Institut National Canadien pour les Aveugles(2005)

3) Handicaps liés aux déficiences visuelles

Un rapport produit en 2000 par l'Agefiph (Association de GEstion du Fond pour l'Insertion Professionnelle des Personnes Handicapées) mentionne qu'en France, au-delà des corrections dues au vieillissement, environ 10 % de la population connaît des difficultés visuelles à des degrés divers. Selon une estimation établie sur le croisement d'informations issues des services de santé, des organismes spécialisés (scolaires et professionnels), des CDES (Commission Départementale de l'Education Spéciale) ou des COTOREP (Commission

Technique d'Orientation et de Reclassement Professionnel) et du ministère des Finances, sur 750 000 naissances annuelle, environ 13% ont ou auront un problème de vision. Malgré les progrès dans le traitement des différentes causes engendrant la déficience visuelle, le nombre de personnes qui ont ou auront des maladies liées à la déficience visuelle est considérable.

Aujourd'hui, les aveugles représentent environ 1 français sur 1 000. On estime leur nombre à environ 77 000 en France, dont 15 000 ont appris le braille et 7 000 seulement le pratiquent. La population en âge de travailler s'élèverait à environ 15 000 aveugles. Parmi eux, 2500 pratiqueraient le braille. Les malvoyants représentent environ 1 français sur 100. La population en âge de travailler serait constituée d'environ 100 000 personnes. La proportion d'étudiants déficients visuels en cycle supérieur est encore trop faible : un millier d'étudiants a été recensé en 1997/1998, 185 aveugles et 810 malvoyants sur un total de près d'1 million et demi d'étudiants. L'ensemble de ces étudiants est inscrit en université, dans les écoles d'ingénieurs, ou en I.U.T.

Très peu de pays ont étudié en profondeur les besoins des personnes non-voyantes. Une vaste étude a été menée au Canada pour tenter de répondre aux problèmes liés à la déficience visuelle. Près de la moitié des personnes à basse vision interrogées par l'Institut National Canadien pour les Aveugles affirment avoir besoin d'aide au quotidien pour contrebalancer la perte d'acuité visuelle, mais les services de soutien offerts sont insuffisants face aux difficultés rencontrées. Les services les plus fréquemment utilisés par les participants sont la formation en orientation et mobilité (apprendre à effectuer des déplacements avec indépendance, sécurité et aisance en tous environnements) (46 %), la formation pour l'utilisation des aides visuelles (40 %) et la formation pour l'acquisition des habiletés de la vie quotidienne (les habiletés de la vie quotidienne représentent les compétences auxquelles faire appel pour relever les défis du quotidien) (38 %). Presque la moitié des participants (41 %) ont déclaré avoir des besoins de services non satisfaits. Le besoin non satisfait le plus fréquemment mentionné est le transport (26 %) suivi de la formation pour l'utilisation du matériel informatique adapté et l'acquisition d'équipement adapté pour une meilleure autonomie. Ce rapport montre donc clairement que la recherche et le développement de systèmes de suppléance en orientation et mobilité représentent des enjeux sociétaux et économiques importants. Pour répondre aux problèmes liés à la mobilité des personnes non-voyantes, des systèmes de suppléance électroniques ont vu le jour

depuis les années 1970, avec un constat très mitigé sur leur utilisation. Ces aides seront dans ce manuscrit classées en deux grandes catégories :

- Les aides au déplacement aident l'utilisateur non-voyant à se déplacer dans son environnement en détectant les obstacles dans un rayon d'une quinzaine de mètres ou en tentant de restituer une représentation de l'environnement proche, dans un repère égocentré,
- Les aides à l'orientation, basées sur des outils de géolocalisation aident l'utilisateur à s'orienter dans l'espace.

Dans ce manuscrit, je présenterai dans un premier temps un état de l'art des différents systèmes de suppléance pour les non-voyants pour la navigation. Les systèmes de suppléances interactifs classiques (non-invasifs) peuvent être classés suivant deux catégories : les systèmes de substitution sensorielle et les systèmes d'augmentation sensorielle. Les recherches en neurosciences font apparaître de nouveaux modes d'interaction entre l'homme et la machine, et plus précisément entre le cerveau et la machine : les neuroprothèses. Je présenterai différents projets de recherche dont l'objectif est de restaurer la vision en stimulant directement différents étages du système visuel humain.

Finalement, comme je le décrirai dans mon introduction, l'état de l'art sur les aides à la navigation pour les non-voyants montre que la plupart d'entre eux ne répond pas au besoin des utilisateurs de par leur utilité et leur utilisabilité. Dans la suite de ce manuscrit je décrirai les travaux que nous avons menés dans le but d'augmenter l'autonomie des non-voyants. Notre hypothèse est qu'il suffit de restituer une petite quantité d'information pour accroître l'autonomie des non-voyants.

Plutôt que de tenter de restituer l'intégralité de la scène visuelle, nous proposons de restaurer une fonction utile du système visuel pour la navigation et la recherche d'objet : la reconnaissance et la localisation d'objets. Je décrirai les étapes de conception et de prototypage d'un système interactif de localisation d'objets pour les non-voyants, incluant un module de vision artificielle. Je montrerai ensuite comment ce type d'approche peut parfaitement s'adapter à la thématique nouvelle des neuroprothèses visuelles. Je décrirai notamment un modèle fonctionnel de neuroprothèse visuelle en cours d'élaboration, auquel j'ai pu contribuer pendant mon doctorat.

Mon travail de doctorat était le premier travail sur cette thématique dans l'équipe et a servi de socle à l'élaboration d'un projet collaboratif plus large d'aide à la navigation pour piétons déficients visuels. Ce projet important initié en Janvier 2009, qui rassemble six partenaires, repose sur la fusion du système interactif de localisation d'objets que je présente ici avec un outil de géolocalisation pour piéton en milieu urbain. Je décrirai ce projet en essayant de montrer comment les perspectives de mon travail de doctorat se trouvent à la base de certains travaux aujourd'hui développés dans ce projet.

CHAPITRE I : LES SYSTEMES DE SUPPLEANCE VISUELLE

La suppléance visuelle pour les non-voyants a pour objectif de restaurer certaines fonctions utiles du système visuel dont l'absence engendre un handicap chez une personne aveugle. Cette suppléance doit fournir une information utile équivalente à celle de la vision par l'intermédiaire de l'audition ou de la somesthésie et s'est focalisée sur deux problèmes principaux que sont la lecture et l'écriture d'une part, et la navigation d'autre part, sur laquelle cette thèse porte plus particulièrement.

Il existe aujourd'hui de nombreux dispositifs commercialisés et de nombreux projets de recherche que nous ne décrirons pas liés à l'accessibilité de l'information (lecture et écriture) pour les personnes non-voyantes. Ces systèmes sont largement basés sur la somesthésie (en convertissant l'image ou le texte présent sous un capteur de vision artificielle (caméra, scanner) en stimuli tactiles (relief)), ou sur l'audition (moteurs de synthèse transformant une chaîne de caractères en parole). L'ensemble des systèmes de suppléance électroniques ont une architecture assez semblable. En effet, ils sont construits suivant le même principe (Figure 2) : une chaîne d'acquisition, une chaîne de transformation de l'information et un module de restitution de l'information traitée.

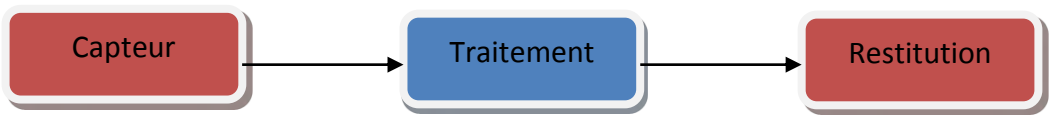


Figure 2: Schéma de fonctionnement d'un dispositif de suppléance. Le capteur acquiert l'information qui est traitée puis restituée à l'utilisateur.

Préambule : quelques notions importantes

1) Substitution sensorielle vs. Augmentation sensorielle

Il existe deux catégories de systèmes de suppléance visuelle. Les systèmes les plus étudiés jusqu'à maintenant sont les systèmes de substitution sensorielle : ils restituent les informations habituellement acquises par une modalité sensorielle vers une autre modalité sensorielle. De nombreux dispositifs ont ainsi vu le jour depuis les années 1970, avec une substitution visuo-tactile puis visuoauditive. Ces systèmes reposent sur l'hypothèse qu'il est possible de restituer suffisamment d'informations visuelles engendrant une perte

d’autonomie à travers une autre modalité sensorielle pour que l’utilisateur les interprète et les traite comme des informations visuelles.

Sur un tout autre principe que les systèmes de substitution sensorielle, depuis les années 1990, de nouvelles aides électroniques ont été créées avec pour objectif de regrouper l’ensemble des technologies disponibles pour répondre aux besoins des utilisateurs. La démocratisation des télémètres lasers ou à ultra-sons a permis d’améliorer un outil utilisé par la plupart des personnes aveugles pour la détection d’obstacles : la canne blanche. Contrairement aux systèmes de substitution sensorielle dont l’objectif est de restituer la vision par une autre modalité sensorielle, ces systèmes augmentent l’information d’une modalité sensorielle pour donner des informations, utiles à la navigation par exemple. Nous appellerons systèmes d’augmentation sensorielle (Kaczmarek, 2000b;Lenay et al., 2003) ces systèmes dont le capteur n’est pas un capteur de vision et la restitution n’est pas une « image » visuelle "traduite" pour une autre modalité sensorielle. L’exemple de la Figure 3 illustre la différence entre ces deux notions : une scène visuelle comportant une voiture, capturée par une caméra peut être restituée soit par substitution sensorielle (Bach-Y-Rita) en ***restituant le motif visuel*** dessiné par la voiture vers une autre modalité sensorielle soit par augmentation sensorielle en ***traduisant l’information pour l’adresser à une autre modalité sensorielle***, par exemple en synthèse vocale. Dans le domaine de la suppléance, la substitution sensorielle est souvent utilisée pour couvrir l’ensemble des systèmes de suppléance visuelle recourant à une autre modalité sensorielle. Tout au long de ce manuscrit, ces deux notions seront présentées de manière distincte.

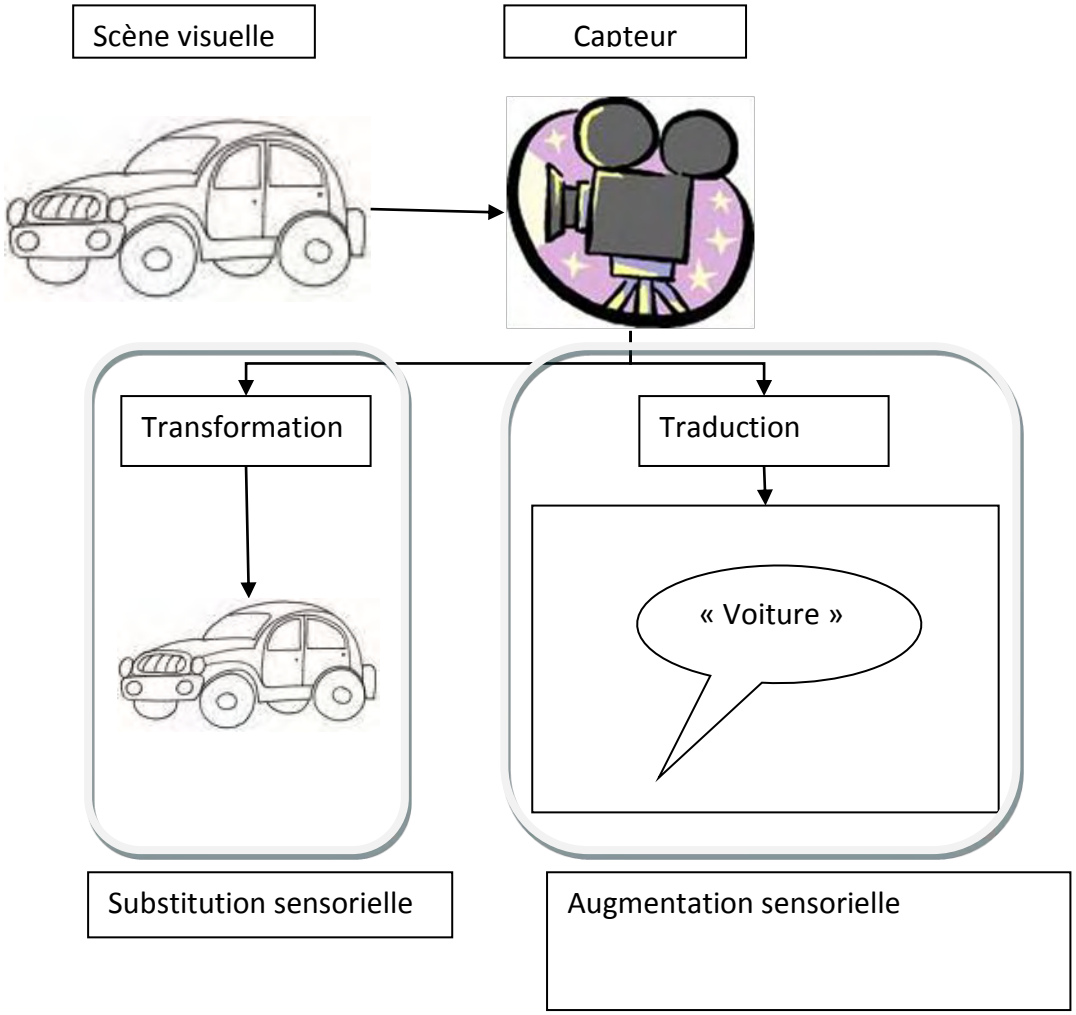


Figure 3: Illustration de la comparaison entre substitution sensorielle et augmentation sensorielle. L'image est soit transformée en une "image" de stimulation électro-tactile, soit en une information de plus haut niveau sur le contenu sémantique de l'image qui peut être restitué par synthèse de la parole par exemple.

2) Aide à la navigation : déplacement et orientation

L'aide à la navigation pour les non-voyants a pour objectif de les aider à se déplacer en sécurité et à s'orienter. Le but des dispositifs d'aide au déplacement est d'améliorer la mobilité des personnes non-voyantes en restituant une information sur la position spatiale des obstacles environnants (sont considérés comme obstacles ici l'ensemble des objets dont on ne connaît pas l'identité) ou des objets (identifiés). De façon plus globale, nous définirons comme « objet » toute entité visuelle perceptible, concrète (ex. un motif visuel sur un mur). Je considérerai tout au long de ce manuscrit cette définition très vaste d'objet. Pour les non-voyants, ces deux catégories de systèmes sont complémentaires pour la navigation dans des environnements inconnus.

Les approches de la substitution et de l'augmentation sensorielles montrent que différentes informations qui peuvent être restituées pour augmenter l'autonomie des non-voyants. Il existe en effet une multitude de capteurs de complexités différentes. Les aides au déplacement (Electronic Travel Aids – ETA) (Farcy and Damaschini, 2000) ont pour objectif de décrire l'environnement proche du sujet par une appréciation du relief ou des objets environnants mais ne permettent pas à l'utilisateur de se situer dans un endroit inconnu. Elles sont en revanche indispensables pour la navigation fine dans des endroits inconnus, puisque l'environnement y est décrit dans un repère relatif à l'utilisateur.

Les aides à l'orientation (Electronic Orientation Aids - EOA) de leur côté ont pour objectif de permettre aux aveugles de se déplacer d'un point à un autre sur un parcours connu ou inconnu en utilisant par exemple des capteurs de localisation absolue de l'utilisateur dans son environnement. Ces derniers semblent très prometteurs mais posent des problèmes d'utilisabilité du fait de leur précision insuffisante pour un piéton non-voyant.

3) Systèmes interactifs électroniques vs. Neuroprothèses

La conception de systèmes interactifs repose sur une conversion d'une information vers une autre modalité sensorielle, tactile ou auditive. Ces systèmes de suppléance non-invasifs, sont aujourd'hui les plus utilisés. Une autre voie de recherche est très active ces dernières années : la conception de systèmes directement connectés au système visuel - les neuroprothèses visuelles - qui auront des applications à plus long terme mais avec un impact potentiel très important sur la qualité de vie des non-voyants.

Les systèmes électroniques de suppléance visuelle

4) Les aides au déplacement

Les systèmes basés sur des caméras vidéo ont depuis les années 1970 beaucoup évolué, permettant d'imaginer de nouveaux outils pour la suppléance visuelle. Avec la miniaturisation et la démocratisation des caméras dans une multitude d'appareils très répandus, ces systèmes sont devenus très bon marché. Les capteurs utilisés pour la vision artificielle (caméras) sont définis par leur résolution, leur fréquence d'échantillonnage des images. Les systèmes de suppléance par substitution sensorielle reposent sur un capteur de vision artificielle, considéré comme capteur de substitution de la vision humaine déficiente. Les informations, brutes ou très peu filtrées, sont alors reproduites sous forme de signal auditif ou tactile. La seconde approche, celle de l'augmentation sensorielle, vise à

augmenter une modalité sensorielle avec des informations qu'elle ne traiterait pas normalement (ex. la localisation d'une cible visuelle par l'audition). Une transformation de l'information provenant du capteur d'entrée est alors nécessaire. Pour un capteur de vision artificielle, ce n'est pas un filtre sur l'image qui est appliqué mais un algorithme pour en analyser le contenu. C'est alors une partie de l'information traitée et filtrée qui est restituée.

Les systèmes de substitution sensorielle

Les systèmes de suppléance basés sur la substitution sensorielle tentent de restituer les informations visuelles brutes issues d'une caméra et échantillonnées en fonction des caractéristiques techniques de celui-ci. Dans ces systèmes, les images en provenance de caméras sont transformées en signaux pour une autre modalité sensorielle en essayant de préserver un maximum d'information sur l'organisation spatiale de l'image.

Substitution Vision – Sens tactile

Comme l'ont montré Paul Bach-y-Rita et collègues en développant un dispositif de conversion des informations visuelles capturées par une caméra en des sensations tactiles à la surface du corps (TVSS) (Bach-y-Rita et al., 1969) (conversion des informations visuelles capturées par une caméra en des sensations tactiles à la surface du corps), il est possible de stimuler le système tactile humain pour percevoir des formes. Bach-y-Rita, a ensuite conçu un dispositif plus léger, le Tongue Display Unit (TDU) permettant de stimuler la langue. La substitution sensorielle pour les non-voyants a été déclinée plus tard sur la langue (Kupers and Ptito, 2004), l'abdomen (Bach-y-Rita, 1983), et le palais (Hui and Beebe, 2003;Hui and Beebe, 2006) . L'étude de ces dispositifs a permis de démontrer de nombreux résultats fondamentaux qui seront présentés dans cet état de l'art.

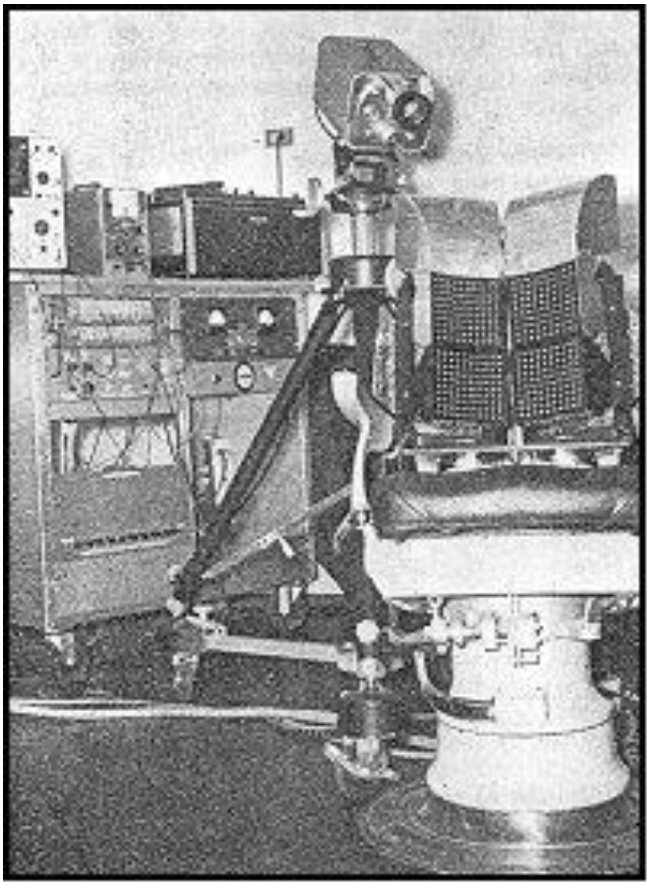


Figure 4: Première implémentation du TVSS avec une matrice de stimulation électro-tactile dans le dos. Le sujet s'asseyait dans la chaise et percevait des sensations à la surface de la peau en relation avec la scène visuelle capturée par une caméra fixe

Dans la première version du TVSS, la grille de stimulation électro-tactile était montée sur une chaise de dentiste (Figure 4). Le sujet percevait ainsi des sensations sur son dos. Dans cette première version, la caméra était fixe et les sujets n'arrivaient pas à percevoir de formes, même très simples. Plusieurs études (Auvray et al., 2007;Bach-y-Rita, 1983) ont montré par la suite que l'interaction avec le capteur était nécessaire pour percevoir des objets provenant de la scène visuelle. Ce dispositif est le premier dispositif de substitution sensorielle tel qu'usuellement décrit : la transformation d'une information provenant naturellement d'une modalité sensorielle vers une autre modalité sensorielle (Kaczmarek, 2000a). Le principal inconvénient du TVSS reposait sur la faible résolution des sensations tactiles à cet endroit du corps.

En 1998, Bach-y-Rita proposera d'utiliser la langue à la place du dos puisque c'est un des organes avec la plus forte densité de récepteurs tactiles, permettant une plus grande résolution de stimulation. La première version de ce dispositif (Bach-y-Rita et al., 1998) avait une résolution de 7x7 électrodes de stimulation (matrice de 49 électrodes : Figure 5).

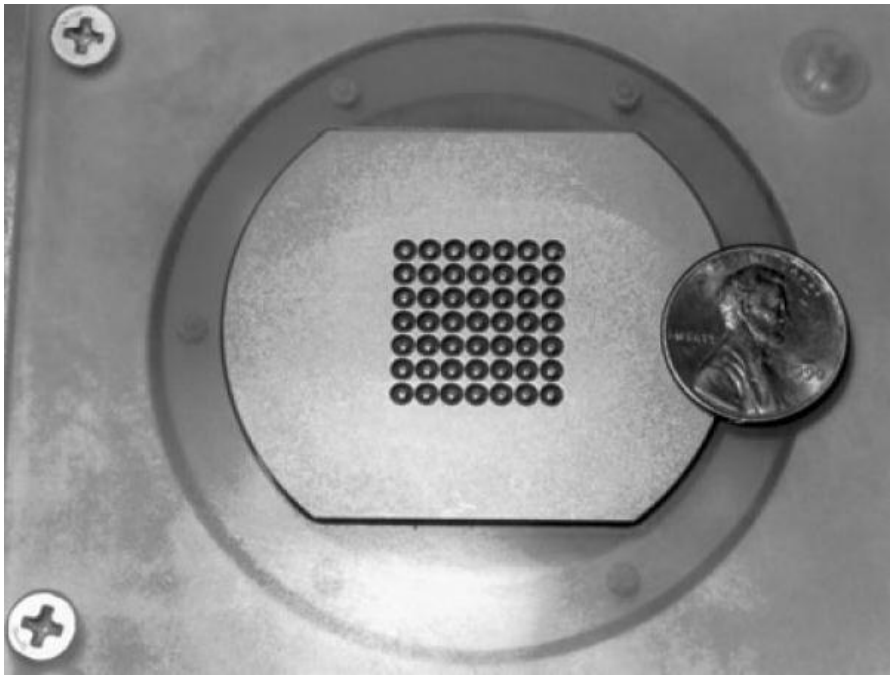


Figure 5 : Première version du TDU (Bach-y-Rita et al., 1998) : une matrice de 49 électrodes électro-tactiles (espacées de 2,54 mm, chacune mesurant 0,89 mm de diamètre) posée sur la langue des sujets. Un penny américain est disposé sur le coté droit de l'image pour donner l'échelle.

Cette version du TDU a été évaluée auprès de cinq adultes voyants. La tâche consistait à reconnaître des motifs simples (Figure 6). Dans cette évaluation, les sujets ne portaient pas de caméras, le motif était directement restitué sur la matrice de stimulation.

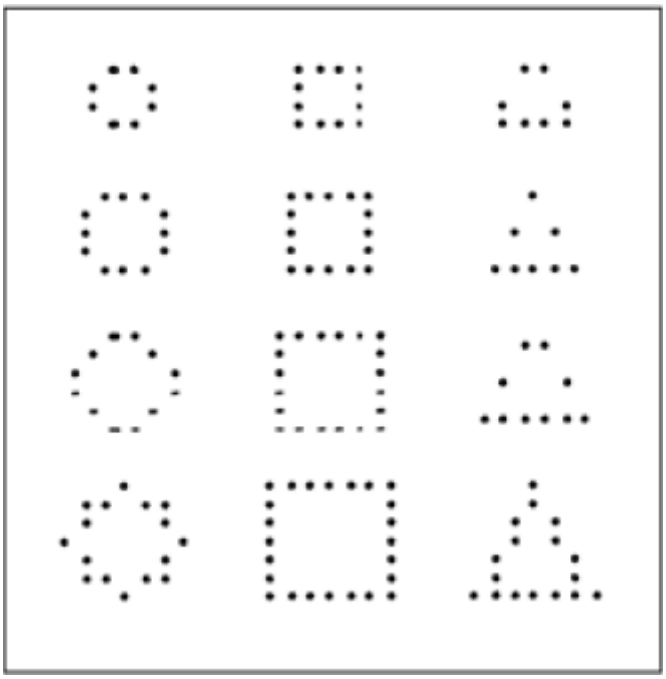


Figure 6 : Motifs à discriminer avec le TDU dans sa première version (Bach-y-Rita et al., 1998)

Les sujets n'avaient aucune contrainte de temps et pouvaient apprendre à discriminer les motifs dans une plus grande taille avant l'expérimentation. Pendant le test, les sujets devaient reconnaître les motifs au cours de quatre séries de 12 motifs chacun. Les résultats ont montré qu'il est possible avec un pourcentage de reconnaissance élevé (80%) de discriminer les motifs dans toutes les tailles. Ces résultats sont comparés aux résultats reportés dans une autre étude (Kaczmarek et al., 1997) avec le même protocole et le même dispositif expérimental (Figure 5) mais une exploration avec le bout du doigt sur la matrice électro-tactile. La fiabilité de reconnaissance des motifs est très proche (90%) de celle obtenue avec une stimulation de la langue alors que cette dernière ne nécessite pas de balayage. Les temps de reconnaissance ne sont pas mentionnés dans l'étude mais sont intéressants puisque l'absence de balayage permet probablement une reconnaissance plus rapide. Au début des années 2000, une nouvelle version du TDU est créée avec 144 électrodes de stimulation (matrice de 12x12). L'acuité "visuelle" avec un tel système au sens du test standard de Snellen[1] a été mesurée en le connectant à une caméra de faible résolution (240x180) et de 54° d'angle de vue (Sampaio et al., 2001). Le groupe de sujets comportait 6 voyants et 6 non-voyants congénitaux, tous naïfs avec ce dispositif. Les stimuli étaient composés de dérivés du 'E' de Snellen (Figure 7c) dans six tailles différentes (5 ; 3,6 ; 2,5 ; 1,8 ; 1,5 et 0,85 cm) et quatre orientations. Les sujets pouvaient manuellement bouger la caméra (à distance fixe contrainte par un bras articulé : 40 cm de la source).

[1] L'acuité visuelle de Snellen consiste en un rapport entre deux nombres caractérisant le test dans lequel l'utilisateur a 100% de bonnes identifications. Le numérateur correspond à la distance à laquelle le test est effectué, le dénominateur représente la distance à laquelle le plus petit optotype ('E') de l'expérimentation est perçu.

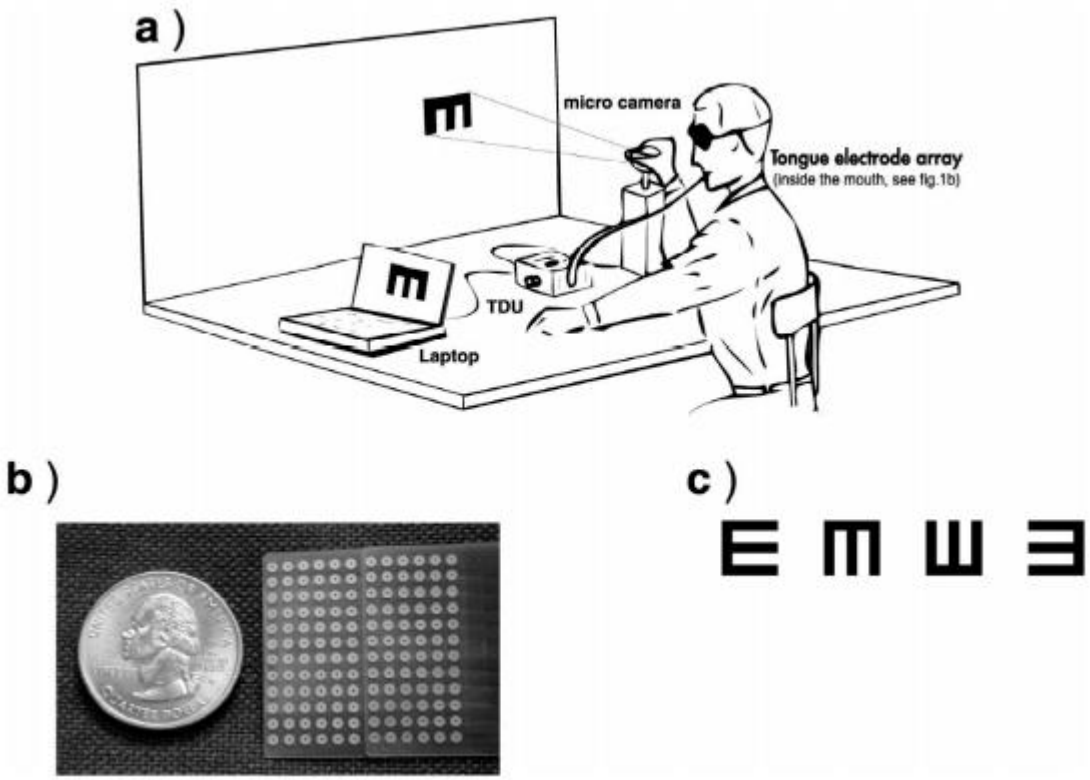


Figure 7 : a) protocole expérimental. Le sujet tient une caméra fixe à 40 cm. La restitution de l'image capturée par la caméra est effectuée par le TDU, l'utilisateur doit établir quand le E de Snellen est reconnu. B) Matrice d'électrodes (12x12) de stimulation de la langue c) Stimuli : E de Snellen dans différentes orientations (Sampaio et al., 2001).

Avant tout apprentissage du système, l'acuité de l'ensemble des sujets (égale dans les deux groupes) était proche de 20/860, c'est-à-dire que le plus petit stimulus visuel reconnaissable mesure 5° d'arc à 860 pieds (260 m). Les stimuli utilisés étaient très grands (environ 35 cm x 35 cm). Un voyant et un non-voyant ont poursuivi pendant neuf heures un apprentissage du dispositif. Cet apprentissage consistait en la détection d'une ligne prenant différentes tailles et orientations ou de deux lignes formant un angle de 45° ou 90°. L'acuité visuelle des sujets après apprentissage avait doublé mais restait très faible : 20/430.

En conclusion, nous notons que la substitution sensorielle visuo-tactile conduit à des taux de reconnaissance de forme simples relativement faibles, ce qui rend les dispositifs de suppléance basés sur cette méthode peu utilisables. Nous pensons que ce problème vient du fait que la résolution de la caméra est bien supérieure à la résolution du système sensoriel. Sur le même principe de substitution sensorielle, de nombreux systèmes de substitution sensorielle visuo-auditive ont été créés et étudiés.

Substitution Vision - Audition

La substitution de la vision vers l'audition convertit l'image en son en préservant un maximum d'information spatiale et lumineuse dans l'image. La position des pixels dans l'image à restituer ainsi que leur intensité lumineuse sont restituées en utilisant 4 propriétés des sons restitués : 1) la fréquence 2) l'intensité 3) le temps 4) les différences inter-aurales.

Quatre principaux systèmes de substitution sensorielle de la vision vers l'audition existent. Le plus connu étant 'The vOICe' (Meijer, 1992), développé depuis 1992 par l'ingénieur Peter Meijer, du laboratoire de recherche de Philips à Eindhoven aux Pays-Bas (les capitales signifiant « Oh I see » pour « Oh, je vois »). Le système EAV (Gonzalez-Mora et al., 2006) est développé à l'Université de la Laguna aux Canaries. Le système PSVA (Arno et al., 1999) a été mis au point en 1999 par Capelle et collaborateurs. Le système le plus récent (2004), 'The Vibe', est issu d'une collaboration entre le laboratoire de Neurophysique et Physiologie du Système Moteur (Sylvain Hanneton) et le laboratoire de Psychologie Expérimentale, tous deux à l'université René Descartes (Sylvain Haupert, J. Kevin O'Regan, Malika Auvray). La principale différence entre ces quatre systèmes de substitution de la vision par l'audition réside dans le codage de l'information.

Ces quatre systèmes fonctionnent donc sur le même principe de conversion d'une image en sons et ont été évalués dans des tâches de localisation et de reconnaissance d'objets.

The vOICe

Le dispositif 'The vOICe' convertit chaque pixel de l'image en transformant sa position spatiale et son intensité lumineuse en son (Figure 8).

1. La position verticale est codée en fréquence (résolution de 64 fréquences différentes). Plus le pattern visuel est haut dans l'image et plus le son est aigu ; plus le pattern visuel est bas, plus le son est grave.
2. La position horizontale est codée de manière temporelle : chaque image est scannée en une seconde et seul un secteur de l'image est sonifié à chaque instant. La résolution auditive de l'image est de 64 pixels en largeur. Chaque tranche de 1/64eme d'image est ainsi sonifiée une fois en une seconde pendant 1/64 secondes.
3. La valeur en niveaux de gris d'un pixel est codée en intensité du son émis : plus un pixel est clair, plus son intensité sera élevée. Par conséquent, un silence veut dire noir et un son fort veut dire blanc. Toutes les intensités intermédiaires correspondent à un dégradé de gris

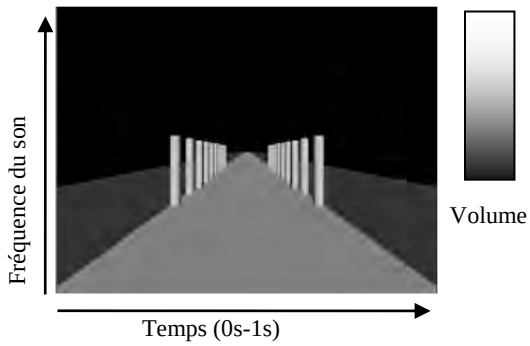


Figure 8: Codage de l'image dans le système 'the vOICe'

The vOICe a été évalué dans des tâches de localisation (une boîte noire de 11x11x1 cm) et de catégorisation d'objets (10 objets avec chacun 9 variantes aux contenus dans l'image très semblables). Les objets sont disposés sur une table blanche et les sujets se tiennent à 1 mètre environ de la table avec une caméra dans la main. Chaque image est convertie en signaux auditifs par le système 'The vOICe'. Les résultats présentés dans la thèse de Malika Auvray (Auvray, 2004) montrent qu'il est possible de localiser et reconnaître des objets : le temps moyen de localisation d'une cible est en moyenne de 100 +/- 70 secondes avec une erreur de 7 +/- 5 cm. Les sujets arrivent à catégoriser les objets en énumérant en moyenne 1 objet et demi en 39 +/ - 27 secondes avant de trouver l'objet correct.

Figure 9 : Exemple de stimuli utilisés dans une tâche de catégorisation reposant sur le système de substitution sensorielle The vOICe (Auvray, 2004)

Cette expérimentation montre qu'il est possible en agissant sur la caméra, de localiser et de reconnaître des objets appris. Ces deux tâches nécessitent cependant beaucoup de temps et les conditions dans lesquelles ont été effectués les tests montrent la difficulté de les utiliser en environnement réel.

PSVA

Le système PSVA fonctionne sur le même principe que The vOICe sans utiliser la composante temps pour coder l'information. En effet, la fréquence du son est utilisée pour le codage (Figure 10) de l'information horizontale rendant la résolution dans les deux dimensions plus faible.

1. La position verticale est codée en fréquence. Plus le pattern visuel est placé haut, plus le son est aigu ; plus le pattern visuel est bas, plus le son est grave.
2. La position horizontale est également codée en fréquence, comme la position verticale. La gamme de fréquence utilisée va croissante de la gauche vers la droite de l'image.
3. La valeur en niveaux de gris d'un pixel est codée en intensité du son émis : plus un pixel est clair, plus son intensité sera élevée.

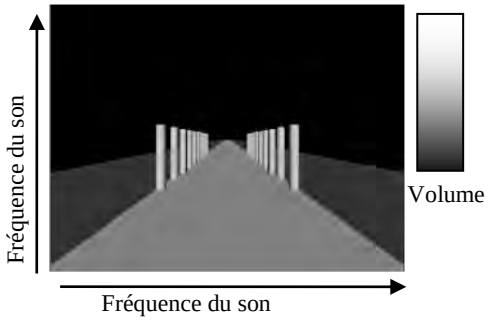


Figure 10: Codage de l'image dans le système PSVA

Sur un principe de restitution très proche, PSVA est développé à l'université catholique de Louvain. Son principe est, comme 'The vOIce', de substituer l'audition à la vision. Comme dans les autres systèmes, le codage vertical de l'information est fréquentiel. Là où le système se distingue, c'est que le codage horizontal est également un codage fréquentiel. Le principal avantage d'un tel codage vient du fait que l'information est restituée de manière instantanée à l'utilisateur sans qu'il ne soit nécessaire de balayer l'image à chaque seconde comme dans 'The vOICe'. En revanche, le choix d'un même mode de codage fréquentiel pour le codage vertical et horizontal réduit considérablement la résolution des images qu'il est possible de restituer par ce système. En effet, 'The vOICe' code 64 positions verticales en fréquence et 64 positions horizontales dans la dimension temporelle alors que l'ensemble des positions des pixels de PSVA doit être codé en fréquence (124 positions possibles). Afin

d'améliorer la précision de l'information au centre de l'image, la résolution y est doublée (Figure 11). Ce système permet ainsi à l'utilisateur de scanner l'environnement à basse résolution et affiner l'information en centrant l'objet souhaité.

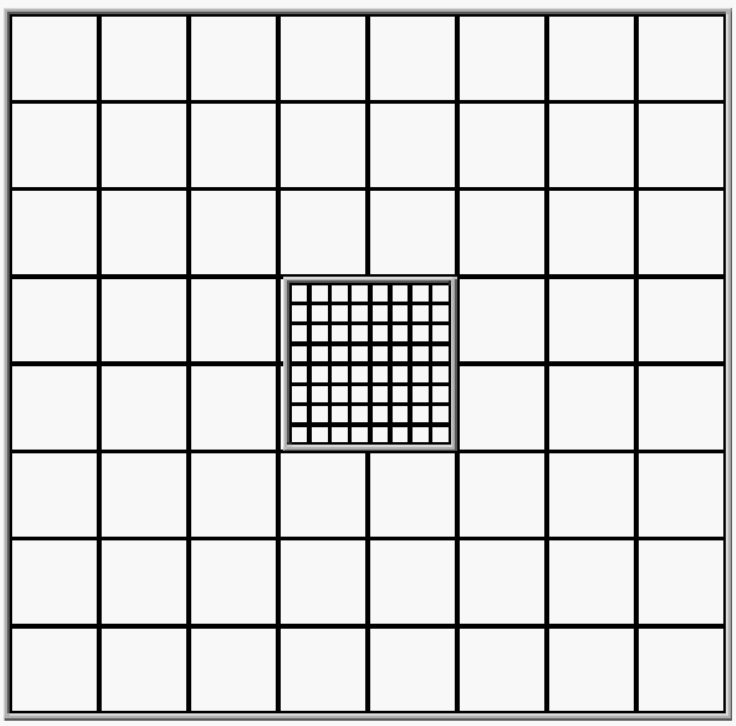


Figure 11: Résolution en deux dimensions d'une image traitée pour être restituée par 'PSVA'. La partie centrale du champ visuel bénéficie d'une résolution doublée.

Le dispositif PSVA développé à l'Université Catholique de Louvain (Arno et al., 1999) comporte une caméra miniature portée sur la tête, qui fournit une image en niveaux de gris à un ordinateur. Le rafraîchissement de la restitution est effectué à 25Hz, ce qui permet à l'utilisateur d'avoir une interaction sensorimotrice dynamique avec l'environnement.

24 sujets voyants, âgés de 19 à 36 ans et ne présentant pas de déficit d'audition ont évalué le système les yeux bandés. Les sujets ont été répartis en deux groupes (le groupe contrôle et le groupe test) de 12 personnes comportant chacun 6 hommes et 6 femmes. L'objectif était ici d'étudier l'impact de l'apprentissage sur la reconnaissance de motifs visuels. Le groupe test a effectué 14 sessions : 10 pour l'apprentissage et 4 pour l'évaluation du système tandis que les sujets contrôles ont seulement effectué les 4 sessions d'évaluation. Avant chaque session expérimentale, les sujets étaient assis à une table, les yeux bandés, en face d'un tableau blanc à environ 30 cm de distance. Avant la première session d'évaluation, l'expérimentateur expliquait le fonctionnement du système, en faisant apparaître un point noir sur le tableau blanc. Il était demandé aux sujets de bouger la tête en explorant l'influence de ces mouvements sur le son produit en réponse à ce stimulus visuel. Une

explication plus précise du lien entre les sons et les stimuli intervenait pendant l'apprentissage. Finalement, il était demandé aux utilisateurs de localiser un motif visuel en utilisant explicitement les différences inter-aurales.

Un total de 50 stimuli dérivés de 15 patterns visuels, dans différentes orientations (Figure 12) a été évalué dans des tâches de reconnaissance et de localisation. Les sessions d'apprentissage pour le groupe test s'étendaient sur environ 6 à 7 semaines (pour 10 sessions) avec des motifs visuels de plus en plus complexes à reconnaître (Figure 12). Durant les sessions d'entraînement, les sujets devaient manuellement réarranger des petits carrés d'aluminium (8,5 mm x 8,5 mm) ou des barres (68 mm x 8,5 mm) pour représenter la forme perçue dans un espace délimité de 180 mm x 180 mm. Lorsqu'une mauvaise réponse était donnée, l'expérimentateur réarrangeait la forme et le sujet pouvait toucher la forme corrigée. Pour chaque session d'une durée d'environ une heure, 10 motifs visuels étaient testés. A la fin de chaque session d'apprentissage, 5 des 10 patterns appris étaient testés et les sujets devaient donner une réponse en les catégorisant en moins de 3 mn chacun, les sujets avaient un retour tactile comme précédemment sur leurs réponses. Une fois l'apprentissage terminé, l'expérimentation était poursuivie par 4 sessions d'évaluation au cours desquelles les sujets contrôle les rejoignaient.

Lors de sessions d'évaluation, 25 motifs visuels étaient choisis au hasard parmi les 50 de la Figure 12 et présentés dans un ordre aléatoire à chaque sujet. Pour chaque motif visuel qui apparaissait, un chronomètre était déclenché et il était demandé aux sujets de dire « STOP » le plus rapidement possible une fois qu'ils avaient identifié le stimulus. Au bout de 3 minutes, le son s'arrêtait si les sujets n'avaient pas dit « STOP ». Dans les deux cas, il leur était demandé d'utiliser les carrés et bandes d'aluminium décrits plus haut pour reconstruire le motif exploré. Les mêmes 25 motifs visuels choisis au hasard étaient présentés durant les 3 sessions d'évaluation suivantes. Les résultats montrent qu'il y a un effet significatif de l'apprentissage sur les performances de reconnaissance et sur le temps de reconnaissance, qui reste cependant élevé dans les deux groupes de sujets (entre 80 s et 110 s pour le groupe expérimental et entre 70 s et 80 s pour le groupe contrôle). Les sujets des deux groupes ont les mêmes performances de reconnaissance (temps et précision) lors de la première session d'évaluation mais le groupe test s'améliore beaucoup plus vite que le groupe contrôle. Le temps de traitement est en moyenne plus élevé chez les sujets contrôle mais avec des scores de reconnaissance plus élevés.

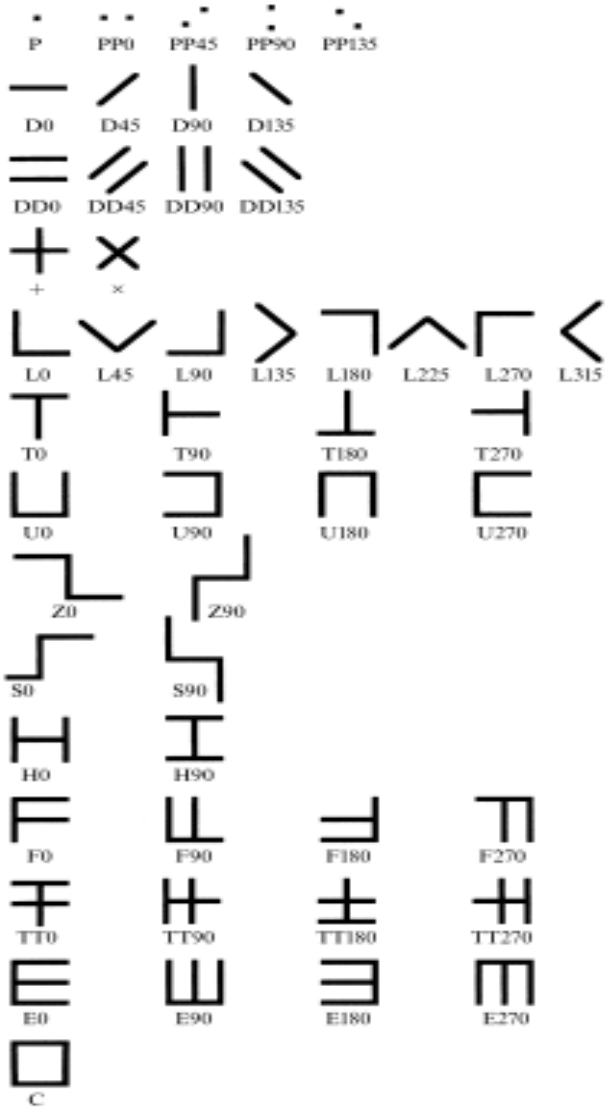


Figure 12: 50 stimuli utilisés dans une tâche de reconnaissance de motif visuel (Arno et al., 1999) avec le dispositif PSVA.

Les résultats obtenus montrent qu'avec un apprentissage long, les utilisateurs s'améliorent mais les temps de reconnaissance de ces motifs très simples (85 s) sont prohibitifs pour une utilisation au quotidien par des personnes non-voyantes. Le principal inconvénient de ce système semble être la faible résolution en fréquence disponible en audition pour un codage des positions de tous les pixels.

The Vibe

Le système The Vibe utilise un système de codage (Figure 13) assez proche des deux autres.

1. La position verticale est codée en fréquence. Plus le pattern visuel est haut dans l'image, plus le son est aigu ; plus le pattern visuel est bas, plus le son est grave.

2. La position horizontale est codée en disparité inter-aurale. Le côté droit de l'image correspond à l'écouteur droit et le côté gauche correspond à l'écouteur gauche.
3. La valeur en niveaux de gris d'un pixel est codée en intensité du son émis : plus un pixel est clair, plus son intensité sera élevée

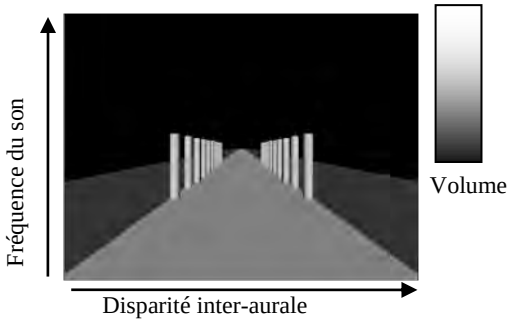


Figure 13: Codage de l'image dans 'The Vibe'

Le système 'The Vibe' (Figure 14) a été évalué en condition réelle dans des tâches de navigation (Durette et al., 2008). 20 voyants âgés entre 22 et 38 ans ont participé à l'expérience, trois d'entre eux avaient déjà participé à une expérience avec une version plus ancienne du dispositif mais n'ont pas eu de résultats significativement différents des autres.

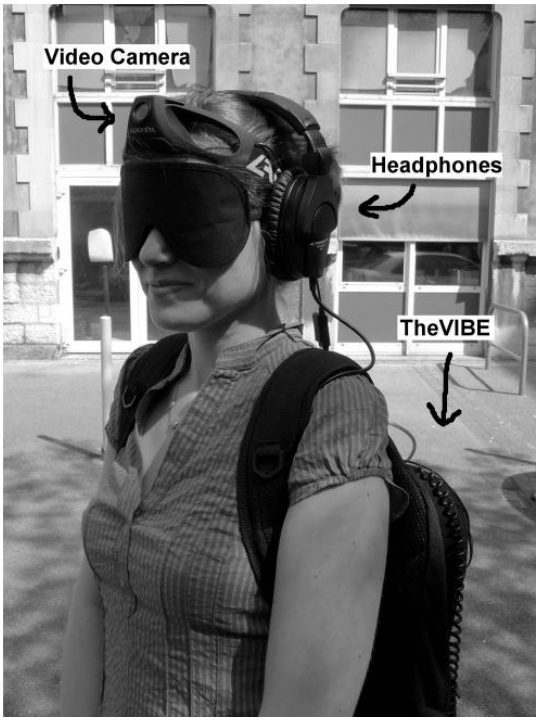


Figure 14 : illustration du système 'The Vibe' : une caméra grand angle est disposée sur le front du sujet. L'image est envoyée au système qui la restitue sous forme de sons dans un casque audio.

Les sujets portaient sur la tête une caméra grand angle (92°) et devaient effectuer un parcours en 'U' (Figure 15) durant quatre sessions expérimentales (espacées d'au moins 24h) en enregistrant le temps total pour effectuer la tâche et le nombre de collisions. Les trois premières sessions étaient des sessions d'apprentissage dans lesquelles l'expérimentateur guidait le sujet avec le bras (première session) ou verbalement par des instructions « gauche » ou « droite » (pour les sessions 2 et 3).

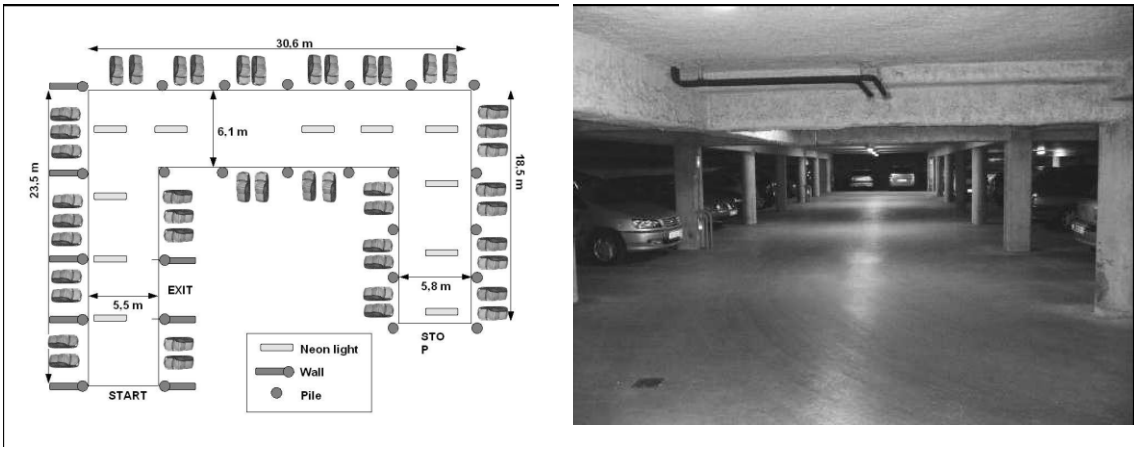


Figure 15: Plan (à gauche) et photo (à droite) de l'environnement dans lequel se déroulait l'expérimentation. Les sujets devaient effectuer le parcours de Start vers Stop.

En moyenne, le temps de parcours au fil des 3 sessions pour lesquelles la tâche était la même diminue de 235 secondes à la première session à 170 secondes à la troisième session. De même, le nombre de fois durant le trajet où le sujet sort de la route prévue diminue en moyenne de 7,2 à 6 entre les trois sessions. La session de test consistait en 3 essais : deux d'entre eux étaient en condition normale d'entraînement et la troisième en condition image inversée : l'image était symétriquement inversée sur l'axe vertical. Les auteurs mentionnent un effet positif du système sur la navigation : les sujets sont plus rapides et font moins de sortie de route avec le dispositif qu'ils ont appris qu'avec ce même dispositif en image inversée.

EAV

Le projet EAV, développé à l'université de La Laguna aux Canaries, s'appuie sur un principe de restitution différent. L'idée consiste à utiliser les capacités spatiales du système auditif pour restituer la position et la forme des objets de la scène visuelle. Pour cela, un capteur stéréoscopique composé de deux caméras montées sur une paire de lunettes permet d'établir les coordonnées 3D des pixels présents dans la scène dans un repère centré sur la tête. L'objectif est ensuite de réduire la résolution de la scène et de synthétiser des sons comme si ceux-ci provenaient de petites enceintes placées à la surface des objets. Cette

méthode présente le principal avantage d'utiliser la capacité des humains à localiser des sources sonores dans l'espace pour indiquer avec précision où se trouvent les objets. Le fait de positionner des enceintes virtuelles sur toute la surface des objets a pour objectif de donner une représentation de la forme présente dans la scène visuelle.

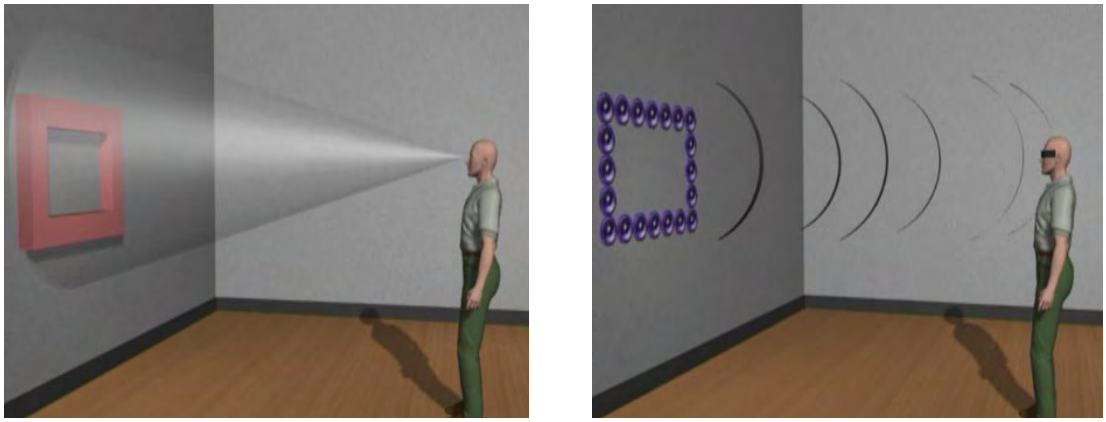


Figure 16 : Illustration du fonctionnement de EAV. L'image de gauche représente l'utilisateur du système face à un cadre resctangulaire accroché à un mur. Cette scène visuelle est capturée par des caméras portées sur des lunettes. L'image de droite représente ce que perçoit l'utilisateur auditivement : des sons spatialisés comme si ceux-ci provenaient de la surface de la forme présente dans la scène visuelle.

Le principal objectif de cette approche est de restituer une image des volumes présents dans la pièce ou à l'extérieur en utilisant les facultés de l'audition pour localiser des sources sonores dans l'espace et ainsi se représenter l'environnement en trois dimensions. La synthèse des sons spatialisés est effectuée en calculant le stimulus sonore de chaque oreille en fonction de la position de la source sonore à synthétiser. Ces fonctions de transfert sont établies pour chaque individu en plaçant deux petits micros intra-auriculaires dans les oreilles des sujets et en enregistrant la différence de signal entre le son émis et le son enregistré dans chaque oreille. Ces fonctions de transfert sont très différentes d'un sujet à l'autre puisqu'elles sont fonctions de la morphologie. Une fois ces fonctions de transfert établies, les utilisateurs se familiarisent avec le système pour apprendre à discriminer des formes en observant différentes lignes blanches de 50 cm dans différentes directions sur un tableau noir placé à 90 cm (Figure 17).

Figure 17 : Familiarisation avec le système EAV pour repérer en deux dimensions des cibles visuelles placées à 90 cm.

Lors de cette première prise en main avec le dispositif, les sujets devaient se concentrer sur la provenance des sons. La description de la scène perçue par les sujets ainsi que la position enregistrée par un suivi magnétique 3D étaient enregistrées. A la fin de l'essai, l'expérimentateur prenait le bras du sujet pour indiquer la position des formes visuelles. Ces expériences ont été effectuées auprès de personnes non-voyantes de différents âges et degrés de déficience. L'ensemble des formes était correctement décrit et spatialisé dans la majorité des cas.

Ces systèmes de substitution sensorielle sont très encourageants car ils permettent aujourd'hui de localiser et reconnaître des objets à partir d'images avec des règles de conversion très simples. Ils ne sont toutefois utilisables que pour reconnaître des motifs très simples et ne peuvent donc être utilisés comme outils de suppléance dans des environnements naturels (Auvray et al., 2007).

Discussion

Tous ces projets ont un même objectif fondamental : substituer la vision par une autre modalité sensorielle. Depuis le TVSS, beaucoup d'études ont montré qu'il est possible de localiser, mais aussi de reconnaitre des objets ou des formes autour de soi. Les premiers dispositifs de substitution sensorielle étaient tactiles sur le dos, sur le torse, au bout du doigt, sur la langue puis sont arrivés les dispositifs à restitution auditive.

Reconnaissance d'objets

Les premières expérimentations sur ces systèmes ont montré un premier résultat fondamental sur la perception et la reconnaissance de formes par les sujets : il n'y avait pas

de perception sans action sur la caméra. En effet, sur le premier dispositif de Bach-y-Rita, la caméra était fixe et la discrimination d'objets très simples était très difficile. En revanche, si la caméra était portée par l'utilisateur et qu'il pouvait balayer la scène visuelle avec la caméra portée dans sa main, alors la perception et la reconnaissance de formes était possible.

Un des principaux freins à ces systèmes de suppléance tactile est la différence entre la résolution nécessaire pour percevoir quelque chose dans une image et la résolution de la modalité sensorielle cible. L'idée ensuite poursuivie par Paul Bach-y-Rita est qu'il est possible de donner plus d'information sur la langue des sujets parce que la résolution tactile de la langue est plus élevée mais aussi qu'elle baigne dans une solution conductrice : la salive. Les expérimentations ont permis de montrer qu'il était possible de discerner des formes géométriques à l'aide de ce système mais, comme avec la version sur la peau du dos ou du torse, la faculté des utilisateurs à reconnaître des formes reste très faible pour une utilisation dans des environnements complexes.

Les résultats sont identiques quant à l'utilisation de systèmes de substitution sensorielle vision-audition : il est seulement possible de discriminer des objets parmi une dizaine, tous présentés sur un fond blanc et le temps de reconnaissance est souvent beaucoup trop long pour être utilisé quotidiennement.

Localisation d'objets

La localisation en 3 dimensions est possible par balayage de la caméra portée sur la tête mais elle reste imprécise et lente. L'idée poursuivie par le projet Espacio Acustico Virtual rajoute un degré d'analyse de scène en calculant les coordonnées 3D de chaque pixel présent dans la scène et en les restituant par des sons 3D. La méthode revient à disposer des petites enceintes à la surface de l'objet pour que l'utilisateur reconnaisse le motif sonore mais aussi le localise. Cette approche permet pour des volumes simples sur un fond blanc, de les localiser en utilisant les capacités de localisation des sons. Cette approche présente le principal avantage de restituer une information tridimensionnelle de manière directe. On ne sait pas en revanche quelle est la capacité des sujets à reconnaître des formes avec sons synthétisés à leur surface.

Faible complexité des algorithmes de conversion

Le principal avantage de ces différentes techniques est le temps de calcul puisqu'il est possible de faire fonctionner le système sur des ordinateurs dotés d'une très faible

puissance de calcul. Ces algorithmes peu coûteux permettent en effet de les porter sur des appareils embarqués comme des téléphones portables ou des PDA. Une version du dispositif 'The vOICe' est d'ailleurs disponible sur plusieurs plateformes et permet une utilisation en temps réel du système.

Conclusion

L'état de l'art sur les systèmes de substitution sensorielle montre que cette approche permet d'acquérir une somme d'information trop importante pour pouvoir être interprétée aisément : si la catégorisation d'objets restreinte à une dizaine de cibles peut fonctionner dans des conditions de laboratoire, il est en revanche extrêmement difficile de reconnaître des objets complexes dans des scènes naturelles. En effet, aucune modalité sensorielle autre que la vision ne permet d'interpréter facilement un signal aussi complexe. La résolution de la modalité sensorielle utilisée pour la substitution est très inférieure à la résolution des yeux ou des caméras. C'est la raison pour laquelle des projets de recherche ont tenté de diminuer l'information à transmettre à la modalité de restitution en recourant à une interprétation de la scène visuelle et non à un simple filtrage de celle-ci. Cette interprétation est possible grâce à des algorithmes d'analyse d'images plus complexes pour n'en restituer qu'une infime partie, à un degré d'abstraction plus élevé, donc facilement utilisable. C'est en cela que le projet EAV est particulièrement innovant en prétraitant les informations visuelles par des algorithmes de stéréovision pour reconstituer l'environnement 3D. Les progrès effectués ces dernières années dans les algorithmes d'analyse d'images et de reconnaissance de formes permettent d'aller encore plus loin dans l'analyse préalable de la scène visuelle en adaptant les informations à restituer en tenant compte des besoins de l'utilisateur.

Les systèmes d'augmentation sensorielle

L'objectif des systèmes d'augmentation sensorielle n'est pas de substituer la vue aux non-voyants par une autre modalité sensorielle mais de restituer certaines fonctions du système visuel parmi les plus utiles aux non-voyants. Les systèmes d'augmentation sensorielle sont aujourd'hui encore très peu développés. La baisse des coûts d'acquisition des télémètres a pourtant permis la conception de détecteurs d'obstacles qui sont aujourd'hui utilisés par les non-voyants avec des retours très positifs. Les signaux visuels provenant d'une caméra représentent un verrou scientifique plus complexe du fait de la complexité du traitement du

signal à effectuer. Ces indices pourraient permettre aux personnes non-voyantes de se représenter l'espace (Pissaloux et al., 2004).

Les télémètres

Les approches basées sur les télémètres visent à restituer la détection d'obstacles, qui est une des fonctions du système visuel les plus importantes pour la navigation. Cette restitution des informations ne correspond pas à une substitution sensorielle puisque le signal de distance à l'objet qui est restitué n'est pas présent dans l'image brute. Nous parlerons donc ici d'augmentation sensorielle. Dans ce cas, une modalité sensorielle reçoit une information de distance qui n'est pas présente immédiatement dans l'image.

Il existe principalement 2 manières d'estimer avec précision la distance à des objets environnants : les télémètres à ultrasons et les télémètres laser. Un signal ultrasonore ou lumineux est envoyé en direction de la surface dont la distance veut être évaluée. Cette surface renvoie le signal à l'émetteur qui calcule ensuite le déphasage du signal pour en estimer la distance. Le signal ultrasonore a une limite d'utilisation d'environ 20 m et la précision de l'estimation de la distance est très dépendante des facteurs environnementaux. Les télémètres laser ont une portée d'environ quelques centaines de mètres, sont très directionnels et plus précis mais beaucoup plus chers et ils ne sont pas utilisables dans des milieux que la lumière traverse trop facilement (une vitre par exemple).

La société Sound Foresight ® (www.soundforesight.co.uk) commercialise déjà des cannes à ultra sons mises au point à l'université de Leeds en Angleterre. Des boutons situés sur la canne permettent à la personne déficiente visuelle qui l'utilise de sentir l'intensité des ultrasons réfléchis. Un signal rapide et puissant signifie ainsi que l'obstacle est proche. Sur le même principe, la canne laser Télétact (Farcy and Damaschini, 2000), développée par René Farcy (laboratoire Aimé Cotton) et Yacine Bellik, (Université Paris-Sud à Orsay), a pour objectif de détecter les obstacles lors du balayage horizontal de la canne par l'utilisateur. Dès que le boîtier pointe vers un obstacle, la distance est calculée et transformée en stimuli tactiles sur 4 doigts représentant des plages de distances. Les auteurs (Farcy et al., 2003) mentionnent que l'expertise acquise sur ce système permettrait même de reconnaître certaines formes par une exploration de celles-ci en les balayant avec le faisceau.

Figure 18 : Illustrations des cannes à ultra sons de sound foresight (à gauche) et canne laser Télétact (à droite). La restitution est tactile pour les deux dispositifs.

Différentes étapes d'apprentissage sont nécessaires avant de devenir un « bon utilisateur » du Télétact. Les utilisateurs débutants commencent leur apprentissage avec un appareil qui se fixe sur leur canne blanche : le tom pouce. Cet appareil est un télémètre laser qui restitue de manière tactile la distance aux obstacles, pour une distance inférieure à 4 m. Après 2 à 3 mois d'apprentissage, les utilisateurs peuvent apprendre à manipuler le Télétact : sur le même principe que le TomPouce mais avec une restitution beaucoup plus fine et une distance de détection des obstacles d'une quinzaine de mètres. Ce dispositif permet ainsi de se faire une idée des volumes, des profils à des grandes distances et de se créer une représentation succincte de l'environnement par balayages successifs. La restitution est toujours tactile dans ces deux premières phases et c'est seulement lorsque l'utilisateur s'est familiarisé avec les deux premières étapes qu'il peut apprendre à utiliser le Télétact en restitution auditive (32 notes de 0 à 15 m, plus la note est aigüe, plus l'obstacle est proche). Cette modalité de restitution est, selon les créateurs, beaucoup plus efficace pour se représenter l'espace mais aussi beaucoup plus difficile à apprendre. Il faut environ 6 mois de formation pour utiliser le Télétact de manière efficace. Cette version augmentée de la canne blanche traditionnelle est un outil particulièrement apprécié et utilisé par les personnes non-voyantes. Ce genre d'outils semble aujourd'hui être le plus mature des outils d'aide électronique au déplacement pour les personnes déficientes visuelles. L'ensemble de ces outils n'aide en revanche pas les utilisateurs à s'orienter dans des environnements inconnus : il n'est pas possible de lire des noms de rue ou des panneaux par exemple. De nombreux systèmes tentent aujourd'hui de répondre à ce besoin d'orientation.

Les systèmes de reconnaissance et de localisation d'objets par vision artificielle

Le traitement des signaux provenant d'un capteur de vision artificielle est très complexe et très coûteux pour en extraire une information qualitative sur les éléments composant l'image. Il existe peu de systèmes d'augmentation sensorielle basés sur la vision artificielle. Deux catégories de systèmes de reconnaissance et de localisation d'objets existent : ceux qui nécessitent au préalable un marquage de l'environnement (codes-barres par exemple) et ceux qui reconnaissent directement les objets par des algorithmes de vision artificielle.

Par marquage préalable de l'environnement

Cette catégorie de systèmes requiert la présence de signes distinctifs facilement reconnaissables sur les objets d'intérêts. Ce marquage peut être invisible à l'œil nu si la discrétion est nécessaire (Makino et al., 1998). Le marquage de chaque objet à reconnaitre peut s'avérer rapidement fastidieux et l'outil n'est plus utilisable dès que la personne se déplace dans un environnement qui ne serait pas marqué.

Par reconnaissance et localisation d'objets

Une approche plus moins contraignante mais également plus complexe du point de vue algorithmique est donc de reconnaitre les objets directement sans avoir à les marquer. La plupart des approches de reconnaissance et de localisation d'objets et d'obstacles utilisés pour guider les non-voyants ont longtemps été dominées par les algorithmes de « pattern matching », c'est à dire par confrontation d'un motif avec l'image dans laquelle il est recherché. Le domaine de l'analyse d'images a cependant fait d'énormes progrès ces dernières années et la plupart des algorithmes modernes sont maintenant implémentés et intégrés à la librairie Open Computer Vision maintenue par Intel.

Un des premiers dispositifs à utiliser des techniques modernes de reconnaissance d'objets a été développé à Stuttgart (Hub et al., 2007) par Andreas Hub. Ce système de suppléance a pour objectif d'orienter les personnes non-voyantes en intérieur, là où les systèmes de géo-positionnement classiques ne sont pas opérants. Pour cela, le prototype comporte un dispositif de capture d'images en stéréoscopie permettant de connaître les coordonnées tridimensionnelles des éléments visuels de l'image.

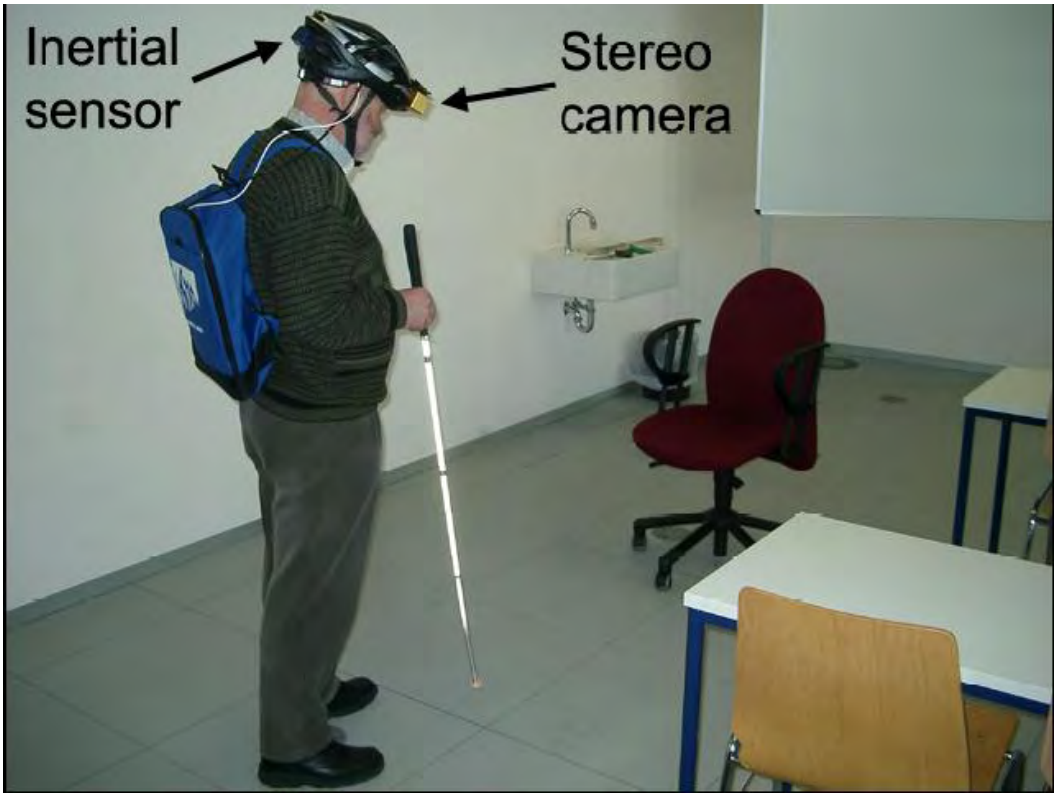


Figure 19: Illustration du dispositif d'aide à la navigation en intérieur développé à Stuttgart (Hub et al., 2007). Le sujet porte un casque sur lequel est disposé un capteur de vision stéréoscopique et un capteur inertiel (Xsens MT9B ; 3 axes). Un petit ordinateur porté dans le dos du sujet effectue les calculs.

L'objectif de ce système est d'augmenter l'environnement par la modalité auditive avec des informations utiles au déplacement. Pour cela, la position de l'utilisateur à l'intérieur d'une pièce est déterminée par triangulation wifi (positionnement en intérieur) et le mobilier de chaque pièce est modélisé en 3D. Le modèle contient des objets mobiles (ex. Chaise), semi-mobiles (ex. une porte : elle peut être ouverte ou fermée) ou fixes (ex. un lavabo). Ce projet est dédié à l'analyse de la scène en intérieur mais ne traite pas comment restituer l'information.

L'algorithme de reconnaissance d'objets utilisé est basé sur une détection de contours puis sur une comparaison avec l'histogramme des couleurs du modèle appris au préalable. L'implémentation de ces deux algorithmes est tirée de la librairie OpenCV évoquée plus haut. L'analyse de la scène permet de localiser des objets mobiles, ou des objets semi-mobiles (une porte par exemple). Les auteurs expliquent différentes manière d'utiliser les informations de reconnaissance et de localisation d'objets. Connaissant la position de l'utilisateur et la position réelle de la porte, si la porte est détectée plus proche de l'utilisateur que dans le modèle où elle est fermée, alors cette porte doit être ouverte (Figure 20).

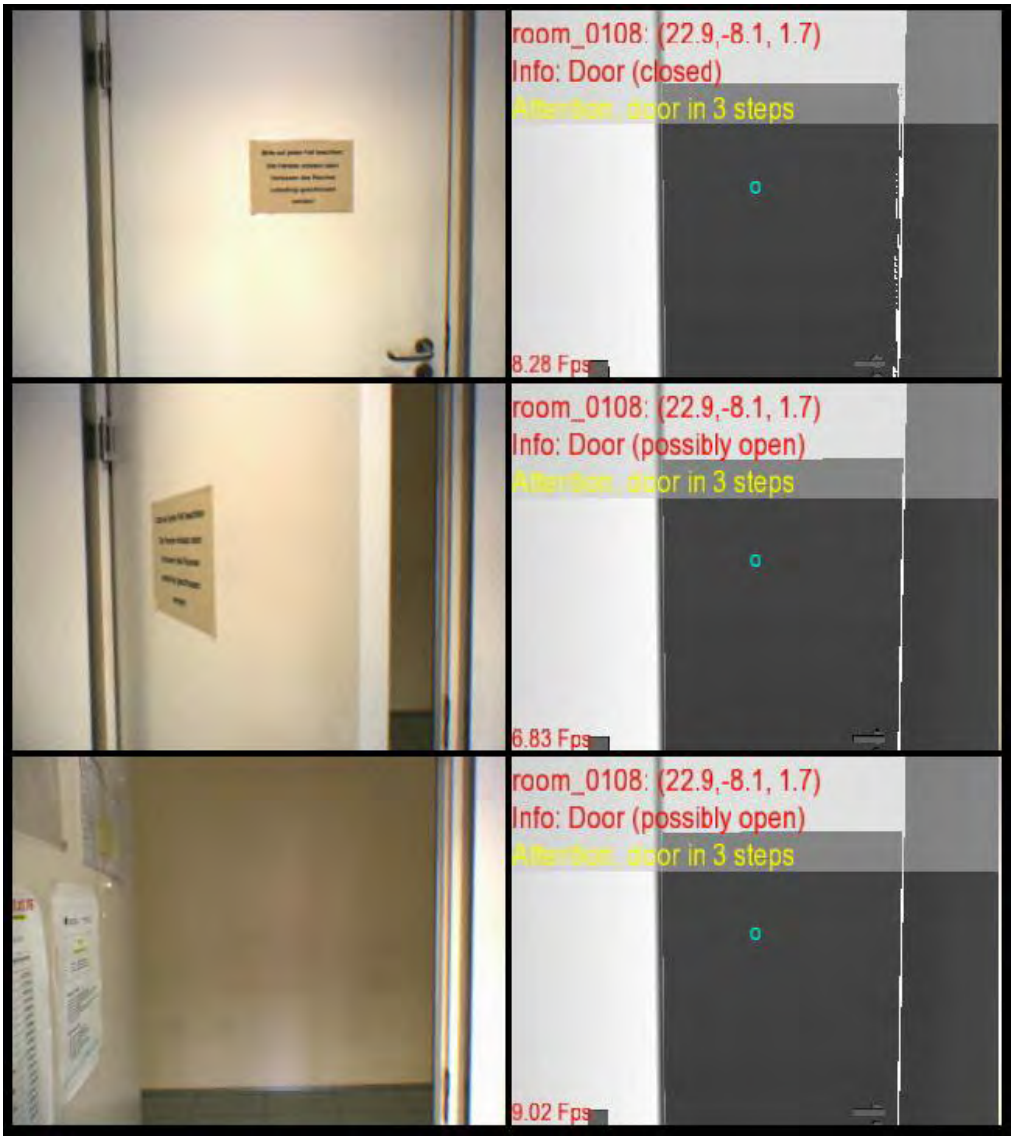


Figure 20 : Illustration de l'utilisation du dispositif développé à Stuttgart pour déduire si la porte est ouverte ou fermée. En haut, la distance à la porte est cohérente avec la distance prédite dans le modèle de la salle en fonction de la position de l'utilisateur (déterminée par triangulation Wifi). Au milieu cette distance est inférieure et en bas, la porte n'est pas détectée, elle est donc probablement ouverte.

Les auteurs ont utilisé un troisième algorithme de la librairie OpenCV pour détecter la présence de visages dans l'image. Cet algorithme permet de reconnaître rapidement des visages et d'en fournir les coordonnées 3D. La reconnaissance d'objets prend environ une à deux secondes tandis que le système de vision peut suivre un objet déjà reconnu à une fréquence d'environ 10 images par seconde (illustration de la Figure 20) et d'environ 5 images par seconde pour la reconnaissance de visages dans une image de résolution 320x240 sur un microprocesseur basse consommation cadencé à 1,6 Ghz. Cette étude apporte une importante discussion autour de la manière d'analyser de manière qualitative une scène visuelle. Les algorithmes utilisés, en grande partie basés sur la couleur, ne

peuvent pas être utilisés dans des environnements où la luminosité peut beaucoup varier mais aussi avec des objets 3D complexes dont le contour va beaucoup changer d'une vue à l'autre. L'algorithme de reconnaissance de visages est basé sur une méthode décrite initialement par Viola et Jones (Viola and Jones, 2001). Son principe est de décomposer une image en primitives de Haar et d'utiliser une cascade d'algorithmes du moins coûteux au plus coûteux en terme de temps de calcul. Chacun d'eux permet de restreindre le domaine d'action des suivants. Les primitives de Haar sont largement étudiées car elles permettent de décrire une image ou un modèle sous la forme de primitives simples, réduisant la résolution d'une image contenant des pixels à une description de l'image par des primitives simples.

Ces différents algorithmes sont implémentés dans la librairie OpenCV et permettent de faire de la reconnaissance et de la localisation d'objets. Le principe est de fournir beaucoup d'images d'un même objet à l'outil d'apprentissage pour que celui-ci apprenne à reconnaître les motifs visuels. Cette méthode n'est pas invariante au rapport d'échelle et à la rotation. Elle permet d'apprendre puis de reconnaître très rapidement des objets génériques (visages par exemple). Les algorithmes basés sur des relations d'invariance spatiale entre points d'intérêts d'une image et d'un modèle à reconnaître sont suivis de très prés depuis leur publication en 1999 (Bay et al., 2008;Lowe, 1999). Le principal atout de ces algorithmes est qu'ils sont invariants au facteur d'échelle et à la rotation contrairement aux algorithmes de convolution classiques d'un modèle dans une image. En effet, ces derniers ont une tolérance à la rotation et au facteur d'échelle qui est faible et pour pouvoir reconnaître un objet dans différentes situations, il faut donc créer beaucoup de modèles. Chaque modèle séparément est très rapide à corréler, mais la multiplication des modèles peut rapidement grever le temps de calcul. Les techniques s'appuyant sur l'extraction de caractéristiques invariantes comme SIFT et SURF tirent donc leur avantage de leur tolérance et de la nécessité de tester un seul modèle pour toutes les orientations et facteurs d'échelles. Un des principaux avantages de ces algorithmes par rapport aux autres est leur indépendance au capteur (caméra) puisque d'un capteur à l'autre, la relation spatiale entre les points d'intérêt est préservée. Ils ne sont en revanche pas du tout adaptés pour reconnaître des objets en 3D puisque le motif visuel d'un objet peut beaucoup changer d'une orientation à l'autre. Dans ce cas, ces algorithmes deviennent très coûteux car chaque face à reconnaître doit alors être apprise puis testée.

Ces algorithmes ont été implémentés dans un système de reconnaissance et de localisation d'objets pour les non-voyants (Cheng et al., 2008). L'objectif de ce système était de reconnaître des objets dans un flux capturé par une caméra (20 degrés d'angle de vue, 320x240). L'algorithme utilisé était SIFT, proposé par David Lowe en 1999 (Lowe, 1999). Les coordonnées projetées sur un plan à une distance fixe sont calculées en fonction des paramètres de la caméra, sa résolution et son angle de vue. L'utilisateur supervise la reconnaissance d'objets par reconnaissance vocale. Les instructions de direction pour le guidage peuvent être restituées par synthèse vocale ou par des sons spatialisés.

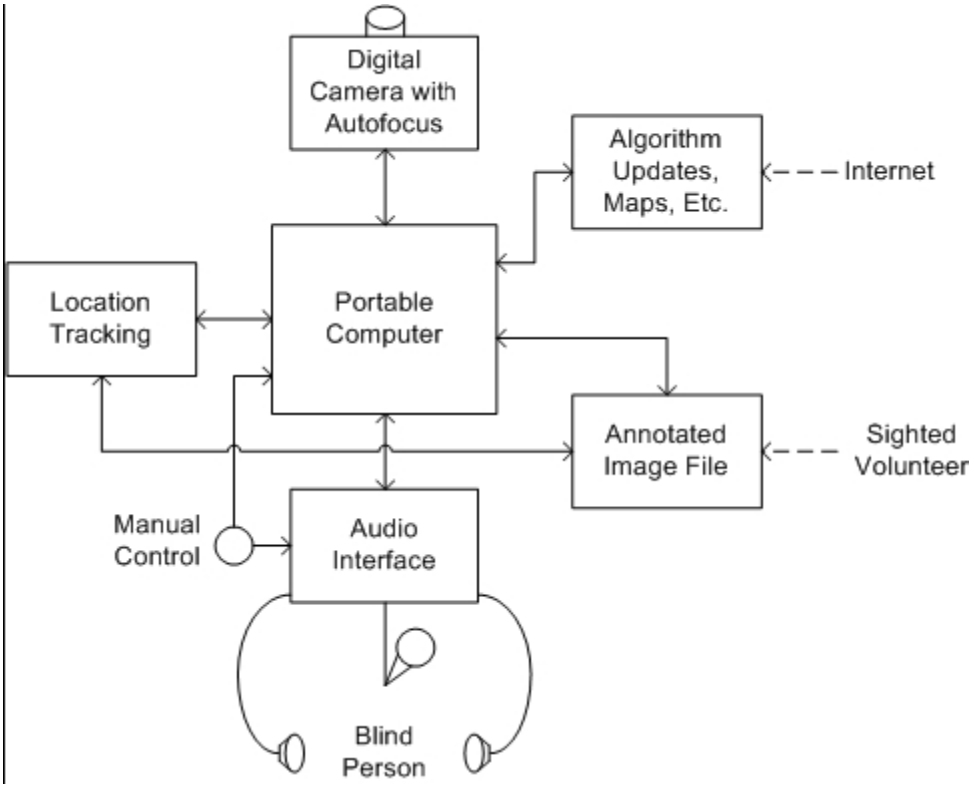


Figure 21 : Illustration du fonctionnement du dispositif créé à l'université du Maryland (Cheng et al., 2008). Une image est capturée, traitée par des algorithmes de reconnaissance d'objets (SIFT) qui sont alors replacés dans un plan distant. L'interface auditive guide les utilisateurs et leur indique la position des objets par un son spatialisé.

Deux modes de restitution ont été étudiés : dans le premier, la restitution de la position de la cible visuelle est effectuée en émettant dans un casque audio un son continu latéralement spatialisé avec la hauteur du son indiquant l'élévation. L'intensité du son devenait de plus en plus élevée avec le recentrage de l'objet dans l'image. Les auteurs mentionnent que le son continu est obligatoire parce qu'avec un simple son court, les utilisateurs ne pourraient pas localiser la source sonore avec précision. Dans une seconde méthode de restitution dite

‘discrète’, les sujets étaient guidés par synthèse vocale avec des instructions comme « haut 20°, droite 4° ».

Des tests préliminaires ont été effectués auprès d’un sujet non-voyant dans une tâche de localisation d’un logo positionné devant lui. Les premiers résultats montrent qu’il est possible quand l’objet est proche, de localiser un objet en moins de 10 secondes avec l’interface restituant des sons spatialisés continus et 16 secondes avec l’interface discrète (synthèse de parole). Les résultats préliminaires de cette étude sont très encourageants bien que le système n’ait pas encore été testé dans des situations moins contrôlées et sur une population plus importante. L’algorithme utilisé, invariant au facteur d’échelle et à l’orientation semble être performant pour reconnaître à la demande un objet précis.

Conclusion

L’état de l’art des algorithmes de vision artificielle fait apparaître que de nombreux algorithmes de reconnaissance d’objets peuvent être utilisés, l’enjeu pour les systèmes de suppléance étant de les utiliser chacun au maximum de leur potentiel en fonction des situations dans lesquelles ils sont les mieux adaptés. Les méthodes de corrélation entre un modèle et une image par convolution du modèle dans l’image (ou d’un modèle simplifié dans une image simplifiée, par réduction de la résolution ou par transformation des deux images en primitives plus simples) sont très peu coûteuses en temps de calcul par rapport aux autres méthodes. Elles présentent le principal inconvénient d’être très peu tolérantes aux changements d’orientation ou d’échelle. Elles sont particulièrement adaptées à la reconnaissance d’objets qui seront toujours dans la même orientation (ex. mobilier urbain) et dont l’échelle dans l’image varie peu (objets lointains par exemple, puisqu’une variation de la distance à l’objet aura peu d’incidence sur la taille de l’objet dans l’image). De plus, ces méthodes très efficaces en temps de calcul permettent de modéliser des objets en 3D en créant des modèles rapides à détecter pour chaque face de l’objet. Les méthodes de reconnaissance et de localisation invariantes à la rotation et au facteur d’échelle sont en revanche beaucoup plus lentes mais plus robustes. En parallèle de la recherche en analyse d’images, les constructeurs mettent de plus en plus à disposition des développeurs des outils de développement sur processeur graphique, qui sont potentiellement bien plus rapides pour effectuer ce type de calculs. Un portage de ces différents algorithmes de reconnaissance d’objets dans une scène visuelle en utilisant ces ressources permettrait d’atteindre des performances beaucoup plus élevées, même sur des plateformes mobiles.

5) Les aides à l'orientation

Dans des environnements inconnus avec comme seules informations les repères visuels, les voyants s'orientent en lisant les panneaux indicateurs, les plaques de rues, en repérant les arrêts de bus, de métro, etc., présents au cours de leur déplacement, c'est-à-dire dans un repère égocentré. Ces indices visuels peuvent être intégrés pour générer une représentation allocentrée de l'espace, dans laquelle la description des éléments est indépendante d'une position relative. La plupart des systèmes d'aide à l'orientation pour les non-voyants fonctionnent aujourd'hui sur la base d'un capteur de géolocalisation, avec des informations de positionnement absolu.

Systèmes d'aides basés sur des capteurs de géolocalisation satellitaire

Géopositionnement

L'arrivée des systèmes de géolocalisation avec le premier satellite expérimental en 1978 puis une constellation de 24 satellites exploitables dès 1995 a fait de ces systèmes un outil incontournable pour le guidage des personnes dans des environnements inconnus. Bon nombre de systèmes de guidage, principalement pour l'automobile, se sont aujourd'hui démocratisés et ont envahi notre quotidien. Si aujourd'hui le système de géolocalisation le plus utilisé est une solution américaine (GPS), chaque Puissance politique tente de développer sa propre solution autonome. En Europe, le système Galileo est en test depuis 2004 mais a pris de nombreuses années de retard pour son lancement définitif. A terme, il est destiné à avoir une couverture et une précision supérieure au GPS. Les premiers récepteurs GPS introduits en 1995 avaient une erreur de positionnement d'environ 25 m en valeur absolue alors que l'erreur sur les récepteurs actuels est d'environ 10m grâce à différents algorithmes de traitement des données acquises par satellite. En couplant le capteur GPS à d'autres capteurs d'attitude (accéléromètres, boussoles, podomètres…), il est possible d'atteindre une erreur inférieure à 5 m 95% du temps.

Les méthodes de GPS différentiel permettent d'atteindre une précision encore bien plus importante en corrigeant les sources majeurs d'erreurs causées par les fluctuations météorologiques. Des stations au sol dont la position est connue calculent en permanence les corrections relatives à appliquer au signal GPS pour faire correspondre leur position réelle avec la position calculée. Ces informations de correction sont ensuite diffusées par radio pour être utilisées par les récepteurs alentours. Cette méthode est donc plus complexe

et coûteuse que le GPS seul puisqu'elle nécessite de positionner des stations émettrices au sol. De plus, les informations émises ne sont pas toujours disponibles dans les environnements urbains, ce qui diminue considérablement leur utilité pour l'aide au déplacement des piétons.

Systèmes de guidage pour les déficients visuels

Sur ce principe, de nombreux projets de recherche (Helal et al., 2001;Loomis et al., 1994;Ran et al., 2004) et de systèmes commerciaux utilisent une géolocalisation basée sur un récepteur GPS (Kapten, Loadstone GPS, BrailleNote GPS, Trekker, Blind navigator, Angeo, Mobile Geo) (voir Figure 22 et Tableau 1). Le choix d'un système GPS commercial se fait selon son expertise et son besoin. En effet, la précision et le mode de guidage sont assez semblables. Ils permettent de connaître sa position, les lieux environnants et de planifier un trajet. C'est en revanche leur implémentation qui est différente, sur des appareils dédiés (BrailleNote, Trekker) ou pour plus de flexibilité des solutions logicielles sur téléphone portable (Mobile Geo). La plupart sont très chers (1000-2000€) et ne répondent pas de manière adaptée aux besoins des non-voyants en orientation et encore moins à leurs besoins d'aide au déplacement. Ces systèmes ne sont pas fiables pour une utilisation par des piétons, qui ont besoin d'une information d'orientation très précise. Si la plupart des constructeurs affichent aujourd'hui des précisions de l'ordre de 5 m 95% du temps, les canyons urbains se révèlent être un verrou technologique important pour la navigation du piéton. Nombreux sont les systèmes pour personnes non-voyantes, qui ont juste été adaptés depuis des systèmes pour voyants si ce n'est pour véhicule. Aucun d'entre eux ne permet de planifier un itinéraire pour piéton en prenant en compte les spécificités des personnes non-voyantes pour la traversée des rues et des carrefours. Ces manques sont dus à deux principaux problèmes : il n'existe pas de base de données cartographique pour piéton ; et si c'était le cas, la précision des systèmes de géolocalisation actuels (5 m) ne serait de toute façon pas suffisante pour les guider.

Trekker, 2002

Kapten, 2008

Angeo, 2010

BrailleNote GPS, 2002

Figure 22 : Projets commerciaux matériels pour la navigation du piéton. Ces dispositifs sont utilisés par les non-voyants et ont été conçus pour les non-voyants hormis le kapten qui est dépourvu d'écran et dont l'interface non-visuelle se révèle très efficace pour les non-voyants.

Tableau 1 : Récapitulatif des différents systèmes d'aide à la navigation par GPS utilisables par les non-voyants

Nom	Prix	Dédié aux non-voyants	Solution	Commande	Restitution
Kapten, 2008	130 €	Non	Matérielle	Vocale	Auditive, TTS
Loadstone GPS, 2006	Gratuit	Non	Logicielle	Clavier	Lecteur d'écran
WayFinder Access, 2003	270€	Oui	Logicielle	Clavier	Auditive, TTS (lecteur d'écran)
BrailleNote GPS, 2002	1700€	Oui	Matérielle	Clavier	Afficheur Braille/ TTS

Trekker, 2002	1600€	Oui	Matérielle	Clavier adapté	TTS
Blind Navigator	1500€	Non	Logicielle	Clavier	TTS
Mobile Geo	1000€	Oui	Logicielle	Clavier	TTS/Lecteur d'écran
Angeo	1500€	Oui	Matérielle	Vocale/Touches	TTS

Un exemple de système de guidage pour non-voyants avec fusion multi-capteurs et guidage auditif

Les systèmes classiques de guidage permettent néanmoins un réel gain d'autonomie pour les personnes non-voyantes (Zabihaylo et al., 2006), en réduisant significativement leur temps de navigation et la distance parcourue, bien qu'il soit aujourd'hui, difficile de les utiliser pour traverser une rue (Maeda et al., 2002) du fait de leur manque de précision. Certaines équipes se sont intéressées au couplage du GPS avec d'autres capteurs pour tenter d'en augmenter la précision. C'est dans cet esprit que l'équipe de Loomis en Californie a conçu un système basé sur un récepteur GPS permettant de géolocaliser une personne avec une précision consolidée de 5 mètres. Ils utilisent un compas en plus du GPS pour connaître l'orientation de la personne et ainsi lui indiquer vers où se diriger avec une plus grande précision. Un prototype a été mis au point au département de psychologie à l'université de Californie à Santa Barbara.

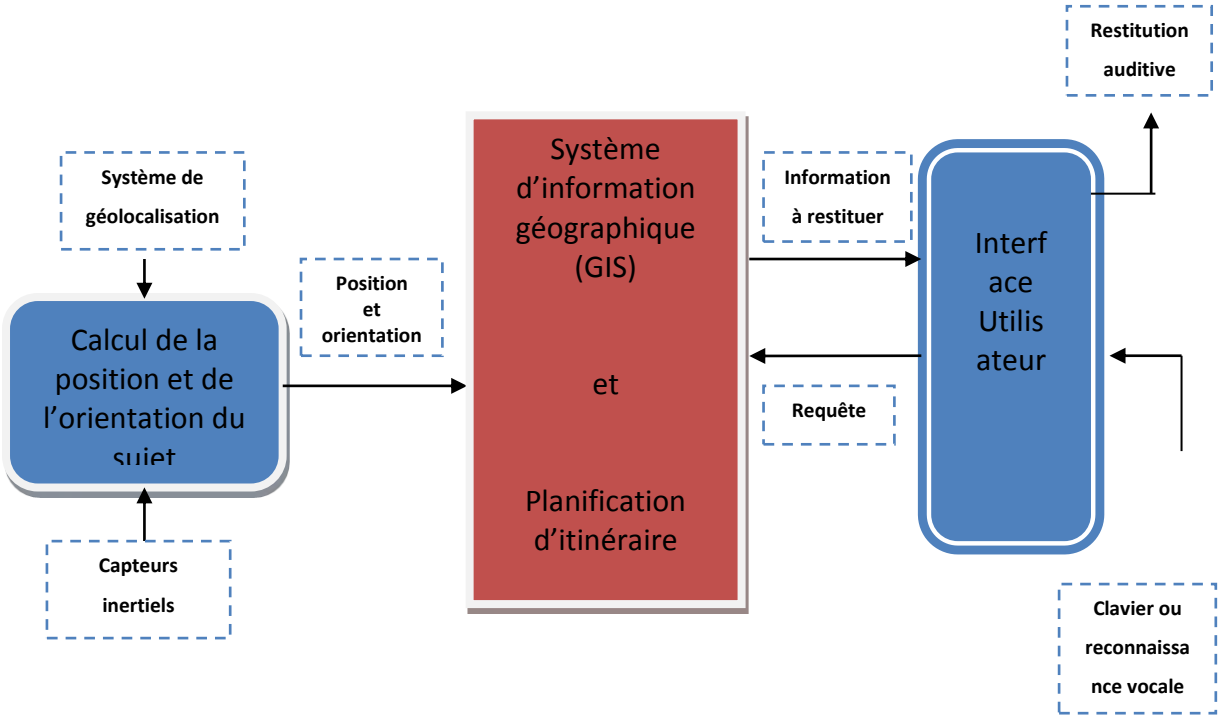


Figure 23 : Le système proposé par M. Loomis en 1998 est composé de trois principaux modules. L'interface utilisateur est pourvue d'un module de reconnaissance vocale (entrée) et d'une restitution auditive (sortie). Le cœur du dispositif traite les informations de position et d'orientation pour aider à la navigation en corrélant la position et l'orientation du sujet avec une base de données géographique. Des capteurs et systèmes de localisation par satellite (GPS) permettent de connaître l'orientation et la position du sujet.

Le dispositif décrit de manière détaillée l'environnement (monuments, bâtiments, etc., à condition qu'ils soient référencés) en les nommant avec un son présenté de manière binaurale (Loomis et al., 1998). Le problème évident de ce dispositif est le référencement des objets. En effet, il est très sensible au changement et nécessite d'actualiser une base de données considérable. Les objets mobiles ne peuvent pas être détectés et s'ils l'étaient, la précision du GPS (10 mètres environ sans capteur supplémentaire) ne serait pas assez bonne pour pouvoir avertir d'un petit obstacle.

Différents modes de guidage auditif (Loomis et al., 1994;Loomis et al., 1998) ont été étudiés en navigation avec le système 'Personal Guidance System'. Le système est équipé d'un outil de géolocalisation et d'une base de données cartographique comportant les routes mais aussi des informations sur les bâtiments publics. Le système d'aide à la navigation a permis de tester différentes méthodes de guidage et de description verbale de l'environnement en ne se basant que sur la cartographie en navigation. L'évaluation de ce dispositif a permis

d'évaluer les différents modes de restitution auditive. Le but était d'évaluer les capacités à suivre un trajet balisé en utilisant un guidage par synthèse binaurale (le son était synthétisé comme s'il provenait des balises), par un guidage avec les mots « gauche » et « droite ou par un guidage « gauche/droite » nuancé (exemple : « gauche 80 degrés »). Le dernier essai concernait le même dispositif que le précédent mais sans compas : l'orientation était déterminée par deux GPS situés au niveau des épaules. Les résultats ne montrent pas de différences significatives entre les différents modes de restitution auditifs. Une préférence est en revanche donnée aux sons spatialisés pour pouvoir donner une représentation spatiale de l'environnement.

Figure 24 : Personal Guidance System développé à Santa Barbara par Jack Loomis. Le système dont le schéma matériel est décrit dans la figure précédente présente les différents modules composant le système : des capteurs inertiels, un GPS, une interaction vocale en entrée et une synthèse sonore pour la restitution des informations.

Les systèmes d'aide à l'orientation pour les non-voyants présentent donc deux principaux verrous technologiques empêchant leur développement pour aider les non-voyants à se déplacer: les bases de données cartographiques peu précises et le manque de précision du capteur. La miniaturisation des capteurs d'attitude (capteurs inertiels, de position …) permettent aujourd'hui de prévoir et d'affiner une position erronée fournie par un système

de géolocalisation mais ces systèmes, même complémentés par des capteurs ne sont pas opérants en intérieur.

Géolocalisation dans des environnements où le GPS est peu fiable ou inopérant

Les centrales d'attitude ont fait leur apparition ces dernières années en mobilité grâce à la miniaturisation des circuits permettant de mesurer l'actimétrie d'une personne. Elles sont constituées par un ensemble de capteurs permettant de connaître l'état d'une personne : sa position, son orientation, l'orientation de sa tête, sa vitesse… Ces données sont primordiales pour étudier le comportement d'une personne mais aussi pour la guider. Les systèmes de géolocalisation actuels ne sont pas adaptés au piéton non-voyant car ils ne sont pas assez précis pour pouvoir guider un non-voyant jusqu'à un passage piéton ou une bouche de métro.

Des équipes se sont intéressées à la navigation dans les espaces dans lesquels le GPS est inopérant. Ces espaces sont cartographiés grâce à des balises RFID (Willis and Helal, 2005) ou des bornes Wifi/Bluetooth (temps d'aller retour sur le réseau (Hub et al., 2007)). Comme nous l'avons abordé précédemment, le principal verrou aux développements d'aides nécessitant un pré-équipement de l'environnement est leur coût d'installation mais surtout de maintenance. Plus le nombre d'émetteurs est important, plus la localisation est précise. Le coût de l'équipement intérieur en capteurs RFID représente la solution la moins coûteuse si l'on désire avoir une précision de quelques centimètres. Malgré cela, l'installation reste lourde et non standardisée dans les différents bâtiments, ce qui rend l'utilisation de ces systèmes faible aujourd'hui.

6) Discussion

Il existe de nombreux capteurs technologiques permettant de capturer de l'information sur une personne ou son environnement. Les systèmes de substitution sensorielle utilisent des capteurs de vision artificielle (caméras) pour restituer l'image capturée vers la modalité auditive ou tactile. D'autres capteurs non « visuels » peuvent aussi être utilisés dans des systèmes de suppléance. Dans ce cas, l'information restituée n'est pas une image au sens d'un motif visuel à interpréter mais une information qualitative sur la scène. On parle alors de système d'augmentation sensorielle.

Capteurs d'actimétrie et d'environnement

Le Tableau 2 synthétise les capteurs qui sont aujourd'hui utilisés et montre qu'il existe une multitude d'informations restituables, chaque modalité sensorielle pouvant interpréter plus ou moins bien chacune d'elles. L'ensemble de ces capteurs rassemble une grande quantité d'informations restituables pour des personnes non-voyantes et l'idée poursuivie dans ce manuscrit est que l'on peut concevoir des systèmes en considérant le capteur pour sa fonction et pas par analogie avec le système visuel défaillant. Ce sont les fonctions de la vision qu'une personne voyante utilise pour naviguer que nous souhaitons restaurer chez une personne qui n'en est pas ou plus pourvue.

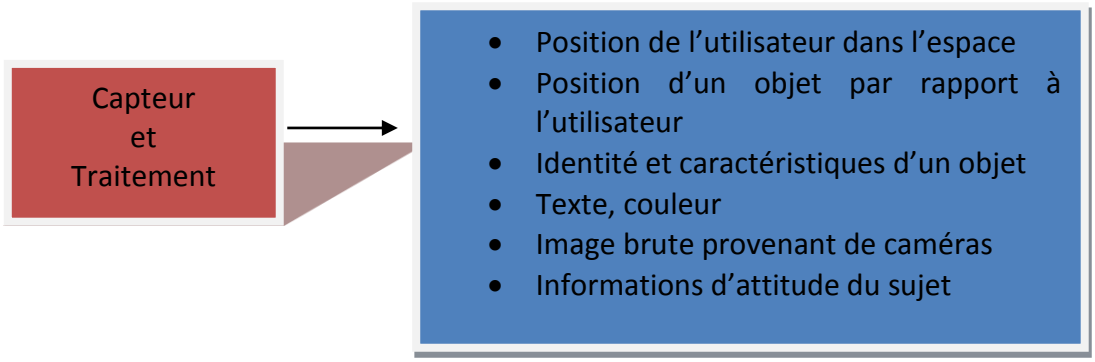


Figure 25: Informations qu'il est possible de restituer dans un dispositif de suppléance visuelle

De par sa nature, la vision humaine est tridimensionnelle : le système visuel humain est capable de localiser et catégoriser des cibles très rapidement (Thorpe et al., 1996;Thorpe, 2002). C'est pourquoi le fait de substituer la vision par des stimulations tactiles ne peut être direct : chaque modalité sensorielle a des propriétés qui ne sont pas bijectives quand elles sont mises en correspondance avec une autre modalité. Il est en revanche possible d'établir des liens relativement directs entre les fonctions des différentes modalités sensorielles. La localisation de cibles est par exemple commune entre l'audition et la vision puisque ces deux modalités réalisent une localisation 3D des objets sur la base des informations qui leur parviennent.

Tableau 2: Principaux capteurs d'entrée pouvant intervenir dans un outil de suppléance pour les non-voyants et informations qu'ils peuvent fournir après traitement.

Capteur	*Information en sortie*
Télémètre (Laser, ultrasons)	• *Distance relative à un objet* • *Précision de 2 mm 95% du temps pour une portée de <500 pour les plus précis* • *Fréquence : pas limitée*
Capteur de géolocalisation	• *Position absolue de l'utilisateur* • *Précision brute de 10 m, pouvant aller jusqu'à 5 m 95% du temps par hybridation avec d'autres capteurs. Quelques cm en GPS différentiel.* • *Fréquence : 1-100 Hz* • *Seulement en extérieur*
Capteurs RFID/Rubee/SIR/Wifi	• *Géolocalisation relative à la position des capteurs* • *Portée :* • *- 3-15 m pour RFID/Rubee/SIR* • *- Jusqu'à 300 m pour le wifi*
Capteur GSM	• *Géolocalisation relative à la position des antennes relais dont la position est connue.* • *Précision : 300 m en zone urbaine et 2 km en zone rurale*
Capteurs inertiels, d'orientation, compas,	• *Orientation d'un membre (tête, bassin…) par rapport au nord magnétique, à la verticale, à l'horizontale ; la vitesse, l'accélération* • *Fréquence : pas limitée*
Capteur(s) de vision artificielle	• *Image visuelle brute* • *Identité d'un objet* • *Position d'un objet* • *Lecture (OCR)* • *Couleur* • *Taille* • *Fréquence : fonction de l'algorithme de traitement utilisé*

Malgré le nombre important de dispositifs existants, peu sont finalement utilisés. Les systèmes de suppléance peuvent être classés en deux catégories : les systèmes de substitution et d'augmentation sensorielle.

Les systèmes de substitution sensorielle

Nous avons décrits les systèmes basés sur la substitution sensorielle comme étant très difficiles à utiliser en raison de la difficulté à interpréter les informations restituées. Les études ont cependant montré qu'il était possible avec ce type de systèmes, d'identifier des formes simples avec une matrice de stimulation tactile posée sur le torse, du moment que l'utilisateur lui-même pouvait déplacer librement le capteur. Ces travaux ont été répliqués sur la langue avec le TDU (Tongue Unit Display) (Bach-y-Rita et al., 1998) en utilisant une matrice électro-tactile de 7x7 puis 12x12 électrodes posée sur la langue. Ce système permettait d'identifier et de localiser des formes simples par des mouvements successifs. Le principal frein à l'utilisation de tels systèmes est la difficulté d'interpréter des scènes complexes par le biais de stimuli auditifs ou tactiles complexes. En revanche, tous ces travaux ont soulevé et soulèvent encore de nombreuses questions en Neurosciences liées à la substitution sensorielle (notamment sur la plasticité cérébrale et l'apprentissage).

Les systèmes d'augmentation sensorielle

Les systèmes d'augmentation sensorielle restituent une information permettant de s'orienter ou se déplacer à un degré d'abstraction supérieur aux systèmes de substitution sensorielle. En effet, l'interprétation de l'information restituée y est facilitée par un filtrage de l'information provenant du capteur. Une multitude de capteurs d'environnement et d'actimétrie existe mais aucun ne permet à lui seul de répondre aux besoins des non-voyants en matière d'autonomie dans le déplacement et d'orientation dans des endroits inconnus. Les aides électroniques aujourd'hui utilisées par les personnes non-voyantes sont basées sur des télémètres et répondent à un besoin spécifique pour la détection d'obstacles. Ces systèmes ne fournissent qu'une information très simple à l'utilisateur, ce qui accroît son utilisabilité. L'utilisation d'une caméra permet pourtant de préserver beaucoup plus d'information sur la scène visuelle. Le principal verrou scientifique à l'utilisation massive de la vision artificielle pour un dispositif de suppléance est la complexité des algorithmes de vision artificielle rendant difficile leur utilisation en temps réel.

Nous avons étudié jusqu'à maintenant les systèmes de suppléances non-invasifs capables d'augmenter l'autonomie des non-voyants en navigation. Les limitations pesant sur ces systèmes de suppléance résident principalement dans la trop faible résolution de la modalité de restitution par rapport à la quantité d'information à restituer. La difficulté de concevoir des systèmes de suppléance réside dans le choix de l'information que l'on souhaite restituer pour répondre à un besoin. Il conditionne la chaîne de traitement de l'information du capteur à la restitution. Les travaux d'analyse du signal provenant d'un capteur permettent d'analyser la scène visuelle et fournir de l'information à un module de restitution de l'information. Cette information peut être restituée de manière auditive ou tactile mais aussi visuelle en stimulant électriquement le système visuel humain dans une neuroprothèse.

Les neuroprothèses

Les progrès en neurosciences nous permettent aujourd'hui d'imaginer de nouvelles interfaces, directement connectées au cerveau. Il est possible d'évoquer des percepts en réponse à des stimulations électriques du système visuel. Ces percepts, les phosphènes, obéissent à des propriétés décrites depuis les premières implantations chez l'homme dans les années 1960 (Brindley and Lewin, 1968).

7) La stimulation du système nerveux

Le système nerveux est composé de neurones constitués eux-mêmes d'un corps cellulaire (ou soma, Figure 26) auquel sont connectées des ramifications dendritiques en entrée et un prolongement axonal, en général unique, en sortie. Lorsqu'un neurone est excité par l'intermédiaire de ses dendrites, une réponse appelée potentiel d'action peut apparaître à la base de son axone et se propager le long de celui-ci (Figure 26). La sommation des potentiels arrivant sur les dendrites conditionne l'excitation ou non du neurone. Si cette somme de potentiels dépasse le seuil d'excitabilité du neurone, un potentiel d'action va être généré au point d'insertion de l'axone et va ensuite se propager dans celui (Figure 27).

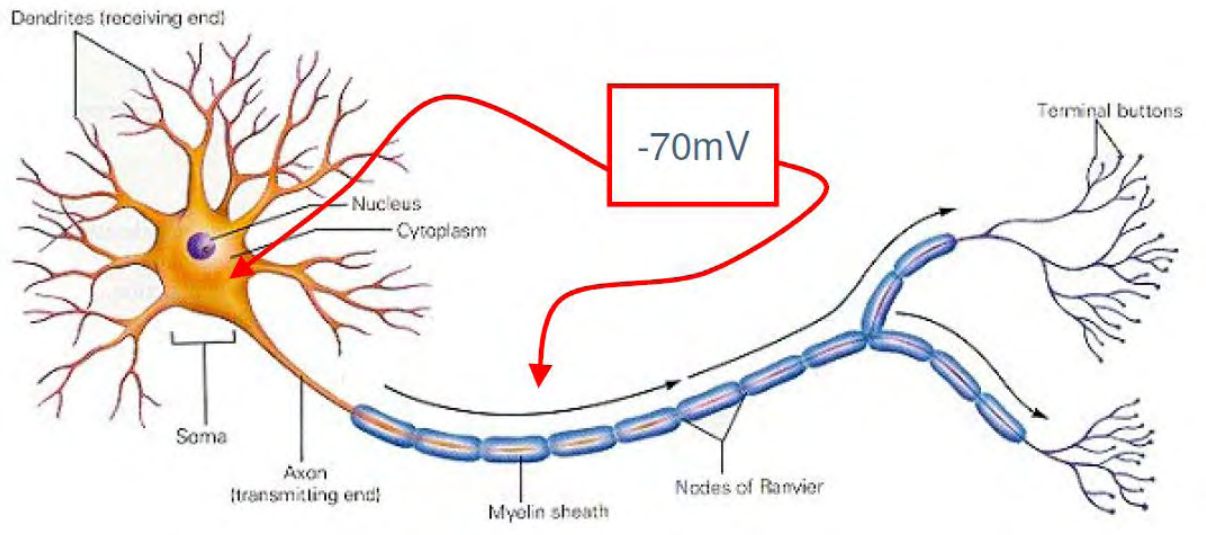


Figure 26: Schéma d'un neurone. Un potentiel d'action est initié au niveau du cône axonal, quand la somme des potentiels provenant des dendrites dépasse un seuil d'excitabilité. Ce potentiel d'action se propage tout le long de l'axone jusqu'aux terminaisons axonales. Le potentiel de repos des neurones est de -70 mV. Ce potentiel devient positif (+40 mV) lors du passage du potentiel d'action.

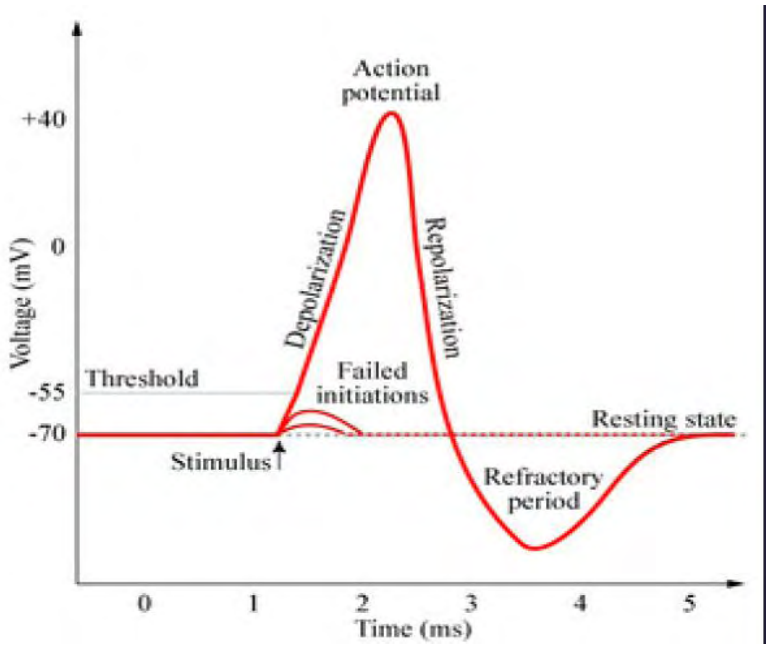


Figure 27: Un potentiel d'action est généré quand un neurone est excité. Le neurone passe par une phase de dépolarisation et d'inversion de potentiel puis d'hyperpolarisation de la membrane avant de revenir au potentiel de repos. Ces étapes sont suivies d'une période réfractaire pendant laquelle le neurone ne peut plus être excité. Le seuil d'excitabilité au-delà duquel un potentiel d'action est généré est d'environ -55 mV.

Il existe différentes manières d'interférer dans le fonctionnement du système nerveux en injectant de manière directe ou indirecte des courants électriques dans celui-ci.

Stimulation de surface : Stimulation Magnétique Transcrânienne

La TMS (Transcranial Magnetic Stimulation) est une technique utilisée en clinique pour le diagnostique de maladies neurologiques. Elle est cependant utilisée de plus en plus comme un outil de recherche en neurosciences. Cette méthode permet de stimuler des zones du cerveau depuis la surface du crâne, en appliquant une stimulation magnétique sur l'encéphale à travers le crâne au moyen d'une bobine métallique. Une variation rapide du champ magnétique (quelques µsecondes) induit un courant électrique perturbant le fonctionnement des neurones de la zone stimulée. Cette méthode permet d'évoquer des percepts visuels (phosphènes) en stimulant le cortex visuel (Kammer et al., 2005).

Figure 28: Illustration de la TMS (Source: Mayo Foundation for Medical Education and Research).

Le fait que la TMS ne nécessite aucune chirurgie en fait un outil de plus en plus étudié pour l'interaction homme-machine. Il n'est en revanche pas possible de stimuler avec précision une petite population de neurones dans une zone du cerveau bien définie puisque les courants électriques induits se propagent sur plusieurs centimètres, impliquant des millions de neurones et de nombreux circuits à chaque stimulation. Il faut utiliser des méthodes plus invasives pour stimuler l'encéphale de manière très précise.

Stimulation intracrânienne : électrodes de surface et intracorticales

La stimulation à l'aide d'électrodes directement en contact avec le système nerveux permet d'accroître la précision et de diminuer le courant électrique de stimulation nécessaire.

Les électrodes de stimulation de surface sont disposées à la surface du cerveau dans une zone que l'on souhaite stimuler. Brindley en 1968 (Brindley and Lewin, 1968) a été le premier à disposer une matrice de 80 électrodes de surface sur l'hémisphère droit, et plus particulièrement sur le cortex visuel.

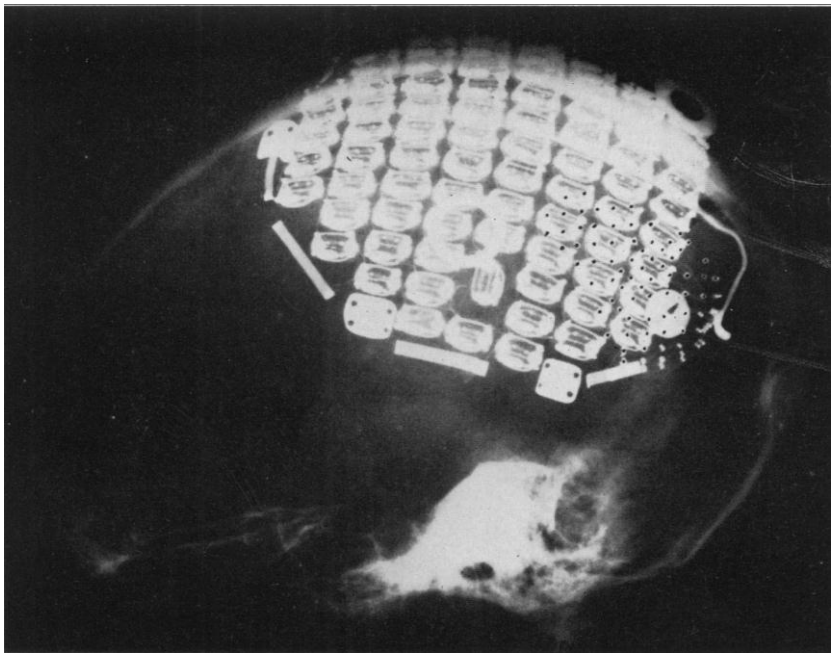


Figure 29: Radiographie d'une matrice d'électrodes posée sur l'hémisphère droit du cerveau d'un patient atteint d'un glaucome aux deux yeux (Brindley and Lewin, 1968)

Le sujet était un homme, 52 ans, myope depuis le jeune âge qui avait un glaucome aux deux yeux ayant abouti à une cécité de l'œil gauche.

En 1996, une autre étude a été faite en implantant 38 électrodes intracorticales à 2 mm de profondeur dans le cortex visuel d'une femme de 42 ans. Cette personne avait un glaucome depuis 22 ans et n'avait plus aucune perception de lumière (Schmidt et al., 1996). (Figure 30).

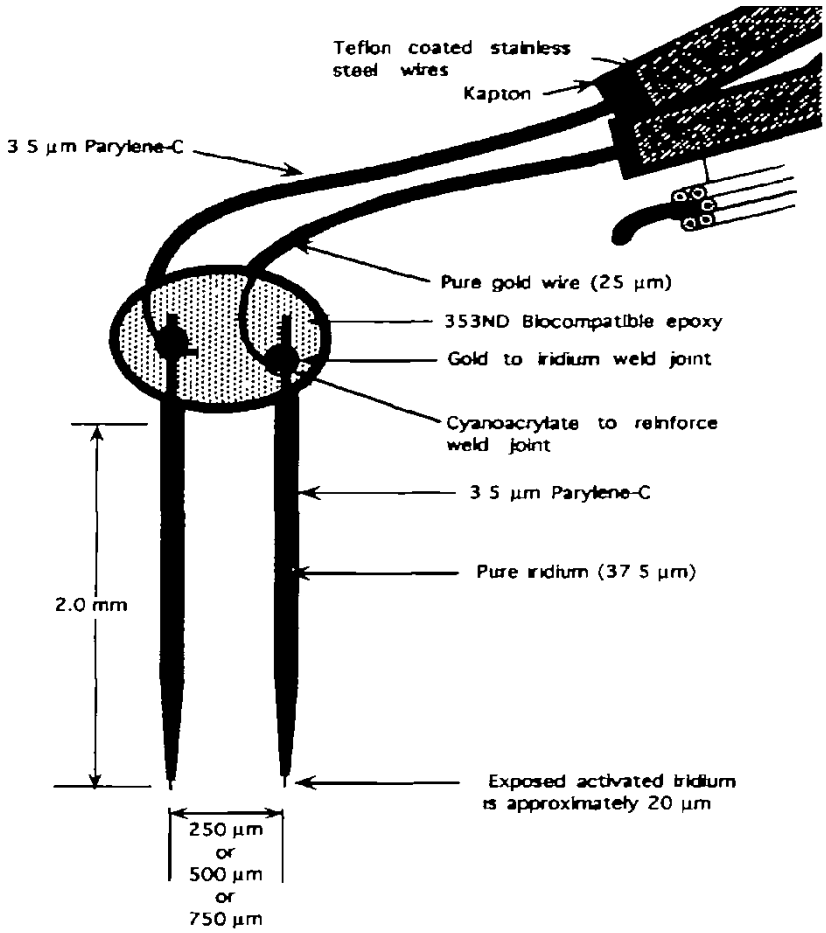


Figure 30: Électrodes de stimulation intracorticales de 2 mm de longueur et 20 micromètres de diamètre. Celles-ci pouvaient être espacées de 250, 500 ou 750 micromètres (Schmidt et al., 1996).

L'objectif de ces deux études était d'établir la faisabilité d'une neuroprothèse visuelle en établissant les règles de transformation d'une image en une série de stimulations caractérisées par leur position, leur intensité et leur fréquence. Dans les deux études, les sujets devaient décrire les percepts induits par les stimulations. Du fait de la cécité d'un seul œil dans l'étude de Brindley, le patient pouvait indiquer grâce à son autre œil la position précise du phosphène et le décrire verbalement. Dans l'étude de Schimdt, la position du percept était indiquée en plaçant une fléchette sur un jeu de fléchettes circulaire divisé en 5 cercles concentriques représentant l'excentricité du percept. Les sujets devaient essayer de garder le regard fixe. La position des percepts était indiquée de manière absolue (pointage) ou relativement à d'autres phosphènes (eg : à gauche du précédent). Dans les deux études une carte des percepts en fonction de la position de la stimulation pu être établie, sans qu'il soit possible de s'assurer de leur reproductibilité étant donné le faible nombre de sujets.

Il y a en revanche de nombreux résultats convergents ou complémentaires entre les deux études.Du centre jusqu'à 10° d'excentricité, les percepts étaient décrits comme des petits

points lumineux et au-delà de 10° comme des grains de riz ou des nuages (Brindley and Lewin, 1968). Les percepts étaient généralement uniques en partie centrale, mais pouvaient comporter jusqu'à une dizaine de percepts pour les excentricités élevées. Lorsque plusieurs percepts simultanés étaient présents pour une seule stimulation, le fait de diminuer le potentiel de la stimulation ne faisait pas diminuer le nombre de percepts mais parfois sa taille (Schmidt et al., 1996). Un potentiel de stimulation trop élevé (trois fois le seuil d'activation du percept provoquait des douleurs profondes dans le crâne (Brindley and Lewin, 1968). La plupart du temps, plusieurs stimulations simultanées évoquaient plusieurs phosphènes identiques à ceux générés par chaque stimulation indépendamment. Les percepts cessaient d'exister dès l'arrêt de la stimulation (Brindley and Lewin, 1968;Schmidt et al., 1996). L'augmentation du temps de stimulation induisait une augmentation de la luminance des percepts (Brindley and Lewin, 1968), ils devenaient également plus plaisants et mieux reconnaissables (Schmidt et al., 1996) et le seuil d'activation minimal (potentiel nécéssaire à la perception d'un phosphène) était plus faible (Brindley and Lewin, 1968). La fréquence de stimulation ne semble pas avoir d'influence sur le percept (entre 200 et 2000 Hz) pour des électrodes de surface (Brindley and Lewin, 1968) alors qu'elle influe sur la durée et la facilité de perception pour des électrodes intra-corticale. L'écart entre l'anode et la cathode faisait varier le nombre de percepts pour une même intensité de stimulation ; un résultat qui semble expliqué par la plus grande population de neurones stimulée lorsque l'écart est plus important. Le seuil d'activation des percepts était beaucoup plus élevé en implantation de surface qu'en intra-corticale. Finalement, il ne semble pas y avoir de fusion entre des phosphènes proches pour former un seul pattern, contrairement aux études sur la stimulation de la rétine (Humayun et al., 1999) où des phosphènes proches peuvent s'assembler pour former une ligne par exemple.

Ces études d'électrophysiologie chez l'homme montrent qu'il est possible d'évoquer des percepts lumineux chez des non-voyants avec certaines propriétés reproductibles. L'avenir des neuroprothèses corticales réside dans l'augmentation du nombre d'éléctrodes et la diminution du nombre de neurones stimulés à chaque stimulation. De nombreuses études sont menées en parallèle chez l'animal (Normann et al., 1999) avec des matrices d'éléctrodes toujours plus petites et plus denses (Figure 31).

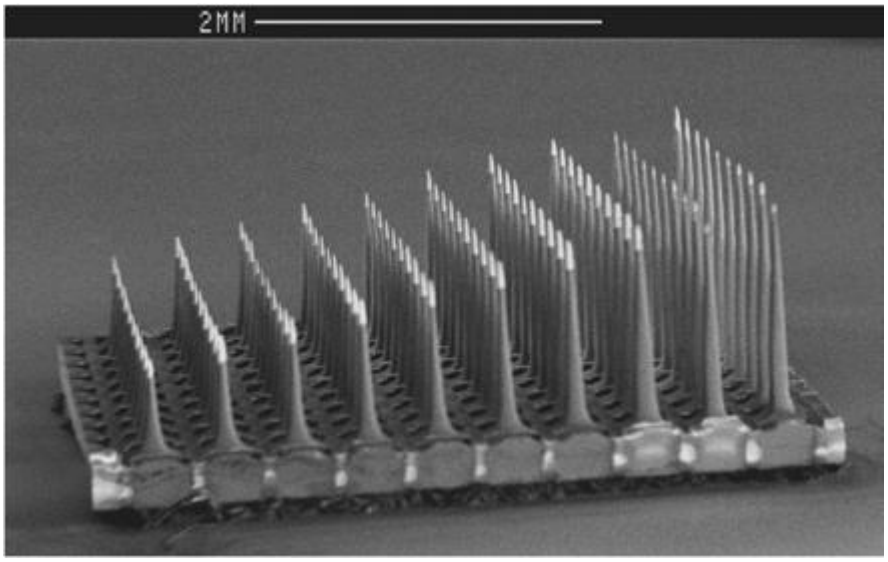


Figure 31: Matrice de stimulation intra-corticale (Utah Slanted Electrode Array)

8) Bases théoriques de la neuroprothèse et systèmes de suppléance

La substitution sensorielle est une manière non-invasive de restituer une information. Les mécanismes de traduction d'une modalité sensorielle vers une autre sont étudiés depuis de nombreuses années. Les recherches en neurosciences ont permis de montrer qu'il était possible de stimuler des zones du cerveau impliquées dans le traitement des informations sensorielles pour évoquer des percepts. Il est alors possible de connecter le cerveau avec un stimulateur pour stimuler directement ces territoires corticaux.

Ces dernières années, les interfaces cerveau-machine ont constitué un domaine de recherche très actif et très prometteur. Les implants cochléaires en sont le premier exemple: en 1957, Charles Eyriès pratique la première opération ayant pour objet d'implanter à demeure des électrodes de stimulation dans la cochlée d'une personne sourde. Lorsqu'un courant est appliqué sur ces électrodes, l'excitation des premiers relais nerveux du système auditif provoque la perception de sensations auditives qui peuvent s'apparenter à des sons. Le principe général de cet implant est simple (voir Figure 32) : capter l'information sonore environnante avec un micro et la transformer afin de la restituer sous forme de stimulation électrique à différents emplacements de la cochlée pour faire percevoir un son intelligible.

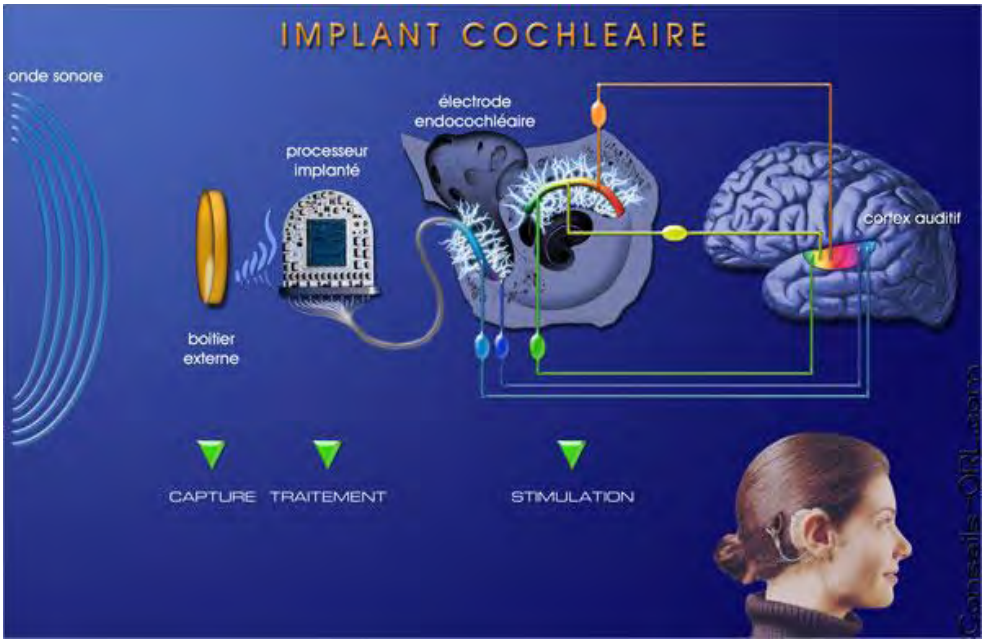


Figure 32: Illustration de la chaîne de fonctionnement d'un implant cochléaire. Le son est capturé, traité par un processeur implanté puis restitué en stimulations de la cochlée pour produire une sensation sonore. (Source : Conseils-Orl.com)

Le système auditif est probablement plus facile à interfacer que le système visuel et les différences dans le développement des neuroprothèses auditives et visuelles le démontrent amplement. Mais depuis quelques dizaines d'années, des chercheurs étudient le fonctionnement du système visuel à tous ses étages : la rétine, le nerf optique, le cortex visuel et ces études commencent à porter leurs fruits.

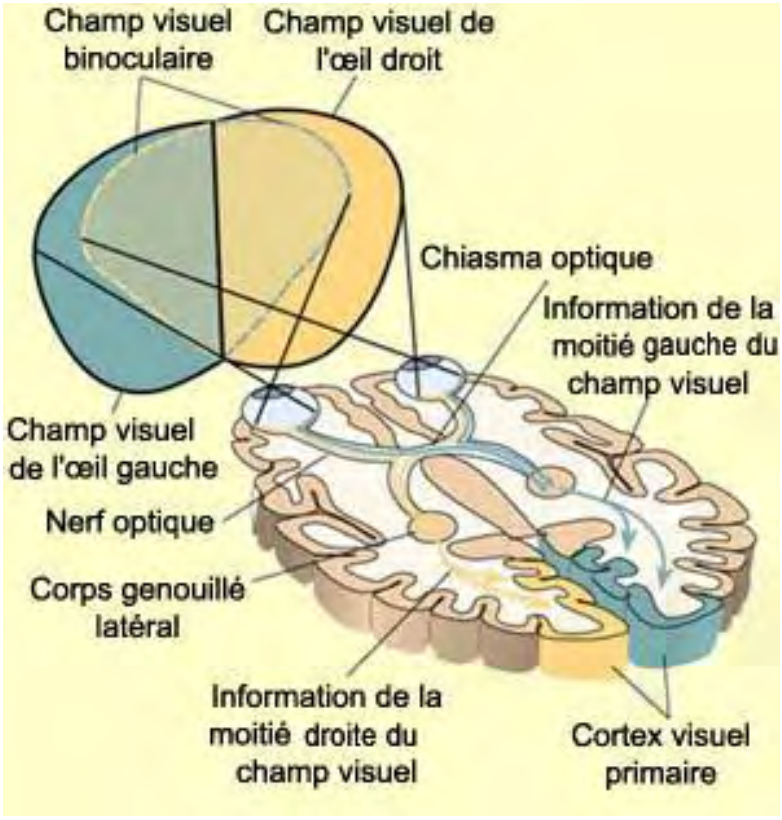


Figure 33: Schéma du système visuel humain. Trois sites de stimulation sont aujourd'hui principalement étudiés : la rétine, le nerf optique et le cortex visuel

Rétine

La majorité des maladies aboutissant à la cécité sont des dégénérescences des cellules de la rétine. La principale cause de ces lésions dans les pays riches reste la dégénérescence maculaire liée à l'âge (DMLA), altérant la vision centrale chez les personnes de plus de 50 ans. A contrario, la rétinite pigmentaire, maladie héréditaire, affecte les cellules photosensibles périphériques de la rétine. Les neuroprothèses rétiniennes sont des matrices d'électrodes implantées directement dans la rétine permettant de remplacer les photorécepteurs lésés. Trois projets se distinguent particulièrement.

Le premier, porté par Alan Chow, propose d'implanter une biopuce dans la rétine (ASR : Artificial Silicon Retina) pour améliorer la perception des contrastes et des couleurs chez des patients atteints de rétinite pigmentaire. La puce est pourvue de 5000 photorécepteurs électroniques reliés à des électrodes, le tout inséré dans la rétine (Chow et al., 2004). La lumière captée est ainsi transformée en signal électrique qui se propage vers la couche voisine de neurones ganglionnaires. Le signal est ensuite transmis au cerveau via le nerf optique comme dans un œil fonctionnel. Aucun rejet n'a été constaté parmi les personnes implantées et des améliorations notoires sont rapportées. Depuis janvier 2000, 10 patients

ont pu être implantés et disent avoir recouvré la perception de certains contrastes et couleurs.

Sur le même principe, le projet allemand (universités de Tübingen et de Regensburg) porté par le professeur Zrenner vise à remplacer les cellules lésées par un système de diodes photoélectriques implanté sous la conjonctive et comportant entre 100 et 1000 photorécepteurs équipés chacun d'une microélectrode. L'implant devrait donner une acuité visuelle d'environ 1/20. Les premiers essais sur des cochons après quatre semaines d'observation n'ont montré aucun rejet (Gekeler et al., 2006). D'autres tests ont de plus montré que les stimulations de ces implants rétiniens chez des chats aveugles (maladies rétiniennes) provoquent une activité du cortex visuel (Eckhorn et al., 2006).

Ces deux projets basés sur le remplacement des cellules lésées par des cellules photosensibles permettent d'augmenter la perception des contrastes en s'appuyant sur les cellules rétiniennes toujours présentes. Une deuxième approche pour restaurer les fonctions visuelles serait de remplacer l'intégralité de la rétine par une rétine entièrement artificielle. Bien que la rétine ait un fonctionnement beaucoup plus complexe qu'une simple matrice de capteurs photo-électriques, les approches actuelles tentent de la comparer à une caméra haute résolution. Il 'suffirait' alors d'allumer les points désirés pour percevoir une image visuelle. C'est l'idée qu'ont eu, à Boston, Joseph Rizzo (Harvard) et John Wyatt (MIT). Ils ont mis au point un dispositif comprenant une caméra (Rizzo et al., 2003) et un processeur montés sur des lunettes, qui envoient les signaux visuels à une rétine artificielle composée d'électrodes de 25 µm de diamètre. Les essais cliniques ont débuté en 2003 et révèlent de nombreuses difficultés. Si l'on pouvait penser que la rétine est l'analogue d'une matrice de pixels, les tests ont montré l'inverse. Cinq patients atteints de rétinite pigmentaire sévère ont reçu des stimulations rétiniennes et ont fait émerger des résultats inattendus. En effet, les patients rapportaient des percepts différents pour une même stimulation délivrée à deux moments différents. Ces résultats variables pourraient s'expliquer par une stimulation trop grossière liée aux difficultés rencontrées pour transmettre un courant d'une intensité suffisante (Rizzo et al., 2003). Ces résultats montrent la complexité de fonctionnement de la rétine et rendent pertinentes les recherches qui reposent sur d'autres sites de stimulations.

Nerf optique

La stimulation du nerf optique est une approche qui peut être utilisée pour les patients ayant des pathologies rétiniennes mais dont le nerf optique et les relais suivants sont intacts.

Elle s'adresse donc aux mêmes patients que les neuroprothèses rétiniennes, ainsi qu'aux patients ayant une rétine totalement non-fonctionnelle.

L'équipe de Claude Veraart de l'université de Louvain en Belgique étudie la stimulation du nerf optique par quatre électrodes implantées autour du nerf. Une micro-caméra fixée sur une paire de lunettes envoie les images dans un petit ordinateur accroché à la ceinture. Les signaux sont renvoyés aux lunettes qui les diffusent vers une antenne située sous la peau du crâne, proche du nerf optique. Ces signaux alimentent une puce de stimulation qui envoie du courant à des électrodes enroulées autour du nerf optique. La première implantation a été réalisée en 2001 chez une patiente de 59 ans atteinte de rétinite pigmentaire. Pour l'instant, cette patiente a appris à distinguer des formes géométriques simples, à les localiser et à aller saisir des objets placés devant elle. L'évaluation de ses performances pour reconnaître des formes et réaliser des mouvements de saisie vers des cibles ont commencé (Duret et al., 2006). Les tests portent sur la capacité à discerner un objet parmi six pouvant se trouver à neuf emplacements différents. La caméra alimentait un stimulateur du nerf optique permettant d'évoquer 109 phosphènes différents dans un champ visuel évalué de 14°x41°. La patiente était assise dans une petite salle aux murs noirs, en face d'une table noire quadrillée en 9 zones (29 cm * 21 cm chacune) numérotées (Figure 34). Les six objets utilisés (grande bouteille, petite bouteille, boite de CD, boite de dentifrice, tasse et couteau) étaient blancs pour limiter le travail d'analyse d'images. L'expérimentation consistait à localiser l'objet, le reconnaitre puis le saisir. Chacune des 5 sessions comportait ces trois tâches et durait environ 40 minutes. Les résultats montrent que cette personne était capable de localiser l'objet après 20 à 30 secondes de balayage. Elle pouvait le reconnaitre en 40 à 50 secondes avec un taux d'erreur en diminution au cours des sessions pour atteindre un score parfait à la dernière session. L'atteinte de la cible une fois localisée n'a pas posé de difficultés : si la patiente avait correctement localisé la cible visuelle, elle arrivait toujours à atteindre la cible. Différentes stratégies pour localiser et reconnaître les objets ont été étudiées suivant la nature de l'objet. Un balayage horizontal puis vertical était nécessaire pour localiser la cible visuelle puis un autre balayage autour de la cible était nécessaire pour la discriminer. Les performances de discrimination se sont beaucoup améliorées au cours des sessions.

Figure 34 : photo du dispositif expérimental pour la reconnaissance, la localisation et l'atteinte d'objets visuels : Le sujet est assis devant une table noire découpée en 9 zones rectangulaires (le chiffre correspond au numéro de ligne ; la lettre R pour Right, C pour centre, L pour Left.

Ces résultats sont très encourageants car peu d'études ont réellement évalué l'utilisation de neuroprothèses dans des tâches de saisie d'objets. Les stratégies de reconnaissance et de localisation montrent que la réalisation de ces tâches n'est pas immédiate et nécessite une exploration prolongée de la zone capturée par la caméra. Avec la rétine, les stimulations du nerf optiques sont celles qui ont montré le plus de reproductibilité et une certaine cohérence dans la distribution spatiale des percepts. Ce projet montre qu'il est aujourd'hui possible de contrôler les percepts évoqués par une stimulation du nerf optique en fonction de l'environnement visuel devant la caméra et d'agir dans un environnement visuel à partir de stimulations du nerf optique.

Cortex visuel

Le cortex visuel est la zone du cerveau où la majeure partie de l'information visuelle est traitée. La stimulation de certaines zones du cortex permet d'évoquer des perceptions visuelles à des patients dont les premiers relais nerveux de l'information visuelle sont manquants (pathologies de la rétine, du nerf optique, ou des relais sous-corticaux). Une neuroprothèse corticale permettrait donc d'élargir le champ d'application des

neuroprothèses visuelles. Un seul projet portant sur l'humain – celui de l'Institut Dobelle à New York - a pour l'instant été conduit mais il est aujourd'hui arrêté. Ce projet repose sur un dispositif comprenant une micro-caméra connectée à un ordinateur, lui-même relié au cerveau du patient par le biais de 68 électrodes disposées à la surface du cortex visuel (Dobelle, 2000a) (Figure 35).

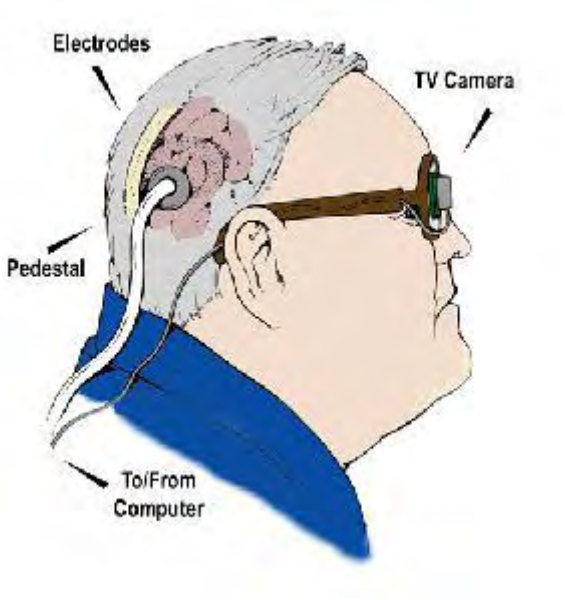


Figure 35 : Schéma illustrant le dispositif du Dobelle Institute. Les informations lumineuses en provenance d'une caméra sont traitées par un ordinateur (filtrage, détection de bords, etc.) et renvoyées sous forme de stimulation électrique vers le cortex visuel.

Un patient de 62 ans, aveugle depuis l'âge de 36 ans, a été implanté (Jerry). Avec ce dispositif, cette personne est parvenue à percevoir des lettres de six centimètres de haut à une distance de six centimètres. Un télémètre à ultrasons miniaturisé, placé au-dessus de la tempe gauche, permettait d'évaluer les distances. Ce dispositif simplifie les images vidéo et les signaux indiquant la distance, qui transmet un courant aux électrodes intracorticales. Les signaux sont transmis au cerveau par le biais de 68 électrodes insérées dans au niveau de l'aire visuelle primaire. Le passage du courant dans ces électrodes provoque une excitation des neurones visuels et entraîne la perception de phosphènes. Le patient apprend à "lire" ces phosphènes, lui procurant une acuité visuelle faible, dans un champ de vision limité, fonction de la position et du nombre d'électrodes de stimulation. Avec de nouvelles améliorations, dont un plus grand nombre d'électrodes (qui permettraient d'améliorer la résolution) et un traitement de l'image plus poussé, le système pourrait fournir un traitement pour la plupart des cécités de l'enfant et de l'adulte. Les travaux menés par l'Institut Dobelle ont démarré depuis plus de 30 ans. En 1978, cette même personne

implantée (Jerry) avait déjà reçu un implant. Il pouvait avec ce dernier distinguer des lettres et chiffres, suivre des lignes. Cependant, il n'existe pas de réel test générique pour comparer le véritable gain de ce système par rapport à d'autres. En plus de ces résultats, ces études montrent que l'on peut implanter une personne sans rejet de la prothèse et la lui changer quelques années après sans difficulté majeur selon les concepteurs de la prothèse.

Ces systèmes ouvrent de grandes perspectives en montrant qu'il est possible de redonner une information visuelle mais nécessitent une chirurgie lourde pour finalement une faible amélioration de l'acuité visuelle. L'utilisation d'une telle chirurgie pour améliorer une vision inexistante n'aura vraisemblablement pas les mêmes cibles et les mêmes usages que les outils beaucoup plus légers, non-invasifs. Les prothèses de restauration de la vision sont en revanche amenées à se développer avec les progrès des neurosciences de la vision, de la neurochirurgie, du traitement d'image et de la maîtrise des interfaces de stimulation. Les phosphènes générés lors de la stimulation du cortex visuel ont été décrits dans les études comme la perception d'étoiles dans le ciel (Brindley and Lewin, 1968). Ils pouvaient être de formes arrondies, allongées ou encore de tailles différentes selon la localisation de la stimulation dans le cortex visuel (Dobelle et al., 1974). Les phosphènes étaient décrits comme rétinotopiques et se déplaçant avec le regard (Brindley, 1982;Dobelle et al., 1974;Dobelle, 2000b;Schmidt et al., 1996). De plus, lorsque la durée ou la fréquence de stimulation étaient modifiés, les sujets disaient percevoir respectivement une variation de la luminance des phosphènes ou un clignotement.

9) Bilan sur les neuroprothèses

Les recherches pour concevoir des neuroprothèses visuelles visent à restaurer la vision : trouver le moyen avec un minimum de phosphènes, de faire voir ou revoir des personnes aveugles. Cette approche s'appuie sur un traitement de l'information visuelle de type « scoreboard » en associant chaque pixel ou groupe de pixels d'une image issue de caméras vidéo à une électrode placée au niveau du cortex visuel. L'information que l'on veut restituer est ainsi une image presque brute sans prétraitement majeur. Les résultats obtenus dans ces études ont montré que pour atteindre cet objectif, il faudrait parvenir à augmenter significativement le nombre d'électrodes à implanter. La faible résolution de cette interface de restitution pose des problèmes similaires à ceux des systèmes de substitution sensorielle. En effet, il est très difficile d'interpréter un signal de type 'score-board' pour représenter une image capturée par une caméra car la résolution de restitution y est trop faible. Comme les

systèmes électroniques de suppléance visuelle, l'enjeu est de sélectionner les informations pertinentes à restituer pour augmenter l'autonomie des non-voyants.

Discussion sur les systèmes électroniques de suppléance visuelle

La France compte aujourd'hui 1,7 millions de personnes déficientes visuelles dont 13% sont aveugles. Les données sur la sociabilisation des personnes atteintes de cécité visuelle montrent que de nombreux progrès sont encore à faire pour une meilleure prise en charge de ce handicap. Pour cela, de nombreux efforts ont été faits pour signaler le mobilier urbain ou rendre accessibles différentes informations pouvant aider les non-voyants à se repérer. En complément de ces efforts pour rendre accessible les informations sans modification de l'environnement, un domaine de recherche connaît aujourd'hui une très forte activité: la conception de systèmes de suppléance pour appréhender l'espace, lorsque celui-ci n'est pas particulièrement adapté au handicap et accessible. La canne blanche est aujourd'hui l'outil le plus utilisé et elle nécessite une phase d'apprentissage courte. Son utilisation permet de signaler son handicap, détecter des obstacles, reconnaître des revêtements de sol ou des objets particuliers pour s'orienter. Son principal inconvénient est qu'elle ne permet pas de s'orienter. Reconnue d'utilité publique en 1981, la Fédération Française des Associations de Chiens Guides d'aveugles a permis de démocratiser la compagnie de chiens guides pour aveugles en créant 10 écoles de formation, un centre d'études, de sélection et d'élevage. Le chien présente le principal avantage de s'adapter et pouvoir guider les personnes en situation de handicap pour trouver leur chemin là où le champ opérationnel de la canne est trop faible. Malgré toute l'aide que celui-ci peut apporter aux personnes déficientes visuelles, la compagnie d'un animal ne convient pas à tout le monde et ceux-ci ne sont pas toujours acceptés dans les lieux publics (restaurants, théâtres, hôpitaux...).

Il existe aujourd'hui un large besoin auquel les systèmes de suppléance peuvent répondre. Si l'accessibilité des outils domestiques, la lecture et l'écriture représentent un domaine de recherche très actif ayant abouti à de nombreuses solutions, nous ne traiterons dans cette thèse que le domaine de la navigation et la reconnaissance d'objets pour les non-voyants. La navigation pour les non-voyants peut se décomposer en deux catégories bien distinctes : les aides au déplacement (détection d'obstacles, restitution de motifs visuels …) et les aides à l'orientation (se localiser et s'orienter dans son environnement). Le premier a été largement étudié depuis les années 1970 par Paul Bach-Y-Rita en proposant des systèmes de substitution vision-tactile qui convertissent les motifs visuels capturés par une caméra en

stimulations tactiles. Sur le même principe, la stimulation de différentes parties du corps a été étudiée, dont la langue (Kupers and Ptito, 2004), l'abdomen (Bach-y-Rita, 1983) ou le palais (Hui and Beebe, 2003;Hui and Beebe, 2006). Depuis les années 2000 sont apparus des systèmes de substitution vision-audition (Arno et al., 1999;Durette et al., 2008;Gonzalez-Mora et al., 2006;Meijer, 1992) qui convertissent les motifs visuels capturés par une caméra vers des sons interprétables. Aujourd'hui, ces systèmes permettent de localiser et reconnaître des objets dans des situations de laboratoire avec des formes extrêmement simples sur un fond uniforme après un apprentissage long et difficile. Un système permettait la localisation rapide d'objets (Gonzalez-Mora et al., 2006) en proposant de synthétiser des sons comme si ceux-ci provenaient de la surface des objets. Ce système présente le principal avantage d'intégrer une notion tridimensionnelle à l'image perçue et permet même si un objet n'est pas reconnaissable, de le localiser.

L'état de l'art sur les systèmes d'aides au déplacement fait apparaître un réel écart entre le besoin des non-voyants et les attentes que l'on pourrait avoir de ces systèmes. La canne laser ou à ultrasons est le seul appareil d'aide au déplacement aujourd'hui réellement utilisé car il répond à un besoin précis en terme de détection d'obstacles et qu'il permet aussi de reconnaître des volumes qui aident à appréhender l'espace. Le coût d'un tel système est de 2300 euros (2010) et le manque d'aide financière pour accéder à ce genre de système est aujourd'hui un frein à sa démocratisation. Les systèmes d'aide à l'orientation pour les non-voyants sont basés principalement sur des systèmes de géolocalisation en intérieur (intensité de signal wifi, RFID …) ou en extérieur (GPS). Le projet de recherche pour l'aide à l'orientation le plus connu a été développé par Loomis à Santa Barbara (Loomis et al., 1998). Différentes méthodes de restitution de l'information adaptées au non-voyant ont été testées pour aider les utilisateurs à se créer une représentation interne de l'espace. De nombreux systèmes de GPS, pas toujours développés spécifiquement pour les non-voyants sont aujourd'hui largement utilisés pour s'orienter. Le Kapten en est une bonne illustration : la société qui développe cet outil a au départ développé un GPS sans écran pour voyants, mais ce système, particulièrement adapté à une interaction non-visuelle, a provoqué un réel engouement dans la communauté non-voyante du fait de son coût très faible (environ 150 euros) par rapport aux autres solutions adaptées à la cécité (1000-2500 euros). Le principal problème de ces systèmes réside dans l'absence de système d'information géographique précis et adapté au piéton aveugle et dans le manque de précision du GPS, surtout en environnement urbain (5 à 10 m). Il est évident qu'un système d'information qui

comporterait l'ensemble du mobilier urbain nécessaire à la navigation non-visuelle ainsi qu'un capteur de géolocalisation très précis serait un outil particulièrement efficace pour éviter les obstacles, trouver la porte d'un bâtiment, un distributeur de billets ou pour un guidage précis d'un point à un autre d'un trajet. Les outils de suppléance interactifs représentent aujourd'hui un enjeu très important et n'ont pas encore atteint leurs limites du fait des progrès encore à faire dans le traitement du signal d'entrée de ces différents dispositifs.

Finalement, les verrous scientifiques à la conception de systèmes de suppléance invasifs et non-invasifs ont de grandes similitudes, principalement à cause de la faible résolution de la modalité de sortie. Les systèmes de suppléance invasifs permettent aujourd'hui d'entrevoir des méthodes pour restituer une perte de vision. Trois principales méthodes sont étudiées : la stimulation de la rétine, la stimulation du nerf optique ou la stimulation du cortex visuel. La littérature contient quelques résultats épars nous renseignant sur les percepts évoqués par la stimulation du système nerveux humain mais il manque aujourd'hui des données fiables et reproductibles pour prédire l'efficacité d'une implantation suivant les patients. Il existe en revanche de nombreuses propriétés reproductibles sur le fonctionnement de ces stimulations à trois étages du système visuel. La stimulation de la rétine semble être actuellement la méthode la mieux maîtrisée. L'implantation d'une rétine artificielle dotée de photorécepteurs permet de stimuler directement la rétine des personnes non-voyantes. Le principal inconvénient fonctionnel de cette méthode réside dans la proportion importante de maladies aboutissant à la cécité dans lesquelles la plupart des cellules de la rétine sont détruites. La stimulation du nerf optique ou du cortex visuel pourrait représenter des alternatives à la stimulation de la rétine en cas de lésions à des étages plus élevés du système visuel. Les résultats récents issus des neurosciences montrent qu'il est possible d'établir des règles stables de prédiction des percepts et permettent aujourd'hui d'entrevoir une certaine reproductibilité dans les l'évocation des percepts. Ces systèmes ne sont aujourd'hui pas matures pour être commercialisés et souffrent d'un manque de connaissance du fonctionnement du système visuel humain mais aussi d'une trop faible résolution. L'approche actuelle consiste simplement à réduire la résolution des images capturées afin de restituer un motif correspondant à ce qui est vu. Cette approche souffre exactement des mêmes limites que les systèmes de substitution sensorielle : la matrice d'électrodes a une résolution beaucoup trop faible pour pouvoir être exploitable pour restituer des informations visuelles point à point.

Les systèmes de substitution sensorielle ont largement étudié les différents modes de restitution d'une information pour les non-voyants par des sensations à la surface de la peau ou de manière auditive. Les systèmes invasifs eux, montrent qu'il est possible d'évoquer directement des percepts visuels mais que leur nombre est très limité du fait du nombre électrodes implantables aujourd'hui. Ces systèmes montrent que les systèmes sensoriels humains ou que les méthodes de stimulation du système nerveux ne permettent pas d'interpréter des motifs complexes avec précision et rapidement. Un enjeu majeur de ces prochaines années réside dans le traitement des informations acquises par un ou plusieurs capteurs pour répondre précisément aux besoins des non-voyants. L'objectif n'est alors plus de restaurer la vision par un autre canal sensoriel mais augmenter un canal sensoriel avec des informations identifiées comme utiles, permettant d'aider à appréhender l'environnement extérieur. Ces informations peuvent par exemple concerner l'organisation spatiale des objets ou du mobilier nous entourant puisqu'elles sont très importantes à la navigation : Le capteur de vision artificielle (caméra vidéo) capture toutes ces informations. L'enjeu des systèmes de suppléance est d'arriver à traiter ces informations visuelles pour ne restituer que celles qui sont nécessaires à un instant donné. Pour cela, l'idée poursuivie dans cette thèse est qu'il est possible avec les algorithmes de vision actuel de traiter l'information visuelle en temps réel et aider les non-voyants à se déplacer et à localiser des objets en restaurant une des fonctions du système visuel humain : la reconnaissance et localisation d'objets. Cette information identifiée comme utile pour augmenter l'autonomie des non-voyants est restituable aussi bien pour les systèmes électroniques de suppléance visuelle que pour les neuroprothèses.

Synthèse et objectifs de la thèse

Je présenterai dans ce manuscrit une étude du besoin effectuée auprès de 54 utilisateurs non-voyants permettant d'établir les principales difficultés des non-voyants et comment y répondre. Convaincu de l'importance d'inclure l'utilisateur final dans l'élaboration d'un système de suppléance, celui-ci tiendra une place centrale dans tout le processus de conception. Les systèmes de suppléance électroniques sont composés d'un module d'acquisition, d'un module de transformation et d'un module de restitution. Ce dernier est directement dépendant de l'information que l'on souhaite restituer et donc des capteurs utilisés. A partir de cette étude du besoin, je montrerai l'importance de concevoir un système basé sur la reconnaissance et la localisation d'objets pour répondre aux principales difficultés des non-voyants dans trois principales catégories du besoin : le déplacement,

l'orientation et la reconnaissance d'objets. Imaginez que vous soyez en train de rechercher un objet, que vous puissiez le demander oralement au système et que celui-ci émette un son dans un casque audio en donnant l'impression qu'il provient de l'objet recherché lui-même.

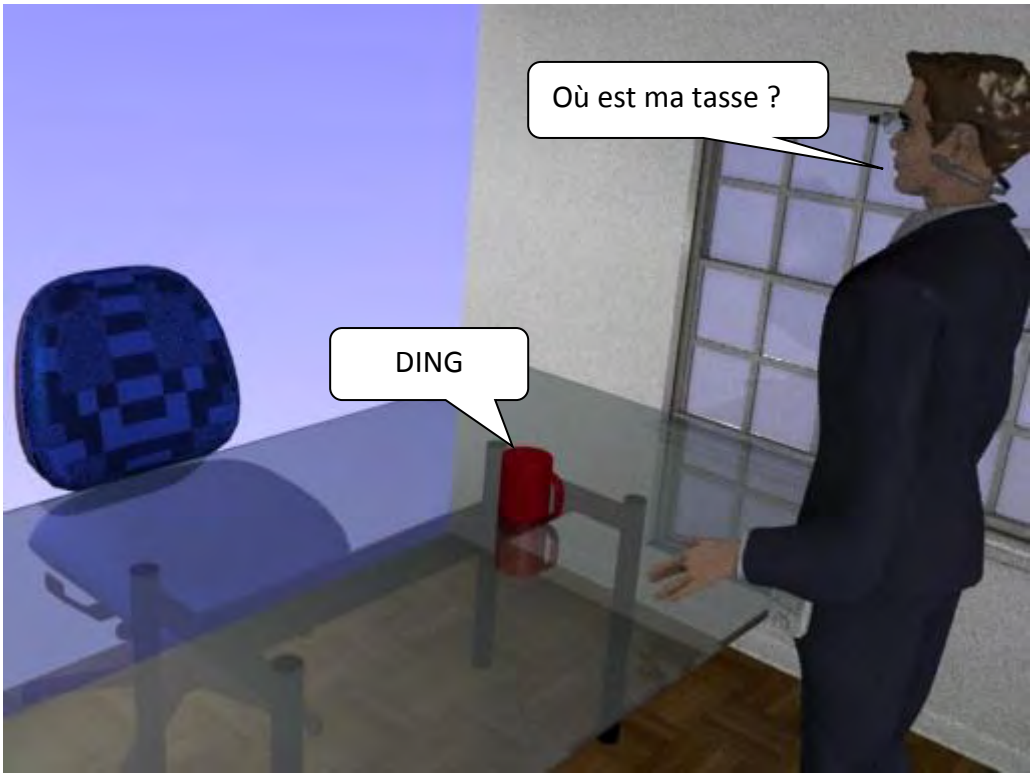


Figure 36: Illustration du fonctionnement du prototype de reconnaissance et de localisation d'objets. L'utilisateur demande au système où est sa tasse, le système localise la tasse par reconnaissance de formes et reconstruction tridimensionnelle. Un son synthétisé comme s'il provenait de la tasse est alors émis dans un casque audio.

La plupart des systèmes basés sur la vision artificielle traduisent les données brutes d'une modalité vers une autre par substitution sensorielle, par des règles simples de conversion d'images en stimuli tactiles ou auditifs. L'idée poursuivie dans cette thèse est qu'il est possible en agissant sur le module de transformation du signal, de filtrer l'information à restituer en la ciblant sur un besoin : la reconnaissance et la localisation d'objets. J'ai donc travaillé sur l'élaboration d'un capteur de vision artificielle pour reconnaître et localiser des objets dans l'espace tridimensionnel. Les algorithmes de vision artificielle qui sont développés aujourd'hui ne sont pour la plupart pas temps réel. Il n'existe pas d'algorithme permettant de reconnaître et localiser de manière parfaitement fiable et en temps réel des objets 3D dans une scène visuelle, mais il est en revanche possible d'utiliser les algorithmes existants pour répondre à certains problèmes des non-voyants. L'approche utilisée dans cette thèse place la nécessité d'un système temps réel au cœur de la conception pour

répondre au besoin. La boucle Perception-Action tient un rôle primordial dans l'utilisabilité d'un tel système. Il faut que l'utilisateur puisse percevoir en temps réel son espace et surtout que le capteur soit lié à ses actions.

Les recherches en neurosciences sur le système visuel humain ont étudié la faculté de l'homme à catégoriser et localiser extrêmement rapidement des objets dans une scène visuelle et ont permis de mettre au point un modèle informatique pour la reconnaissance et la localisation d'objets bio-inspiré : Spikenet. Cet algorithme de reconnaissance et de localisation d'objets temps réel constitue la première brique du système de vision artificielle utilisé. Une étude empirique du fonctionnement de ce système pour établir son périmètre d'action pour un système de suppléance sera étudiée. L'objectif de cette étude est de déterminer les champs d'applications de cet algorithme pour une utilisation fiable en temps réel. La reconstruction de la position 3D d'un objet reconnu dans une image sera effectuée à l'aide d'un capteur stéréoscopique composé de deux petites caméras portées sur la tête du sujet. La méthode de reconstruction 3D utilisée sera développée dans la description d'un premier prototype du capteur de stéréovision fabriqué à partir de deux webcams et une calibrage manuelle des caméras. Un deuxième prototype du module de vision a également été développé à partir d'un dispositif commercial permettant de reconstruire plus précisément l'environnement visuel 3D. Couplé avec un outil de reconnaissance d'objets dans une image en temps réel, ce module nous permettra la position tridimensionnelle et l'identité des objets visuels. Cette fonction particulièrement utile à la navigation et à la recherche d'objets dans l'environnement visuel servira de base à l'élaboration d'interfaces de restitution novatrices adaptées au besoin.

Les systèmes de suppléance pour les non-voyants sont basés sur une restitution tactile (distance à un objet, braille) ou auditive (Synthèse vocale, substitution sensorielle). Un des besoins primordial défini par l'étude du besoin des personnes non-voyantes est la nécessité de pouvoir localiser des objets d'intérêt en temps réel. Aucun de ces modes de restitution ne permet de restituer de manière instantanée la position d'une cible dans l'espace. Une des propriétés du système auditif humain est la faculté de pouvoir localiser avec précision des sources sonores dans l'espace. La localisation binaurale dans les deux dimensions angulaires (azimut et élévation) a été largement étudiée et décrite, mais il subsiste en revanche une méconnaissance des performances de localisation dans l'espace proche de l'utilisateur en prenant en compte la perception de la distance. Nous avons pour cela évalué

et comparé la précision de localisation de stimuli auditif réels dans l'espace péripersonnel (<1m) selon deux conditions acoustiques de salles et différents stimuli au contenu différent. Les résultats nous permettront de définir un stimulus le plus adapté possible à la restitution d'une position tridimensionnelle dans l'espace. Ces résultats issus d'une évaluation avec les utilisateurs finaux seront intégrés à une interface de restitution sonore basée sur la synthèse binaurale.

Un autre mode de restitution à plus long terme a été étudié et des résultats préliminaires encourageants seront présentés. Les neuroprothèses ont pour objectif à long terme de restaurer la vision des non-voyants en stimulant électriquement le système visuel. Sur la base d'un recueil de propriétés issues de la littérature sur l'évocation de percepts visuels issus de la stimulation du système visuel humain, un modèle de neuroprothèse visuel a été créé. Pour cela, des percepts sont évoqués dans un casque de réalité virtuelle pour simuler l'utilisation d'une neuroprothèse pour étudier les performances comportementales dans diverses situations de reconnaissance et saisie d'objet. Une expérimentation préliminaire a été conduite auprès de 10 sujets voyants pour évaluer la possibilité de se diriger vers une cible. Ces sujets devaient évoluer dans un monde virtuel et atteindre des cibles représentées sous forme de phosphènes similaire à ceux qui seraient générés par une neuroprothèse. Cette étude exploratoire est une première étape dans la conception et l'évaluation d'une véritable neuroprothèse visuelle. Nos résultats sur la localisation de cibles sont très encourageants.

Les travaux présentés ici constituent la base d'un projet financé par l'Agence Nationale pour la Recherche et la Région Midi-Pyrénées pour l'aide à la navigation pour les personnes non-voyantes. NAVIG est basé sur le système de vision présenté dans cette thèse en couplant la reconnaissance d'objets avec un capteur GPS pour le guidage du non-voyant. Les travaux sur la vision artificielle et la restitution auditive pour répondre au besoin des non-voyants seront présentés dans le cadre plus large de ce projet.

CHAPITRE II : CONCEPTION D'UN SYSTEME DE RECONNAISSANCE ET DE LOCALISATION D'OBJETS POUR LES NON-VOYANTS

L'état de l'art sur les systèmes de suppléance montre qu'il existe une vraie demande des personnes non-voyantes à laquelle les systèmes actuels ne répondent pas. Les outils d'aide à l'orientation sont peu adaptés au piéton du fait de deux problèmes majeurs : le manque de système d'information géographique adapté et la précision insuffisante des outils de géolocalisation. Les outils d'aide au déplacement par substitution sensorielle ne sont pour la plupart pas utilisés du fait de la faible résolution des modalités de restitution tactile ou auditive.

La plupart des travaux de cette thèse s'inscrivent dans le projet NAVIG. Il a pour objectif d'augmenter l'environnement avec des informations auditives utiles à la navigation et à la localisation d'objets. Chaque module est développé et testé séparément en plaçant l'utilisateur au cœur de chaque cycle de conception. Ce chapitre décrit la conception d'un système de suppléance pour les non-voyants basé sur la reconnaissance et la localisation d'objets par vision artificielle et une restitution de la position spatiale de ces objets par synthèse binaurale. Plus précisément, je présenterai une étude du besoin effectuée auprès de 54 non-voyants dont les résultats montrent clairement la pertinence d'un système de reconnaissance et de localisation d'objets pour les non-voyants. Pour cela, un système de vision artificielle pour la reconnaissance et la localisation d'objets a été développé. Les performances du système de reconnaissance de formes seront évaluées dans une application à la reconnaissance de billets de banque testée auprès de sujets non-voyants mais aussi dans une évaluation exhaustive du système dans des images naturelles. Nous étudierons dans cette thèse deux modes de restitution différents pour restituer la position d'un objet dans l'espace : par synthèse binaurale et dans une neuroprothèse. Je présenterai pour cela une étude psychophysique sur les capacités des voyants et des non-voyants à localiser une source sonore dans l'espace proche de l'utilisateur (<1m). Ces résultats seront utilisés pour concevoir un module de restitution de l'information de position de l'objet par synthèse binaurale. Cette même information de position d'un objet sera aussi évaluée dans un casque de réalité virtuelle simulant au plus près des percepts pouvant être évoqués dans une neuroprothèse. Finalement, je présenterai un système de reconnaissance et de

localisation d'objets pour les non-voyants basé sur la reconnaissance d'objets et une restitution de la position des objets par synthèse binaurale.

Méthodes de conception

Un des défauts majeurs dont souffrent aujourd'hui les méthodes de conception des systèmes de suppléance est le manque d'implication des utilisateurs finaux. La conception doit-être pensée dès les premières phases de modélisation de l'application finale en intégrant les étapes d'évaluation et de retour à la conception. Nous avons intégré les utilisateurs au cycle de développement du dispositif par une analyse du besoin, point de départ fondamental à la conception du système. Les résultats des études psychophysiques présentées dans ce manuscrit font partie intégrante de la méthode de conception centrée utilisateur.

Par ailleurs, afin de faciliter la conception et l'évaluation de chaque module du dispositif, nous avons utilisé une méthode de conception modulaire basée sur la conception d'agents autonomes communiquant par échange de messages sur un bus logiciel réseau.

10) Conception modulaire et prototypage rapide

Concevoir des systèmes interactifs expérimentaux est une tâche rendue complexe par l'hétérogénéité des environnements logiciels et matériels utilisés pour leur développement. Bien souvent, la technique est un frein à la conception participative avec des temps de retours pour l'intégration de nouveaux résultats et de nouvelles idées trop élevés : les environnements de développement sont pour la plupart très cloisonnés et limités à des plateformes logicielles ou matérielles spécifiques. Comment faire communiquer plus efficacement tous ces systèmes ? Deux tendances se dégagent : choisir un seul environnement de développement cohérent mais au prix de potentialités restreintes ou bien opérer sur différentes plateformes mais avec les problèmes de communication et d'intégration de ces développements au sein d'un système interactif performant. Il est essentiel que notre environnement de développement favorise l'obtention d'un retour d'expérience sur les fonctionnalités des prototypes, et permette le suivi d'un cycle de développement incrémental. Une des réponses possibles à ce problème consiste à utiliser un bus logiciel (middleware) permettant une communication sur un mode événementiel selon une architecture totalement distribuée. Pour ces raisons, nous avons choisi d'utiliser le bus Ivy développé sous licence LGPL par la DTI/SDER (DGAC) [http://www.tls.cena.fr]. Ivy fonctionne avec l'échange de messages textuels d'un agent à un autre. Aucune structure de

donnée complexe et typée ne peut être envoyée. Cette restriction aux seules chaînes de caractères permet de rendre compatible les données sous la forme d'un standard disponible sur toutes les plateformes et dans tous les langages. Les modules envoient et reçoivent les messages sur une adresse de broadcast. Ainsi chaque module peut s'abonner et écouter des messages filtrés par un préfixe de message (textuel) et invoquer une fonction évènementielle à chaque réception de messages. Il est ainsi possible de supprimer un module, le simuler ou le remplacer très facilement du moment que les modules sont sur le même sous-réseau.

Les trois modules du système (acquisition, traitement et restitution) sont des agents totalement autonomes et pouvant être testés ou simulés séparément. Ce fonctionnement facilite l'évaluation indépendante de chaque module et permet d'évaluer sur une même base logicielle, différentes méthodes d'acquisition, de traitement et de restitution sans changer l'architecture du système.

11) Conception participative, évaluation des modules

La conception participative est une méthode de travail utilisée principalement en conception de logiciels interactifs. Sa principale caractéristique est la participation active des utilisateurs au travail de conception. Il s'agit donc d'une méthode de conception centrée sur l'utilisateur où l'accent est mis sur le rôle actif des utilisateurs. Dans le cadre de ma thèse, j'ai intégré les utilisateurs finaux dans le développement de chaque module du système de suppléance pour les non-voyants. Je commencerai par une étude du besoin qui m'a permis d'isoler des besoins spécifiques des personnes non-voyantes : la catégorisation et la localisation d'objets. Ce résultat m'a permis d'établir les spécifications des trois modules du système depuis le capteur jusqu'à la restitution adaptée à ce besoin.

Je proposerai un système d'analyse de la scène visuelle adapté au besoin utilisateur : la vision artificielle couplée à des algorithmes de reconnaissance d'objets. Le module de restitution des informations visuelles doit répondre à un besoin de localisation 3D et a été conçu en plaçant l'utilisateur au cœur de la conception. Des tests de psychologie expérimentale ont été effectués pour concevoir les caractéristiques d'une interface sonore adaptée à ce besoin pour les personnes non-voyantes et dans un deuxième temps pour concevoir une interface de restitution par la modélisation d'une neuroprothèse adaptée pour présenter cette information de position des objets. Les deux modes de restitution (synthèse binaurale et évocation de phosphènes par réalité augmentée) ainsi que le capteur

et le module de traitement associé seront évalués indépendamment. L'évaluation de chaque module sera réinjectée tout au long du cycle de conception. Je commencerai par investiguer les besoins des utilisateurs non-voyants par une étude auprès de sujets non-voyants et tenterai d'y répondre par un dispositif capable de reconnaître, localiser des cibles visuelles et les restituer.

Étude du besoin utilisateur

Dès le début de ma thèse, j'ai réalisé une enquête auprès de 54 non-voyants au moyen d'un questionnaire électronique accessible sur l'Internet (cf. Annexe). La totalité des sujets ayant répondu ont indiqué être disponibles pour toute autre évaluation future. Ce lien va ainsi nous permettre à chaque étape de conception du système, d'avoir des retours des utilisateurs. Ce questionnaire nous a permis de mettre en valeur les besoins réels, auxquels ne répondent pas les systèmes actuels. 53 d'entre eux avaient entre 20 et 63 ans. Un seul d'entre eux utilisait un outil électronique d'aide à la mobilité (le Télétact).

Une seule personne utilise un outil d'aide électronique à la navigation mais mentionne un coût trop élevé par rapport à son utilité. Les cannes blanches sont utilisées par 76 % des sujets interrogés car elles permettent de détecter les obstacles proches sur le sol. Elles sont en revanche inopérantes dans le cas d'obstacles dont la hauteur est supérieure à celle des genoux. Leur principal atout est leur coût mais aussi leur rapidité d'apprentissage. 28% des sujets interrogés ont abandonné la canne blanche le jour où ils ont eu un chien guide du fait de leur faculté à s'adapter, apprendre de nouveaux trajets, reconnaître de nouveaux points d'intérêts et tout cela, avec une très faible charge cognitive de la part de l'utilisateur. Il n'est pas étonnant que la plupart des non-voyants interrogés considèrent une personne accompagnante comme l'aide la plus pertinente. Quelle que soit le type d'assistance utilisé, 80% des participants sortent seuls dans la rue au moins une fois par jour et 87% au moins une fois par semaine. Seulement 7,5% des personnes n'utilisent aucun système d'aide à la navigation mais elles ne sortent jamais seules. Cela confirme la nécessité des aides pour la mobilité. Les transports en commun sont très utilisés, 38% d'entre eux les utilisent au moins une fois par jour, 85% au moins une fois par semaine et seulement 7,5% ne les utilisent jamais. Il est à noter que ces derniers sont aussi ceux qui n'utilisent aucun système d'aide et ne sortent jamais seul. Les transports en commun, bien qu'ils ne soient pas assez accessibles semblent être primordiaux pour l'autonomie. Derrière ces données descriptives sur les

usages des non-voyants, cette étude a permis de mettre en valeur des besoins qui sont rarement satisfaits par les outils de suppléance existants.

L'analyse des résultats du questionnaire met en évidence trois catégories de besoins exprimés par les non-voyants : la navigation dans des environnements inconnus, la localisation d'obstacles et d'objets et la catégorisation d'objets semblables.

1) La navigation dans des environnements inconnus

La plupart des personnes qui utilisent un chien d'aveugle le trouvent beaucoup plus utile que la canne blanche car comme décrit plus haut, il est beaucoup plus adaptable à différentes situations. Les chiens-guide peuvent trouver le moyen le plus court pour arriver à une destination, emprunter les passages piétons, les trottoirs et adapter un parcours en fonction des obstacles distants et en hauteur, contrairement à la canne. Un sujet précisait que la seule chose qui manque au chien est sa capacité à interpréter les numéros de rue, les panneaux de signalisation, les feux tricolores. Ils ne sont en revanche pas acceptés par toutes les personnes : certaines personnes interrogées ne souhaitent pas avoir la compagnie d'un chien. De plus, ils ne sont pas acceptés dans tous les lieux publics (hôpitaux, commerces …). Le principal intérêt du chien pour aveugles est le fait qu'il aide simultanément les personnes non-voyantes à la fois pour l'orientation et le déplacement.

2) Localisation d'obstacles et d'objets

Un problème important est la détection des obstacles, spécialement ceux qui sont à hauteur de tête (extincteurs, bennes de camion, etc.) car ils ne sont pas détectés par la canne et parfois ignorés par les chiens guides. Les formes d'obstacles les plus difficiles à éviter sont longues et fines, verticales (ex. arrête d'une porte ouverte), ou horizontales (ex. le plateau d'une table, un toit). Les participants rapportent que les obstacles les plus dangereux en navigation sont les ruptures brutales de pente et le mobilier urbain mobile comme les véhicules.

3) Catégoriser des objets semblables

Localiser un objet par son motif ou sa silhouette est utile, mais peut parfois ne pas être suffisant lorsque les objets sont structurellement proches. Est-ce une boîte de petits pois ou d'ananas ? Est-ce ma quittance de loyer ou mon relevé d'identité bancaire ? Est-ce un billet de 10 ou de 20 euros ? Ces questions sont revenues de manière récurrente dans notre enquête et font ressortir le problème de la lecture, de l'acquisition d'information précise sur

un objet, un livre, une feuille de papier, un nom de rue, les chiffres d'un digicode, un distributeur de billet, un numéro de bus, un nom du magasin, un arrêt de bus …

La structure n'étant pas forcément pleinement informative, la couleur (feux de circulation, couleur des habits, etc.), et la luminosité (lumière allumée ou pas dans le cas d'un bouton poussoir ou d'un va-et-vient, temps qu'il fait dehors, etc.) sont des informations qui peuvent s'avérer utiles pour certains objets.

4) Discussion

Le besoin des sujets non-voyants réside dans de nombreuses tâches de la vie quotidienne impliquant différentes fonctions du système visuel. L'objectif est de répondre à ce besoin très large des personnes déficientes visuelles en restituant une ou plusieurs de ces fonctions particulièrement efficaces du système visuel pour l'aide à la navigation. L'élément commun de la vision humaine à la majorité de ces besoins, déficiente chez les personnes non-voyantes, est la faculté à parcourir et interpréter l'environnement de manière instantanée. C'est ce à quoi ont essayé de répondre les différents systèmes de substitution sensorielle qui ont été conçus depuis les années 1970. S'il est possible grâce à eux de parcourir l'environnement de manière instantanée, ces systèmes en revanche ne permettent pas d'interpréter une scène visuelle complexe. L'interprétation de la scène visuelle par les personnes voyantes est effectuée en reconnaissant des objets connus et en les localisant dans un espace tridimensionnel. Cette faculté permet aux personnes voyantes de s'orienter en prenant le plus court chemin entre deux points visuels, lire des noms de rue, reconnaître et localiser des objets et des obstacles. La navigation tient une grande part dans l'accroissement de l'autonomie des non-voyants. La restauration d'une fonction de localisation et de reconnaissance d'objets permettrait donc de répondre en partie au besoin utilisateur. En se trouvant sur une place, le fait de pouvoir instantanément se créer une représentation spatiale des différents objets du mobilier urbain et des bâtiments permettrait ainsi de s'orienter au point de destination, en créant son propre itinéraire et sans risque.

Nous pensons qu'il est possible d'utiliser la localisation de cibles visuelles dans l'environnement pour aider les personnes non-voyantes à naviguer par la reconnaissance d'amers visuels (métro, porte, passage piéton, …). Cette information, si elle est disponible à tout moment permet de garder un cap en restituant par la localisation continue d'un objet pour le suivre (ex. passage piéton), l'atteindre ou l'utiliser comme point de repère intermédiaire dans son parcours. Le fait de restaurer cette fonction du système visuel

permet de répondre en même temps à un deuxième besoin en reconnaissance d'objets de formes semblables mais dont seul le motif visuel à la surface de l'objet est différent.

L'idée que nous poursuivrons dans ce travail est donc qu'il est possible de répondre au besoin en restaurant une fonction du système visuel : la localisation d'objets par vision artificielle.

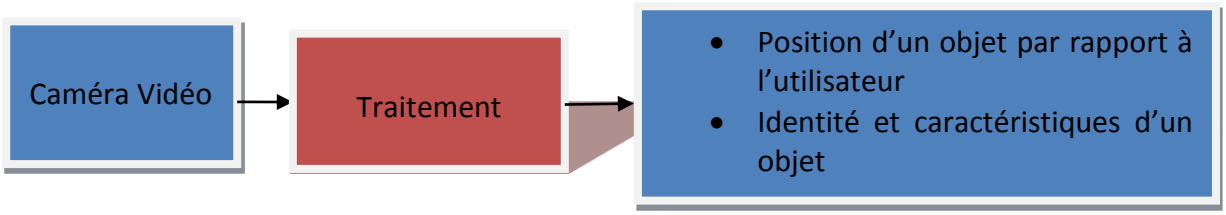


Figure 37: Informations qu'il est possible de restituer avec un dispositif de localisation de cibles visuelles

Nous allons dans un premier temps traiter la conception d'un système de suppléance pour les non-voyants en étudiant une manière de restituer de manière instantanée et fiable les deux informations définies comme indispensables à la navigation : l'identité d'un objet et sa position dans l'espace.

Analyse de la scène par vision artificielle

Un système de suppléance comporte toujours un capteur en entrée du système, pour recevoir les informations relatives à l'environnement ou au sujet. La caméra vidéo représente un capteur très proche de l'œil humain et permet ainsi d'acquérir l'essentiel de l'information non accessible par un non-voyant. Le signal ne peut être directement transformé pour une autre modalité sensorielle en raison de la grande quantité d'information qu'il contient. La scène doit être analysée pour ne restituer qu'une fraction des informations, utiles et interprétables par la personne non-voyante. Nous allons tenter de répondre au besoin utilisateur par l'identification et la localisation d'objets dans la scène visuelle. Cette information de localisation peut être restituée de manière simple par synthèse binaurale. La localisation d'objets à partir de caméras nécessite des algorithmes de traitement de l'image rapides et robustes. Nous avons vu dans l'état de l'art qu'il existait différentes catégories d'algorithmes permettant la reconnaissance d'objets, chacune avec ses avantages et ses inconvénients. En plaçant la nécessité d'un système temps réel pour guider les personnes non-voyantes, les méthodes par convolution nous sont apparues adaptées à la reconnaissance d'objets du mobilier urbain, souvent dans une même orientation ou à la reconnaissance d'objets en 3D (toutes les faces de l'objet sont modélisées). Nous discuterons de l'utilisation des autres algorithmes de vision dans

différentes tâches de reconnaissance et localisation où elles seraient plus adaptées que celle que nous allons présenter dans ce manuscrit. C'est à partir de l'observation du système visuel humain et de sa capacité de catégorisation très rapide d'objets dans des scènes naturelles (Thorpe et al., 2001a;Thorpe et al., 2001b;Thorpe and Fabre-Thorpe, 2001) qu'a été mis au point Spikenet (Thorpe et al., 2004), permettant de reconnaître et localiser des objets dans une image.

1) Fonctionnement de Spikenet

Pendant longtemps, la vision artificielle était inexploitée, principalement parce qu'elle nécessite des algorithmes complexes et très coûteux en calcul qui les rendaient inutilisables. L'interprétation d'une scène visuelle reste cependant un problème très complexe malgré la progression conjuguée des capteurs, des algorithmes et des calculateurs. L'apprentissage d'un classifieur d'objet est la plupart du temps supervisés (ils requierent un ensemble d'apprentissage). Le principal enjeu dans le temps de traitement des applications temps-réel est le compromis entre l'information extraite et le temps qu'il faut pour l'extraire. La plupart des algorithmes d'extraction robustes de cibles d'intérêt sont très coûteux en temps de calcul et ne sont pas appropriés à de l'analyse de scènes en temps réel. A l'inverse, les algorithmes simples et rapides ne fonctionnent pas dans des conditions naturelles où l'environnement lumineux change tout le temps.

Les principes par lesquels la reconnaissance et la localisation de cibles peuvent être accomplis sont étudiés à la fois en neurosciences et en vision artificielle. A l'interface de ces deux disciplines, S. Thorpe a proposé un modèle de codage de l'information, basé sur des observations physiologiques, permettant la reconnaissance rapide et la localisation d'objets dans la scène visuelle. Les résultats qu'il a obtenu avec son équipe en psychophysique et en modélisation (Thorpe et al., 2001b;Thorpe and Fabre-Thorpe, 2001) montrent que le système visuel est capable de détecter de façon extrêmement rapide une cible donnée. Les résultats de ces travaux ont donné naissance à un noyau de reconnaissance d'objets (SpikeNet). Cet outil est particulièrement efficace pour reconnaître des modèles et nous l'avons évalué dans le cadre d'un système de suppléance pour les non-voyants permettant de localiser et reconnaître des objets égarés ou bougés par des tiers mais aussi pour reconnaître des amers en navigation. Nous allons dans un premier temps présenter une application de Spikenet à la reconnaissance d'objets d'intérêt de forme similaire dans une situation très contrainte : la reconnaissance de billets de banque. Nous poursuivrons par une

évaluation de Spikenet dans un cadre plus général sur la base d'une évaluation empirique du comportement de Spikenet en situation contrôlée et dans des images réelles.

2) Détecteur de billets

Bien que différentes méthodes et outils existent pour aider les non-voyants à identifier les billets de banque, la plupart d'entre eux trient leurs billets à la maison afin de pouvoir les utiliser plus rapidement en situation. Certains systèmes dédiés sont capables de reconnaître les billets sur la base de capteurs vidéos [23] (Liu Xu, 2008). Dans le but d'évaluer notre approche de l'utilité de la vision dans un dispositif pour les non-voyants, nous avons donc réalisé un dispositif de reconnaissance de billets basé sur un capteur vidéo et utilisant Spikenet pour reconnaître les billets.

Matériel et méthodes

Le matériel de l'expérimentation était composé d'un UMPC (Ultra Mobile Personal Computer) SONY cadencé à 1,33Ghz et doté d'1 Go de mémoire vive. Le dispositif devra être utilisable en situation de mobilité, en complément d'un outil d'aide à la navigation porté sur la tête ou intégrée dans un téléphone portable. Une étude préalable a été nécessaire pour paramétrer Spikenet pour qu'il soit suffisamment sélectif aux billets et qu'il n'y ait aucune fausse détection. Nous avons donc fait apprendre au système les quatre billets les plus courants : 5, 10, 20 et 50 euros. La première étape a été d'isoler les régions caractéristiques des billets pour Spikenet : celles qui sont uniques sur chaque billet et qui comportent le plus de saillances. La couleur n'a aucune importance ici, seul le motif en niveaux de gris est utilisé. Il faudra de plus que cette zone soit facilement identifiable par les non-voyants pour pouvoir orienter le billet à reconnaître et rendre l'identification plus rapide. Les euros présentent la particularité d'avoir une bande verticale distinguable de manière tactile. Les zones d'intérêt pour le système se situent juste à coté de cette zone plastique qu'il suffit de repérer pour orienter le billet. La détection d'autres billets en euros ou dans d'autres monnaies n'a pas été évaluée ici mais serait aisément intégrable dans le prototype.

[2] knfbReader mobile by knfbReading Technology, Inc.
[3] Note Teller 2 by Brytech Inc. http://www.brytech.com/noteteller/

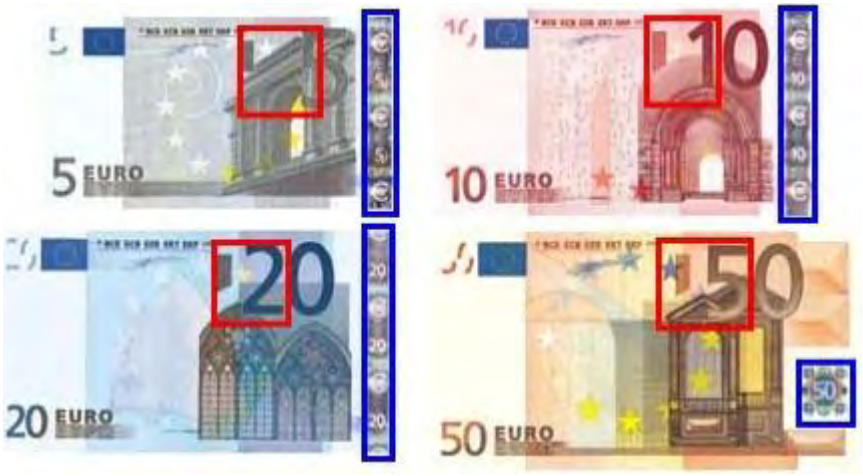


Figure 38: Les billets de 5, 10,20 et 50 euros. En bleu la bande plastique distinguable en passant le doigt, en rouge, les zones caractéristiques apprises et reconnues des billets

Figure 39: Les zones apprises et reconnues des 4 billets

L'apprentissage de ces zones résulte de choix techniques qui ont été faits pour rendre le système utilisable avec un temps de réponse court. Afin de rendre la détection invariante à la rotation, il faut générer autant de modèles qu'il y a d'orientations possibles pour couvrir 360°. Ceci est réalisé suivant un pas d'échantillonnage qui doit être judicieusement choisi, puisqu'il entraîne un compromis précision/nombre de modèles. Spikenet étant faiblement tolérant à la rotation, un pas de 12° a été choisi, ce qui correspond à une génération de 30 modèles pour chaque zone caractéristique. Le détecteur comptera donc ici 120 modèles activés simultanément. Le rapport d'échelle doit aussi être réglé pour permettre une tolérance à la distance entre le billet et la caméra. Les modèles Spikenet sont caractérisés par

- Le motif visuel qu'ils codent
- Le détail – Correspond à la quantité d'informations que l'on souhaite préserver dans l'image modèle.

- Le seuil de détection – Correspond à la qualité minimale de mise en correspondance entre une partie d'une image et le modèle pour qu'il soit considéré comme une détection.
- La taille minimale à laquelle ils seront reconnus (en pourcentage de l'image d'apprentissage)
- La taille maximale à laquelle ils seront reconnus (en pourcentage de l'image d'apprentissage)

La société Spikenet-technology ® préconise de ne pas utiliser de modèles dont la taille serait inférieure à 30x30 pixels pour que l'information contenue dans le modèle soit suffisamment sélective. Pour cela, la distance entre le billet et la caméra ne doit pas être trop grande pour la résolution de la caméra utilisée (320x240). Dans cette application, les modèles ont été paramétrés avec un seuil élevé (61), un niveau de détail très faible (35) permettant d'éliminer les hautes fréquences dans le modèle (traces de pliage, usure…). La tolérance à la taille est comprise entre 44% et 100% de la taille initiale (les modèles ont été appris à taille maximale : plein champ de la caméra).

Le dispositif permet la détection des billets à une fréquence de 4Hz sur l'UMPC indiqué plus haut et à une distance comprise entre 3 et 10 cm. Il n'y a eu aucun faux positif pendant l'expérimentation, ce qui illustre la fiabilité de la réponse du système. L'utilisateur a été au centre de la conception du système avec des entretiens et des tests réguliers avec des non-voyants tout au long du processus de développement. L'expérimentation s'est composée de quatre étapes : un questionnaire pré-expérimental, une phase d'apprentissage puis l'évaluation du dispositif et un questionnaire post-expérimental.

Questionnaire pré-expérimental

Tous les sujets sans exception possèdent un téléphone portable, et six sur sept d'entre eux utilisent couramment un ordinateur. Des logiciels et dispositifs spécialisés permettent une exploitation correcte de ces appareils par les non-voyants. Trois sujets possèdent et utilisent couramment d'autres dispositifs tels que des détecteurs de lumière, de billets, de couleur, ou encore des GPS pour piétons. Les autres sujets ont généralement déjà essayé ces dispositifs mais n'ont pas été satisfaits pour des raisons en accord avec l'expression de leur besoin. Les outils existants souffrent selon les personnes interrogées de trois principaux inconvénients :

- Un prix souvent prohibitif, du fait d'une concurrence assez limitée dans ce secteur d'activité, des faibles volumes de vente et des subventions parfois importantes dont peuvent bénéficier les non-voyants pour l'acquisition de dispositifs spécialisés,
- Une utilité qui n'est pas toujours justifiée, en particulier pour les personnes qui ne vivent pas seules, mais de manière plus générale tous affirment pouvoir faire sans,
- Un encombrement qui peut devenir pénalisant, puisque la plupart des dispositifs mobiles sont spécialisés dans une tâche, ce qui multiplie les appareils.

Apprentissage

Une phase d'apprentissage est effectuée par l'utilisateur en présence de l'expérimentateur qui lui explique le fonctionnement du système. L'utilisateur s'exerce alors à trouver rapidement la bande plastique à coté de laquelle la zone à identifier se trouve. Pour cela, les entretiens lors de la conception du dispositif ont permis d'élaborer des techniques particulièrement rapides et fiables. L'utilisateur plie ainsi le billet en quatre en laissant la partie du billet indentifiable visible. L'identification peut se décomposer en quatre étapes :

1. Repérer la bande plastifiée sur le billet
2. Plier le billet deux fois en gardant vers soi la bande plastifiée
3. Saisir la caméra en allongeant son index ou majeur sur la face arrière du billet pour en apprécier l'orientation
4. Coller l'objectif de la caméra au centre du billet plié et s'éloigner progressivement du billet jusqu'à ce que sa valeur soit énoncée

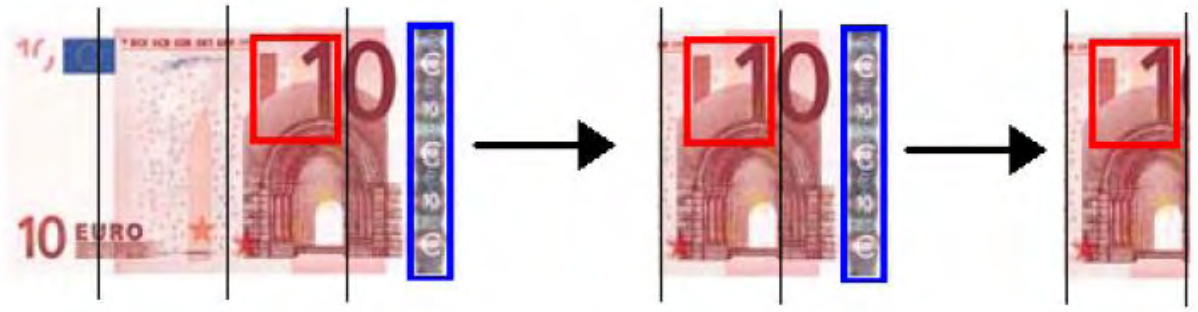


Figure 40: Etapes du pliage du billet en quatre dans l'ordre chronologique de gauche à droite. Le but est de le plier en deux puis encore en deux en gardant toujours la bande plastique vers soi. La zone identifiable est alors en dessous.

Les billets sont souvent très froissés, cette façon de procéder permet de rendre la partie à identifier plus rigide, sans pliage ni ombrage et ainsi augmenter la fiabilité de la détection. La phase où les sujets apprennent à utiliser le dispositif dure environ cinq minutes.

Expérimentation et évaluation du dispositif

Le sujet est assis devant une table et 5 billets sont empilés devant lui. La tâche consiste à classer les billets dans l'ordre croissant le plus rapidement possible. Les variables étudiées sont le temps mis pour reconnaître les cinq billets et le nombre d'erreurs d'identification. Chaque session expérimentale durait environ 30 minutes.

Résultats

Il a fallu moins d'une minute pour classer les cinq billets pour 5 sujets sur 7, un sujet a mis 1 minute et 10 secondes (sujet 7) et le dernier près de 2 minutes (sujet 3). Ce dernier a eu des difficultés lors de la prise en main du système, effectuant les manipulations trop rapidement pour que celui-ci ait le temps de détecter la zone. La moyenne des temps de reconnaissance par billet et par sujet (Figure 41) était d'environ 10 secondes. La plus grande partie de ces 10 secondes était allouée à l'identification tactile de la bande plastique.

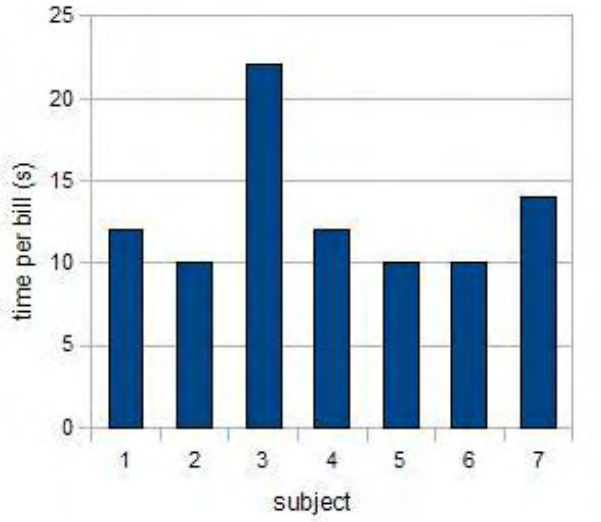


Figure 41: Moyenne du temps d'identification d'un billet, par sujet

Le principal facteur ayant rendu l'utilisation du dispositif difficile pour le sujet 3 est la détection de la bande plastique. Dans l'ensemble, les résultats sont très concluants malgré quelques remarques faites par les utilisateurs après l'expérience.

Questionnaire post-expérimental

A l'issue de ce test, une courte série de questions est posée aux sujets afin d'étudier leur expérience avec le système.

Que pensez-vous de l'utilité d'un tel outil ?

L'un des sujets est capable de reconnaître le montant d'un billet sans aucun appareil, en appréciant directement la taille du billet. Cette personne affirme donc n'avoir évidemment aucune utilité d'un tel dispositif. Les autres sujets répondent cependant unanimement qu'un tel dispositif est très intéressant pour améliorer leur autonomie. Deux personnes ont cependant soulevé le fait que si un tel dispositif de détection d'objets existait, ils l'utiliseraient principalement à domicile. Ils expliquent en effet qu'en extérieur, dans un hypermarché ou dans un restaurant, il est bien plus naturel de s'informer autour de soi que de s'isoler par l'utilisation du dispositif.

La prise en main de cet outil a-t-elle été difficile ?

La plupart des sujets ont eu une prise en main rapide et relativement naturelle. Deux d'entre eux ont mis un peu plus de temps mais le mode opératoire a été apprécié par la majorité des utilisateurs.

Le temps de réponse de l'outil vous paraît-il satisfaisant ?

Tous les sujets sont allés au terme de l'expérimentation et ont trouvé le temps de réponse tout à fait cohérent avec ce que l'on peut attendre d'un tel outil.

Quels sont les défauts évidents que vous auriez pu constater ?

Cinq sujets ont émis des remarques sur la qualité sonore des sons qui sont joués à la détection d'un billet. Ces sons ont en effet été enregistrés à l'aide d'un micro-casque classique dans une atmosphère relativement bruyante, et sont effectivement de qualité médiocre. L'évaluation portait plus sur la faisabilité d'un système de reconnaissance de billets par vision artificielle plutôt que sur le mode de restitution. Les sujets avaient cependant conscience qu'il ne s'agit pas d'un réel défaut technique, et que les choses peuvent être améliorées très rapidement concernant ce point.

Les premiers sujets auxquels aucun conseil particulier n'a été donné à priori sur le placement de la caméra ont tous relevé la difficulté de manipulation due à la liberté de mouvement, qui peut entraîner un mauvais cadrage. Les sujets à qui on a indiqué une procédure optimale pour la reconnaissance n'ont pas fait de remarques à ce sujet et ne trouvent aucun défaut majeur au dispositif. On pourra remarquer qu'un des sujets exprime une préférence pour un retour par vibrations (une pour cinq euros, deux pour dix euros, etc.) plutôt que par le son, principalement dans un souci de discrétion.

Que pensez-vous de faire fonctionner une telle application sur un téléphone portable ?

Tous les sujets sont très enthousiastes quant à la perspective de faire fonctionner une telle application sur un téléphone portable. Exceptés deux sujets qui possèdent et utilisent couramment un outil du marché et le sujet autonome dans la reconnaissance des billets, tous les sujets ont affirmé qu'ils feraient l'acquisition du programme si son prix était en accord avec sa qualité.

Conclusion

Les utilisateurs ont été très enthousiastes à l'utilisation d'un tel système malgré quelques réticences à l'utiliser au quotidien principalement à cause de leurs expériences passées avec d'autres systèmes de suppléance. Plusieurs d'entre eux ont en effet mentionné qu'ils ne l'utiliseraient que s'ils étaient seuls ou chez eux plutôt que de s'isoler en public avec un tel appareil. Cette expérimentation conforte notre idée qu'il est possible avec très peu

d'apprentissage, d'utiliser une caméra pour reconnaître des billets. Un système porté sur la tête par exemple et avec lequel d'un seul geste en le passant devant la caméra on pouvait en connaître le montant serait alors utilisé fréquemment en autonomie, seul ou avec des personnes autour. Il est utile de relever aussi de grandes différences inter-sujets dans la durée de prise en main du dispositif en lui-même mais aussi dans la méthode de pliage qui s'est avérée difficile pour un sujet (identification de la bande plastique) mais qui, avec l'usage, deviendrait très rapide. Cette méthode permet de lever 2 difficultés principales : le cadrage du billet est facilité, le billet est plus rigide et présente ainsi moins de pliures qui pourraient gêner l'identification par reconnaissance d'image.

3) Reconnaissance et localisation de cibles

La conception de l'outil de suppléance que nous proposons repose sur la localisation d'objets pour les non-voyants. Nous proposons un système de reconnaissance et de localisation d'objets en 3D, basé sur un capteur vidéo. Le système de reconnaissance et de localisation de cibles visuelles dans une image 2D repose sur l'utilisation de Spikenet. Les coordonnées 3D de l'objet reconnu sont calculées par reconstruction tridimensionnelle dans le référentiel de la caméra dans un modèle sténopé. La Figure 42 rappelle le schéma de fonctionnement d'un outil de suppléance visuelle avec un capteur d'entrée, le traitement de l'information capturée et sa restitution.

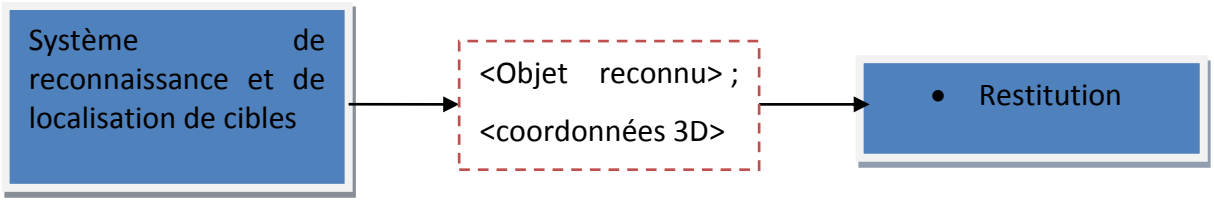


Figure 42: Informations disponibles à la restitution dans un dispositif de localisation de cibles visuelles

Nous allons dans cette partie développer une méthode d'évaluation de l'outil Spikenet pour une application dans un contexte de suppléance visuelle puis décrire la méthode de reconstruction tridimensionnelle utilisée. Nous avons choisi une méthode d'évaluation empirique pour étudier le fonctionnement de l'algorithme afin d'évaluer de manière fonctionnelle l'utilisabilité de ce système de suppléance pour les non-voyants.

Évaluation d'un outil de reconnaissance de formes en 2D: Spikenet

L'utilité d'un système de reconnaissance de formes pour la localisation d'objets dans un flux vidéo repose sur sa capacité à le localiser dès qu'il apparaît dans l'image avec un temps de traitement acceptable pour notre application. En situation réelle, de nombreux facteurs réduisent la fiabilité des systèmes de vision artificielle : des reflets, des ombrages, du flou, du bruit, etc. L'évaluation de l'outil de reconnaissance de forme est primordial pour notre application puisqu'elle détermine le type d'objets qu'il est possible de localiser et les situations dans lesquelles le système pourra ou ne pourra pas être efficient. L'étude présentée ici a été effectuée en collaboration avec Rémi Parlouar, un stagiaire de Master 2 Informatique que j'ai encadré.

Matériel et méthodes :

Les tests ont été effectués sur un ordinateur doté de 2 Go de mémoire RAM, sur un pentium bi-cœur de 2 GHz, sous Windows XP. Spikenet se présente sous la forme d'une bibliothèque, intégrée dans un environnement d'évaluation programmé en C++. Le chargement et le traitement des images sont effectués par la librairie Open Computer Vision, spécialisée dans l'analyse d'images. Nous avons voulu, en utilisant l'outil d'apprentissage supervisé Spikenet, connaître la tolérance, la précision et la qualité des détections dans différentes situations. Nous utiliserons dans la suite de cette étude, différentes conventions :

- Le taux de vrais positifs (True Positive Rate – tpr) est le taux de détections d'un modèle, là où il est effectivement présent : la détection est alors valide.
- Le taux de faux positifs (False Positive Rate – fpr) est le taux de détections d'un modèle là où il n'est pas présent : la détection est abusive.
- Une détection est valide si son centre se situe dans le polygone du modèle d'apprentissage et que le centre de la forme à reconnaître se situe dans le polygone de la forme reconnu dans l'image. Une deuxième contrainte concerne la taille de la détection : la différence entre la taille du modèle reconnu et celle de la région à reconnaître ne doit pas dépasser 20% de la taille de la région à reconnaître.

La précision de détection est le quotient du nombre de vrais positifs par la somme du nombre de vrais et faux positifs. Une précision valant 0 signifie que toutes les détections sont des faux positifs. Une précision de 1 indique que toutes les détections sont des vrais positifs. Ce taux ne peut pas être calculé si aucune détection n'est apparue dans l'image (division par 0).

- La qualité de détection est une valeur renvoyée par Spikenet pour estimer le niveau de confiance pour une détection donnée. Le seuil de détection permet de filtrer les détections en fonction de la qualité de détection. Seuls les modèles dont la qualité est supérieure à Seuil/100 ne son conservés.

Les tests ont été effectués en trois parties : la première évalue l'impact des paramètres intrinsèques à Spikenet (seuil, niveau de détail, taille de modèle) sur la détection, la localisation et le temps de traitement. Pour cette première partie, le corpus d'images de test est composé d'un logo de la Poste, de 2 silhouettes humaines, d'une voiture, d'une plaque de rue et de 6 formes géométriques (Figure 43). Dans le but d'évaluer la robustesse des détections avec Spikenet, le corpus de test est le même que celui d'apprentissage. Les tests nous renseignent sur la tolérance de l'algorithme aux déformations de l'image ou du modèle.

Figure 43 : Corpus d'images de test. Les zones d'apprentissage sont détourées en rouge dans les 5 premières images (de gauche à droite et de haut en bas : le logo de la poste, un piéton, la signalétique d'un passage piéton, une voiture de profil, et une plaque de rue). Le dernier corpus d'apprentissage concerne des formes géométriques primitives.

La deuxième partie des tests évalue la robustesse de l'algorithme à différentes altérations qui peuvent affecter l'image. Dans une troisième partie, nous étudierons la fiabilité de l'algorithme dans des images en extérieur.

Résultats

L'établissement d'une matrice de confusion permet de mesurer la fiabilité d'un classifieur. De nombreuses méthodes existent pour visualiser et analyser une matrice de confusion. Le coefficient de Matthews, calculé à partir d'une matrice de confusion est un bon indice de la fiabilité et de la précision d'un système. Dans cette étude, la fiabilité du système sera exprimée en calculant de coefficient :

$$Mcc = \frac{\mathrm{TP} * \mathrm{TN} - \mathrm{FP} * \mathrm{FN}}{\sqrt{(\mathrm{TP} + \mathrm{FP})(\mathrm{TP} + \mathrm{FN})(\mathrm{TN} + \mathrm{FP})(\mathrm{TN} + \mathrm{FN})}}$$

Avec TP le nombre de vrais positifs, TN le nombre de vrais négatifs, FP le nombre de faux positifs et FN le nombre de faux négatifs. Cette mesure est comprise entre 1 et -1 : la valeur 1 est donnée pour une classification parfaite, -1 pour une classification exactement inverse et 0 pour une classification de type aléatoire. Le nombre de faux négatifs dans l'image est l'ensemble des positions dans l'image où l'objet est présent et qu'il n'est pas détecté. Ce nombre est très faible ce qui fait que FP*FN ne sera jamais supérieur à TP*TN. Le coefficient de Matthews ne sera donc jamais négatif dans notre cas.

Influence des paramètres internes à Spikenet sur la qualité des détections

Les tests de l'influence des paramètres seuil et détail du modèle ont été effectués en prenant comme modèle le cercle, le feu pour piéton, la voiture de profil et la ligne horizontale. Les résultats obtenus confirment le fonctionnement connu de ces paramètres (Figure 44) :

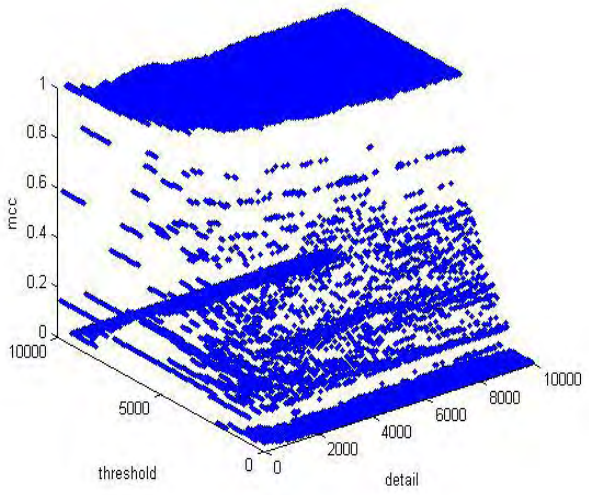


Figure 44: Coefficient de Matthews en fonction du seuil et du détail pour la détection et la localisation du feu pour piéton dans l'image source. Nous obtenons des résultats semblables avec les 3 autres modèles.

- Le seuil de détection du modèle est conçu pour filtrer toutes les détections dont la qualité est inférieure à Seuil/100. La fiabilité du système augmente avec le seuil : les fausses détections sont éliminées avec l'augmentation du critère de qualité de détection.

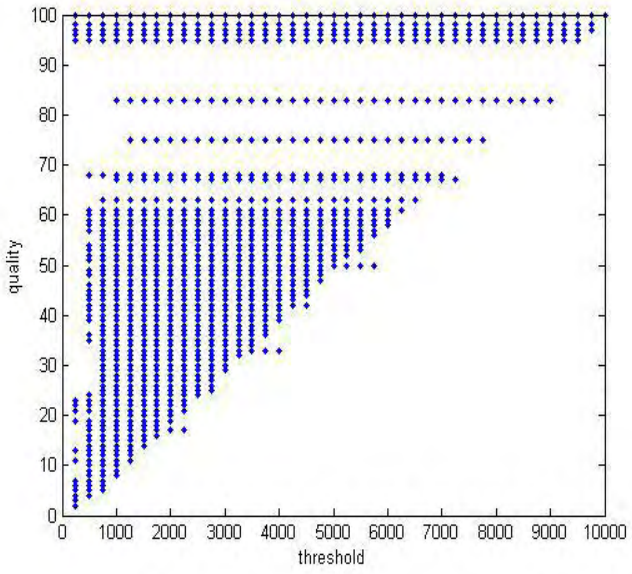


Figure 45: Qualité de détection du feu pour piéton en fonction de chaque valeur de seuil possible (comprise entre 0 et 10000), par pas de 250, pour toutes les valeurs de détail possibles (entre 0 et 10000) par pas de 250. Nous observons bien un seuillage de la qualité : le modèle est accepté si la qualité est supérieure à seuil/100.

La Figure 45 montre que le seuil joue bien un rôle d'exclusion de mauvaises détections. Lorsqu'une détection est effectuée, l'algorithme fournit un pourcentage de qualité de la détection, qui est un très bon indice sur sa fiabilité. C'est sur cette valeur que le seuil joue le rôle de valeur limite d'acceptation. Les détections sont valides si $Qualité > \frac{Seuil}{100}$.

- Le détail du modèle influe sur la résolution de celui-ci : la quantité de saillances à utiliser pour caractériser la forme à reconnaître. La Figure 46 montre qu'un niveau de détail faible diminue la marge de détection entre les vrais positifs (qualité supérieure à 80 par exemple) et les faux positifs. Si on voulait seuiller cette composante pour ne garder que les modèles à partir d'une certaine qualité, la marge de seuillage serait beaucoup plus faible avec un détail faible qu'avec un détail élevé. Il semble donc qu'un détail élevé soit garant d'une meilleure fiabilité de détection, essentiellement pour rejeter les faux positifs par une méthode de seuillage.

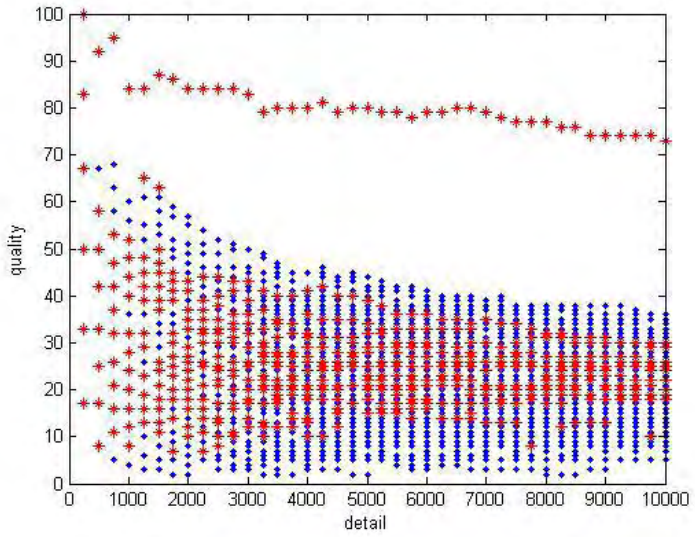


Figure 46: Influence du détail sur la qualité des détections pour le seuil le plus bas : sans aucun filtrage sur la qualité des détections. Les vrais positifs sont affichés en rouge et les faux positifs en bleu. Le nombre de vrais positifs augmente considérablement comme le nombre de faux positifs car certains d'entre eux sont dans le polygone entourant l'objet. Ils sont en fait des faux positifs mais considérés comme des vrais positifs puisqu'ils sont localisés sur le motif à reconnaître.

Le seuil et le détail influent donc bien sur les détections. Le seuil n'influe pas sur la qualité des détections mais ne retient que celles dont la qualité est suffisante. Le détail en revanche, agit sur la qualité des détections, rendant les faux positifs de meilleure qualité pour un détail faible et augmentant ainsi la difficulté d'un seuillage optimum. Un seuil trop élevé (100 par exemple) rejetterait toutes les détections. Cette valeur doit être déterminée de manière

optimale pour rejeter les faux positifs et conserver la majorité des vrais positifs. La valeur de détail permettant la meilleure classification entre les vrais positifs et les faux positifs est sa valeur maximale : 10000.

Temps de traitement

L'utilisabilité d'un système de suppléance est extrêmement dépendante de sa réactivité pour répondre au besoin. Les paramètres internes à Spikenet (seuil et détail) influent sur le temps de traitement. Un seuil bas engendre une hausse du temps de traitement. Ce résultat peut être expliqué par le fait qu'un seuillage bas engendre une liste de modèles reconnus beaucoup plus grande et donc un traitement plus long.

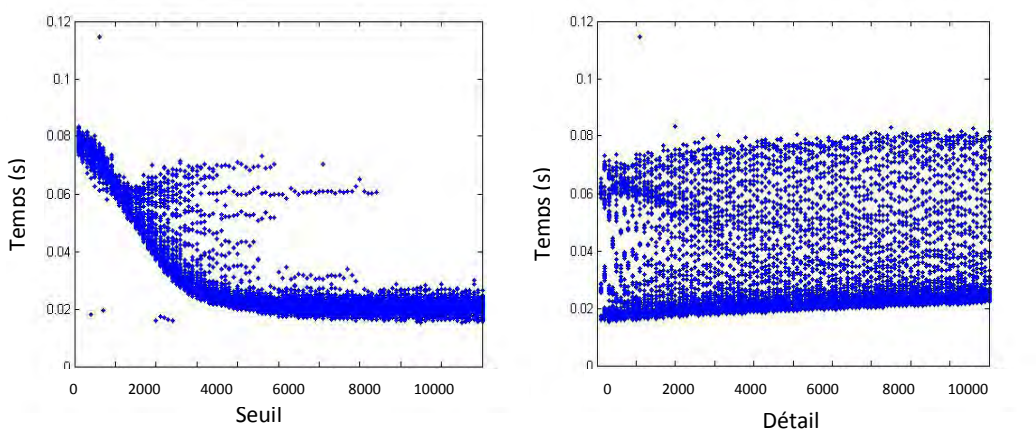


Figure 47: A gauche : influence du seuil pour toutes les valeurs de détail (entre 0 et 10000 par pas de 200) sur les temps de traitement. A droite, influence du détail pour toutes les valeurs de seuil (entre 0 et 10000 par pas de 200) sur les temps de traitement. Le temps de calcul diminue avec l'augmentation du seuil et la diminution du détail.

L'augmentation du détail fait légèrement augmenter le temps de traitement. Nous avons vu précédemment que le détail du modèle était proportionnel à la quantité de saillances retenues pour le caractériser. Un haut niveau de détail entraîne une augmentation du nombre de critères de comparaison, ce qui explique ce résultat. Nous pouvons noter que la différence de temps de traitement pour le modèle testé (voiture) entre les deux valeurs extrêmes de détail est relativement faible.

La localisation d'objets dans l'espace environnant nécessite de connaître l'influence du nombre de modèles à reconnaître sur le temps de prétraitement et de traitement. Le

prétraitement est défini pour Spikenet comme les opérations préparatoires sur l'image avant toute reconnaissance. Ce temps est donc par définition indépendant du nombre de modèles chargés. Le temps de traitement est quand à lui linéaire par rapport au nombre de modèles à reconnaître dans une image. Ces deux résultats se retrouvent dans les figures Figure 48 et Figure 49.Pour obtenir ces résultats, le temps de traitement et de prétraitement a été étudié indépendamment dans un cas particulier : la reconnaissance de <nbModels> occurrences d'un même modèle dans une même image. Le temps de traitement augmente linéairement avec le nombre de modèles activés. Comme on pouvait s'y attendre, le temps de prétraitement n'est pas fonction du nombre de modèles activés puisqu'il correspond au temps de préparation de l'image pour la reconnaissance.

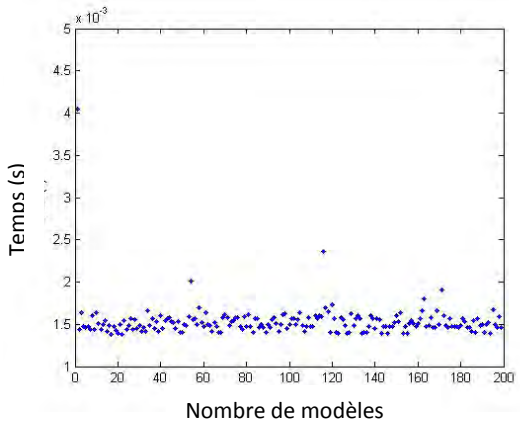


Figure 48: Temps de prétraitement en fonction du nombre de modèles identiques activés

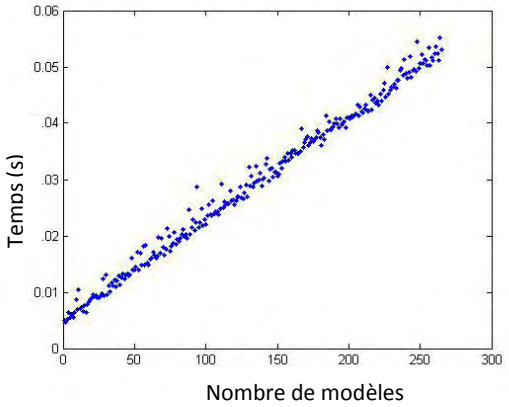


Figure 49: Temps de traitement en fonction du nombre de modèles identiques activés

Les résultats obtenus dans les deux figures précédentes confirment les hypothèses que nous avions quant au comportement du temps de prétraitement et de traitement dans Spikenet. Nous pouvons noter que le comportement est le même avec n'importe quelle image et n'importe quel modèle. Nous avons ensuite voulu savoir si nous ne nous trouvions pas dans un cas particulier avec l'ajout d'un modèle identique au premier. Nous avons pour cela construit quatre modèles particuliers : un premier modèle simple, un deuxième modèle de la même taille que le premier mais plus complexe, un troisième modèle identique au premier mais deux fois plus grand, et un quatrième modèle identique au deuxième mais deux fois plus grand. Les graphes suivants ont été construits en ajoutant 50 fois le premier modèle, puis 50 fois le second, etc. (modèles ajoutés un par un). Les résultats sont présentés dans la Figure 50 et la Figure 51.

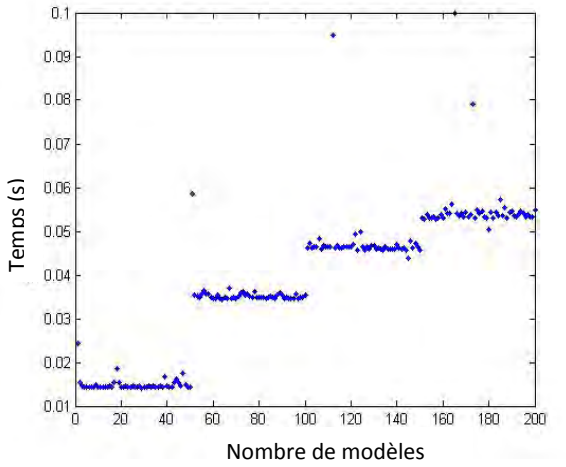


Figure 50: Temps de prétraitement en fonction du nombre et du type de modèles activés. 50 modèles sont ajoutés un par un pour chacun des 4 types successivement. Contrairement au résultat précédent (Figure 49), le temps de prétraitement n'est pas constant.

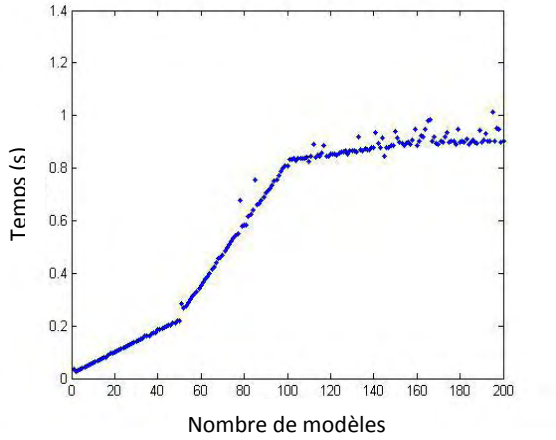


Figure 51: Temps de traitement en fonction du nombre de modèles activés. Contrairement à l'exemple précédent, quatre modèles différents ont été choisis : en faisant varier la complexité et la taille des modèles. Chaque modèle a été ajouté cinquante fois (un par un) puis le suivant avec le même mode opératoire jusqu'au 4^{ième} (200 modèles en tout).

Le temps de traitement est donc bien linéaire par rapport à l'ajout d'un même modèle mais pas en fonction du nombre de modèles activés : cela dépend de la complexité de chaque modèle. De même, le temps de prétraitement qui était défini comme constant dans une même image ne l'est pas et dépend du nombre de modèles activés. Nous pouvons remarquer que le temps de traitement est linéaire par blocs de modèles identiques et la pente d'augmentation du temps de traitement est plus faible pour les 100 derniers ajouts (mêmes modèles que les 100 premiers mais deux fois plus grands). Afin de ne pas se placer dans le cas particulier de l'ajout de modèles différents mais ayant un lien étroit, 5 modèles très différents ont été choisis et étudiés suivant le même mode opératoire que précédemment. Cette fois, chaque modèle est ajouté un par un 30 fois au lieu de 50 fois. Les résultats présentés dans la Figure 52 et la Figure 53 confirment les résultats obtenus précédemment.

Le temps de prétraitement n'est pas constant pour une image et il dépend du nombre et du type de modèles à reconnaître.

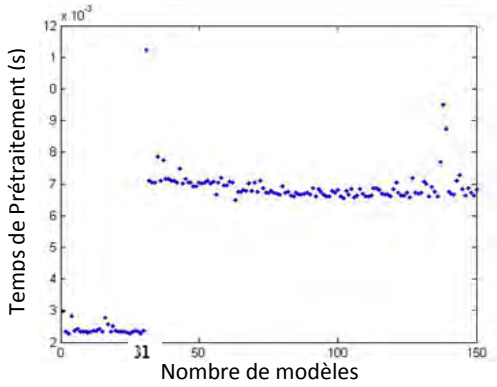


Figure 52: Temps de prétraitement en fonction du nombre et du type de modèles activés. 5 modèles différents sont activés un par un par tranche de 30.

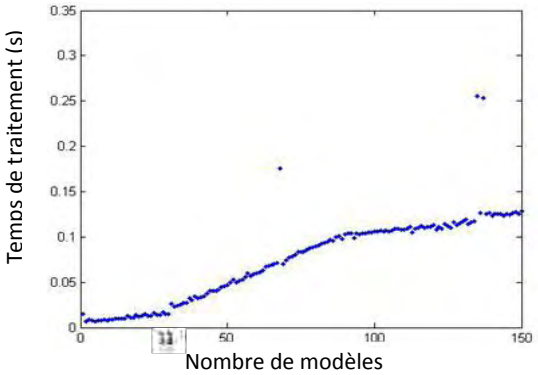


Figure 53: Temps de traitement en fonction du nombre et du type de modèles activés. 5 modèles différents sont ajoutés un par un par tranche de 30.

Nous observons sur la Figure 52 et la Figure 53 quatre blocs de croissance linéaire des temps de traitement, ce qui signifie que parmi les cinq modèles utilisés, deux ont un temps de traitement suffisamment proche pour que la croissance du temps reste linéaire (il s'agit de l'intervalle [30;90] en abscisse). Nous constatons cependant sur la Figure que l'ajout du deuxième modèle (abscisse 31) provoque un 'saut' du temps de prétraitement, ce qui n'est pas le cas pour l'ajout des trois modèles suivant, alors que les modèles sont tous différents. Le passage à un nouveau palier n'est donc pas systématique pour chaque ajout d'un nouveau modèle et n'est pas prévisible à ce stade du fonctionnement de Spikenet. Nous pouvons donc conclure cette partie avec les affirmations suivantes :

- Le temps de prétraitement total n'est pas la somme des temps de prétraitement de chaque modèle activé. Il peut être constant, ou évoluer brutalement suite à l'ajout d'un modèle de taille et complexité très différentes. Le temps de prétraitement étant le temps de préparation de l'image source pour la reconnaissance, les opérations faites sur l'image dépendent du type de modèles activé. Nous verrons en particulier par la suite, l'influence du facteur d'échelle du modèle sur le temps de prétraitement.
- Le temps de traitement total équivaut globalement à la somme des temps de traitement de chaque modèle activé. Ceci n'est vrai que sur un processeur mono-

cœur ; une parallélisation des processus de prétraitement et de traitement sur des processeurs multi-cœurs permettrait de réduire les temps de reconnaissance.

Le temps de traitement et de prétraitement semble donc très dépendant des caractéristiques du modèle activé. Parmi celles qui définissent un modèle, nous allons à présent étudier l'impact des paramètres de taille du modèle sur le temps de traitement et de prétraitement. Le modèle est testé dans toutes les positions possibles de l'image. Il paraît alors trivial que le rapport entre la taille de l'image et la taille du modèle activé joue un rôle très important dans le temps de traitement. Nous faisons l'hypothèse qu'une réduction de 50% de chaque dimension du modèle augmente de 4 fois le temps de traitement en quadruplant le nombre de positions possibles pour le modèle dans l'image. Autrement dit, le temps de traitement serait inversement proportionnel au carré de la taille du modèle.

La Figure 54 et la Figure 55 montrent que les temps de prétraitement et de traitement sont bien inversement proportionnels au carré de la taille du modèle. Ces tests ont été effectués pour un même modèle dans une même image. Le temps de traitement et de prétraitement est étudié en faisant varier la taille du modèle recherché de 20% à 200% de sa taille initiale par de 20%.

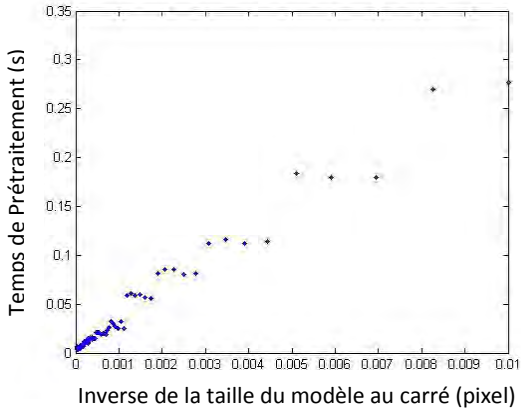


Figure 54: Temps de prétraitement en fonction de l'inverse de la taille du modèle au carré. La taille du modèle a été testée de 20% à 200% par pas de 10%.

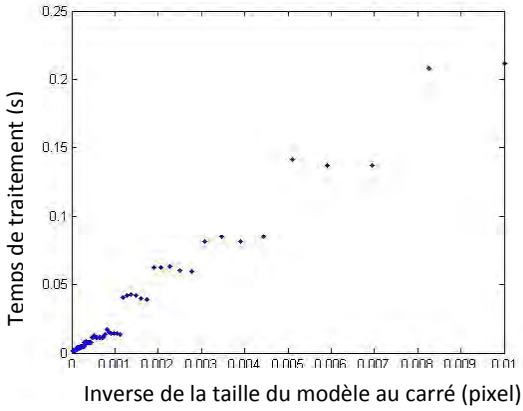


Figure 55: Temps de traitement en fonction de l'inverse de la taille du modèle au carré. La taille du modèle a été testée de 20% à 200% par pas de 10%.

Il aurait été intéressant d'étudier les temps de traitement en fonction du rapport entre l'image et le modèle à reconnaître pour pouvoir les prédire. Il n'a pas été possible d'établir une telle formule, les temps n'étant pas liés seulement à ce rapport, la formule est

extrêmement complexe à établir. Nous pouvons en revanche remarquer que le temps de traitement et de prétraitement affichent des résultats par paliers en fonction de la taille des modèles. Nous pouvons émettre l'hypothèse que le paramètre de taille des modèles agit en réalité avec des valeurs prédéfinies et que, toutes les valeurs de tailles ne sont pas testées (seules des tailles de référence sont testées).

Influence et tolérance de la rotation sur la reconnaissance de forme

Pour pouvoir reconnaître un objet dans toutes les situations possibles, il faut être tolérant à la rotation tout en restant sélectif au modèle. Nous pourrons ainsi conclure sur le nombre de modèles nécessaires pour couvrir toutes les orientations sur 360° de rotation. Les modèles testés sont les mêmes que précédemment dans les mêmes images. Nous avons dans un premier temps voulu mesurer l'impact de la rotation sur la qualité des détections. Comme nous pouvions nous y attendre, un cercle est bien invariant à la rotation et la qualité de détection ne varie pas avec l'angle.

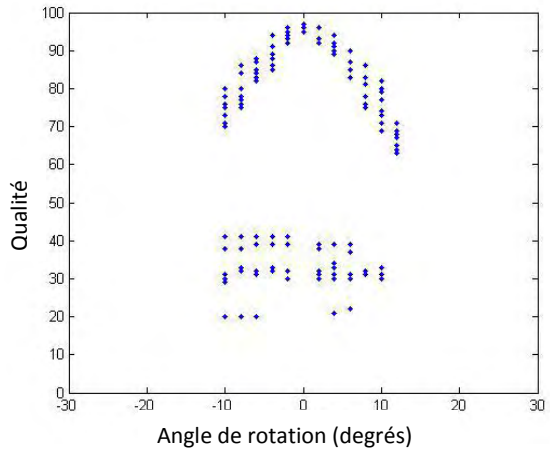


Figure 56: Le carré : Qualité de reconnaissance en fonction de l'angle de rotation

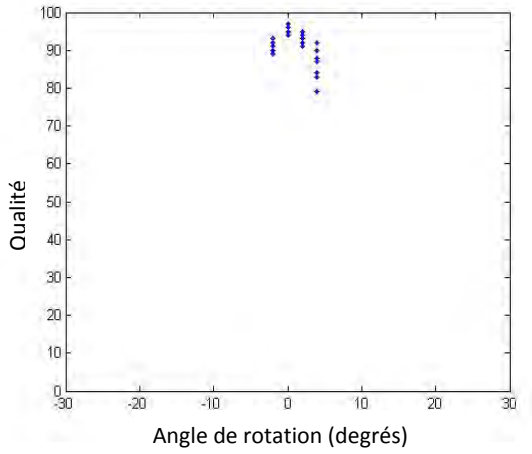


Figure 57: La ligne horizontale : Qualité de reconnaissance en fonction de l'angle de rotation

Nous pouvons remarquer que pour le carré (Figure 56) et pour la ligne horizontale (Figure 57), la rotation influence la qualité des détections de manière symétrique. Plus l'angle de rotation du modèle est grand, plus la qualité du modèle diminue. Nous pouvons voir que la tolérance à la rotation n'est pas la même pour tous les modèles. Pour un seuil de qualité à 70, la tolérance à l'orientation sera de +-10° pour le carré et +-4° pour la barre horizontale. Plus le modèle sera allongé ou asymétrique, moins Spikenet sera tolérant à la rotation. La principale difficulté réside dans le fait que la tolérance à la rotation est très dépendante du modèle et il sera donc impossible de prédire avec précision avec quel pas de rotation il faudra créer des modèles pour qu'un objet soit reconnu à 360°. Cet échantillonnage devra

donc être effectué indépendamment pour chaque modèle. Nous savons cependant que le temps de reconnaissance est affecté par l'ajout de nouveaux modèles, c'est pourquoi il est important de minimiser le nombre de modèles à générer. Pour déterminer le nombre de modèles optimal, il faut diviser une rotation complète par l'intervalle de tolérance à la rotation. L'intervalle de tolérance correspond à l'intervalle continu de valeurs de rotations au dessus d'un seuil choisi. Ce seuil est choisi le plus bas possible en faisant en sorte qu'il n'y ait pas de faux positifs. Nous avons cependant mis en évidence que le seuil et le détail ont une forte influence sur cette tolérance, et que plus le réglage de ces paramètres est bas, plus la tolérance est importante (au risque de faire apparaître des faux positifs). Pour reconnaître la ligne horizontale un seuil à 7000 (seules les qualités supérieures à 70 sont prises en compte) se traduit par un intervalle de tolérance en rotation de 8°. Il faudra donc 360/8=45 modèles pour le reconnaître dans toutes les orientations. Il n'en faudra en revanche que 16 pour reconnaître le carré avec un seuil de 6000 car sa tolérance à la rotation est plus grande.

Influence du facteur d'échelle sur la reconnaissance

Comme pour la rotation, le facteur d'échelle d'un modèle Spikenet est très important puisqu'il permet de le reconnaître, non seulement à l'échelle à laquelle il a été créé. Cela revient à créer des modèles différents à différentes tailles avec un pas permettant une reconnaissance continue entre les différentes échelles. Les figures suivantes (Figure 58, Figure 59, Figure 60, Figure 61) affichent la qualité des détections en fonction du facteur d'échelle pour la ligne horizontale, le cercle, la croix et la plaque de rue. La ligne horizontale et la croix sont invariantes en qualité pour un facteur d'échelle supérieur à 1. On peut en effet toujours trouver une croix dans une croix plus grande ou une ligne dans une ligne plus longue. Ce n'est pas le cas pour un facteur d'échelle inférieur, ce qui explique une baisse de qualité linéaire pour ces deux modèles pour un facteur d'échelle inférieur à 1. Un modèle d'objet plus complexe qu'un carré ou qu'une ligne droite comportera une multitude de détails qui le rendra moins robuste au facteur d'échelle. Le cercle et la plaque de rue sont plus représentatifs de l'utilisation de Spikenet en situation réelle pour reconnaître des objets. Si l'on choisit un seuil de qualité de 50 (seuil en-dessous duquel la plaque de rue crée de nombreuses fausses détections), la tolérance au facteur d'échelle est alors d'environ 10% maximum.

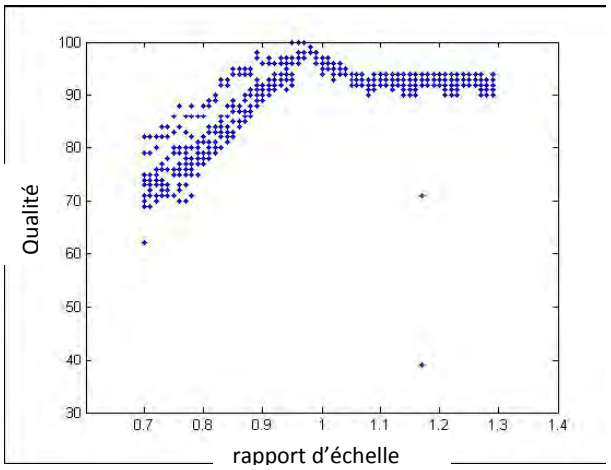


Figure 58: Ligne horizontale; Qualité des détections en fonction du facteur d'échelle. Les points de faible qualité correspondent à des fausses détections.

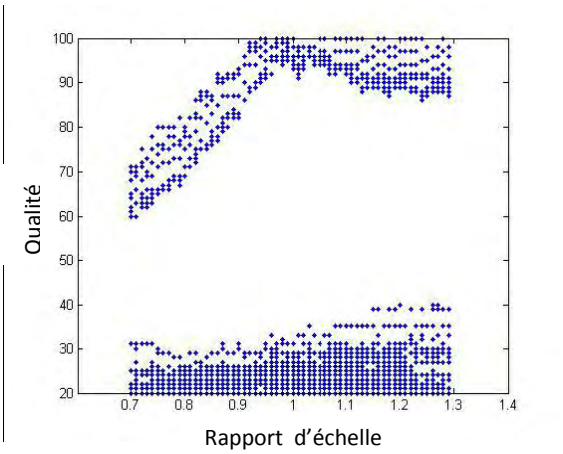


Figure 60: Croix; Qualité des détections en fonction du facteur d'échelle. Les points de faible qualité correspondent à des fausses détections.

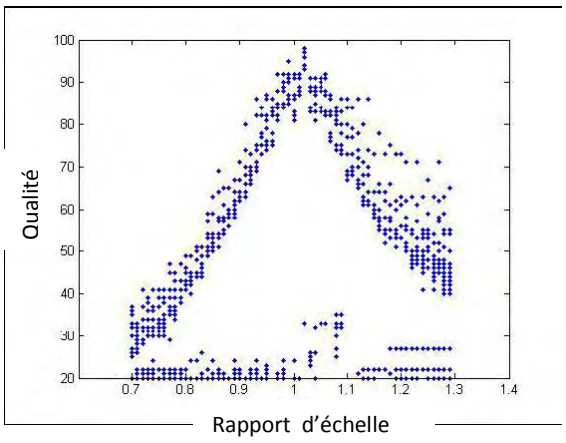


Figure 59: Cercle; Qualité des détections en fonction du facteur d'échelle. Les points de faible qualité correspondent à des fausses détections.

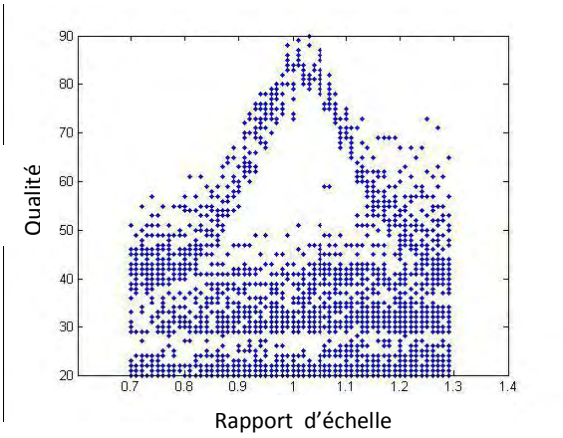


Figure 61: Plaque de rue; Qualité des détections en fonction du facteur d'échelle. Les points de faible qualité correspondent à des fausses détections.

L'étude de la qualité des détections en fonction du rapport d'échelle pour une tolérance de l'algorithme de reconnaissance variant entre 10% et 190% par pas de 1% nous renseigne sur le fonctionnement interne de ce paramètre. En effet, l'étude de la qualité des détections (coefficient de corrélation entre la détection et le modèle Spikenet associé) va nous permettre de comprendre comment ce paramètre fonctionne. La Figure 62 et la Figure 63 montrent la qualité des détections avec un facteur d'échelle testé variant entre 40% et 130%. Le modèle à reconnaître est testé dans l'image contenant l'objet recherché, à toutes

les échelles entre 40% et 130% par pas de 1%. Une qualité de 100% représente un maximum de corrélation entre le modèle recherché et la détection. Ces deux figures montrent que la courbe de la qualité en fonction du rapport d'échelle du modèle pour une tolérance continue entre 40% et 130% est en fait cyclique avec un maximum tous les pas de 20% environ.

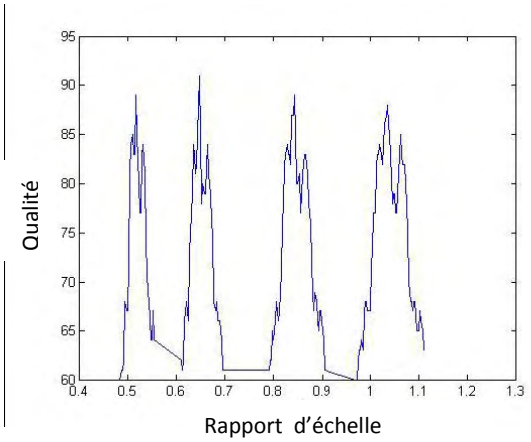


Figure 62: Qualité de détection en fonction du rapport d'échelle de l'objet à détecter (ici le cercle). SizeMin et sizeMax ont été fixés à 10% et 190%.

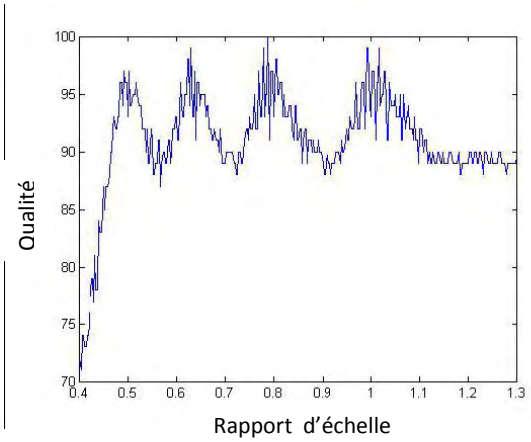


Figure 63: Qualité de détection en fonction du rapport d'échelle de l'objet à détecter (ici la plaque de rue). SizeMin et sizeMax ont été fixés à 10% et 190%.

Ces résultats montrent que ces paramètres sizeMin (rapport minimal de la taille du modèle initiale (en pixels) à laquelle l'algorithme pourra reconnaître l'objet) et sizeMax (rapport maximal de la taille du modèle initiale (en pixels) à laquelle l'algorithme pourra reconnaître l'objet) donnent un intervalle au sein duquel le rapport d'échelle n'est pas testé de manière continue. Pour un seuil de tolérance de qualité inférieur à 60, le modèle sera donc reconnu de manière continue mais avec une qualité variable. Avec un seuil de tolérance supérieur, le modèle ne sera reconnu que de manière discontinue. Spikenet teste donc le modèle dans l'intervalle [sizeMin ; sizeMax] de manière discrète par pas de 20% avec pour origine 100%. Il ne servira donc à rien d'avoir un paramètre sizeMin ou sizeMax compris entre 90% et 110% puisqu'une seule échelle sera évaluée : 100%.

Temps de traitement en fonction du ratio Modèle/Image

Comme nous l'avons vu précédemment, il n'est pas possible avec les connaissances que nous avons, de prédire le temps de prétraitement et de traitement en fonction de la taille du modèle et de la taille de l'image à traiter. Nous allons donc nous baser sur des relevés effectués sur un ordinateur de bureau équipé d'un processeur Intel Xeon 3GHz et de 1 Go de mémoire vive, sous Windows XP SP2 avec 3 images de contenu quelconque et de résolutions différentes. Les trois images étaient respectivement de résolution 320*240, 864*648 et 2067*2923. Les Figure 64, Figure 65, et Figure 66 illustrent le temps de traitement total relevé en fonction du ratio Image/modèle pour les 3 images. Le ratio correspond à la taille dans chaque dimension du modèle par rapport à l'image.

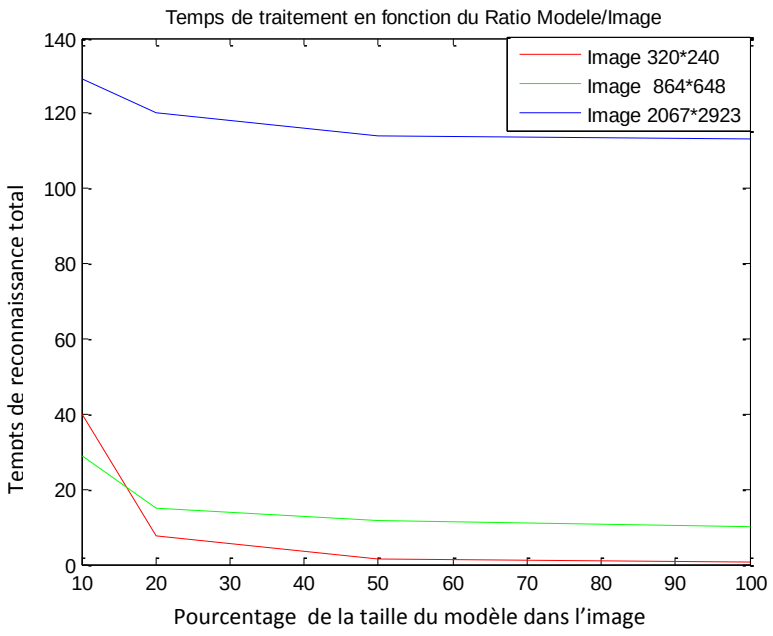


Figure 64: Temps de reconnaissance total en millisecondes en fonction du pourcentage de la taille du modèle dans l'image

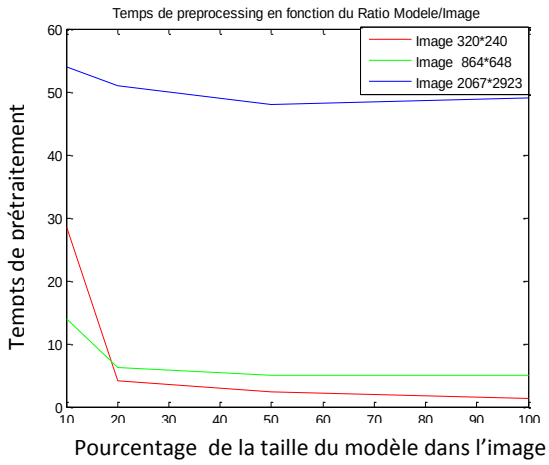


Figure 65: Temps de prétraitement en millisecondes en fonction du pourcentage de la taille du modèle dans l'image

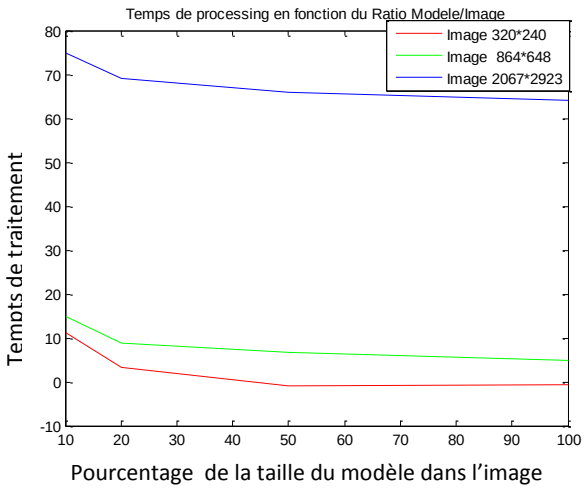


Figure 66: Temps de traitement en millisecondes en fonction du pourcentage de la taille du modèle dans l'image

Le temps de traitement étant exponentiel en fonction du ratio taille du modèle / taille de l'image, la recherche d'objet devient inutilisable pour notre application pour des ratios inférieurs à 10%. Seuls les résultats pour des ratios supérieurs à 10% seront donc utilisés. Le temps de traitement pour les 3 images diminue quand la taille du modèle dans l'image augmente avec des pentes différentes selon que le ratio est inférieur à 20% ou supérieur. Il semble en effet dans les 3 images que le temps de traitement augmente de manière exponentielle avec un ratio inférieur à 20%. Le temps global et le temps de prétraitement est plus élevé pour l'image 1 (320*240) que pour l'image 2 (864*648). Nous avons vu que la complexité de l'image et des modèles jouait un rôle dans le temps de traitement, ces images avaient un contenu différent. Cette remarque peut expliquer ce résultat pour des ratios très bas mais n'explique pas pourquoi le temps total et de prétraitement est ensuite ordonné en fonction de la résolution de l'image.

Nous confirmons ici les résultats précédemment présentés :

- Le temps de traitement varie très peu en fonction du ratio modèle/image
- Le temps de prétraitement n'est pas constant pour une image mais il est fonction du modèle à détecter : sa taille et sa complexité.

Modélisation du temps de traitement de Spikenet dans un cas réel : la détection d'un panneau de 80 cm de côté

Afin de se placer dans un cas pratique, nous avons effectué des statistiques à partir des temps relevés précédemment pour la détection d'un panneau de signalisation de 80 * 80 cm, filmé avec une caméra suivant les 3 résolutions précédemment utilisées, avec un angle de vue de 100°.

La largeur L en pixel d'un objet de taille (largeur) LR (m), à une distance d (m) dans une image dont la résolution en largeur est W pixels, avec un champ de vue alpha est donnée par la formule :

$$L = LR * W / (2d * \tan(alpha/2))$$

La largeur en pixels en fonction de la distance à l'objet ainsi que le temps de traitement extrapolé en tenant compte de cette largeur en pixels est présenté dans les 3 figures suivantes (Figure 67, Figure 68, Figure 69) pour les 3 images de résolutions différentes.

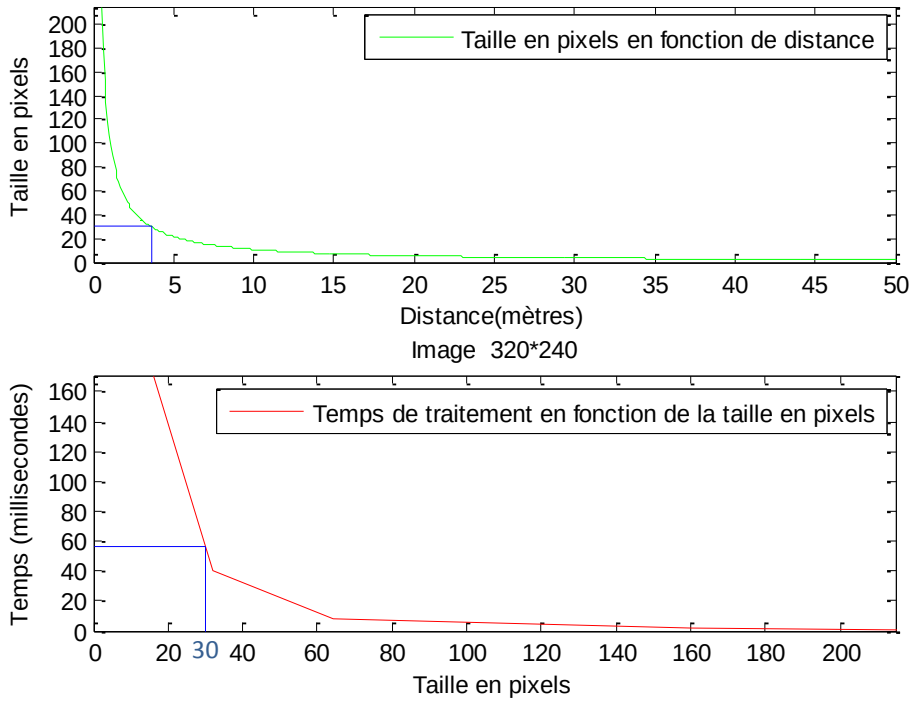


Figure 67: En haut : Largeur en pixel d'un panneau stop (80cm) en fonction de la distance. En bas : Temps de traitement en fonction de la largeur en pixels du modèle dans l'image de 320*240. La ligne bleue correspond dans le graphique du haut à la distance maximale (4m - abscisse) à laquelle le panneau stop pourra être reconnu par Spikenet (>30 pixels de largeur). L'équivalence taille en pixel/temps de reconnaissance est affichée dans le graphique du bas avec la ligne bleue permettant de connaître le temps de traitement nécessaire (55 ms – ordonnée) pour un modèle de 30 pixels.

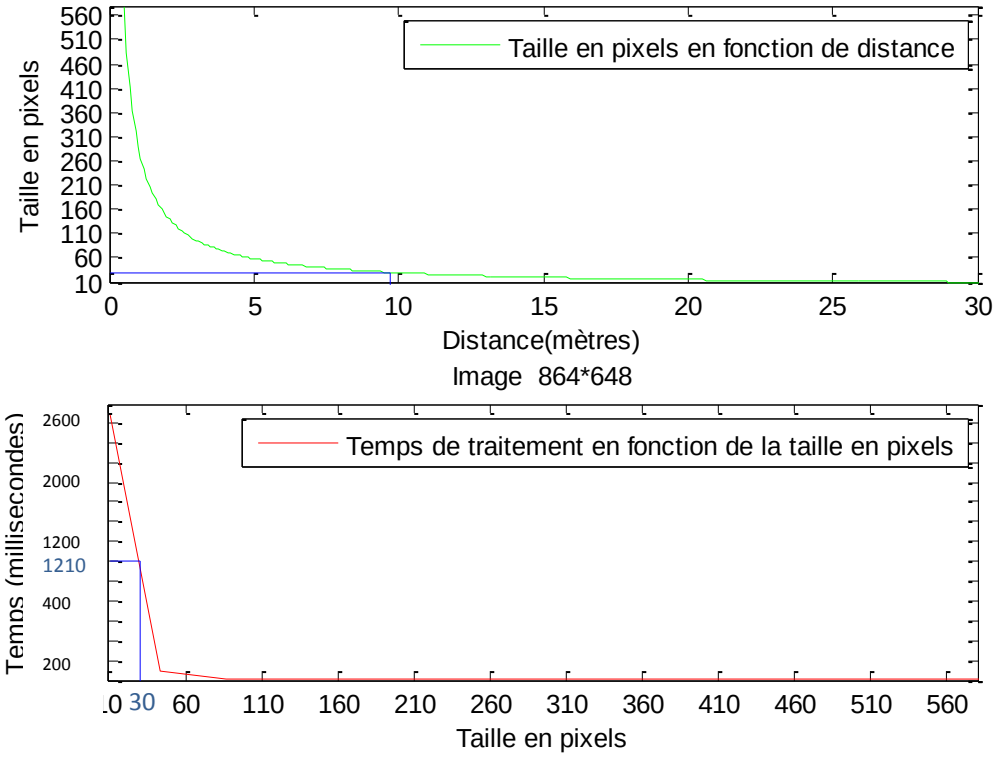


Figure 68: largeur en pixel d'un panneau stop (80cm) en fonction de la distance (en haut); Temps de traitement en fonction de la largeur en pixel du modèle dans l'image de 864*648. La ligne bleue correspond dans le graphique du haut à la distance maximale (9m - abscisse) à laquelle le panneau stop pourra être reconnu par Spikenet (>30 pixels de largeur). L'équivalence taille en pixel/temps de reconnaissance est affichée dans le graphique du bas avec la ligne bleue permettant de connaître le temps de traitement nécessaire (1210 ms – abscisse) pour un modèle de 30 pixels.

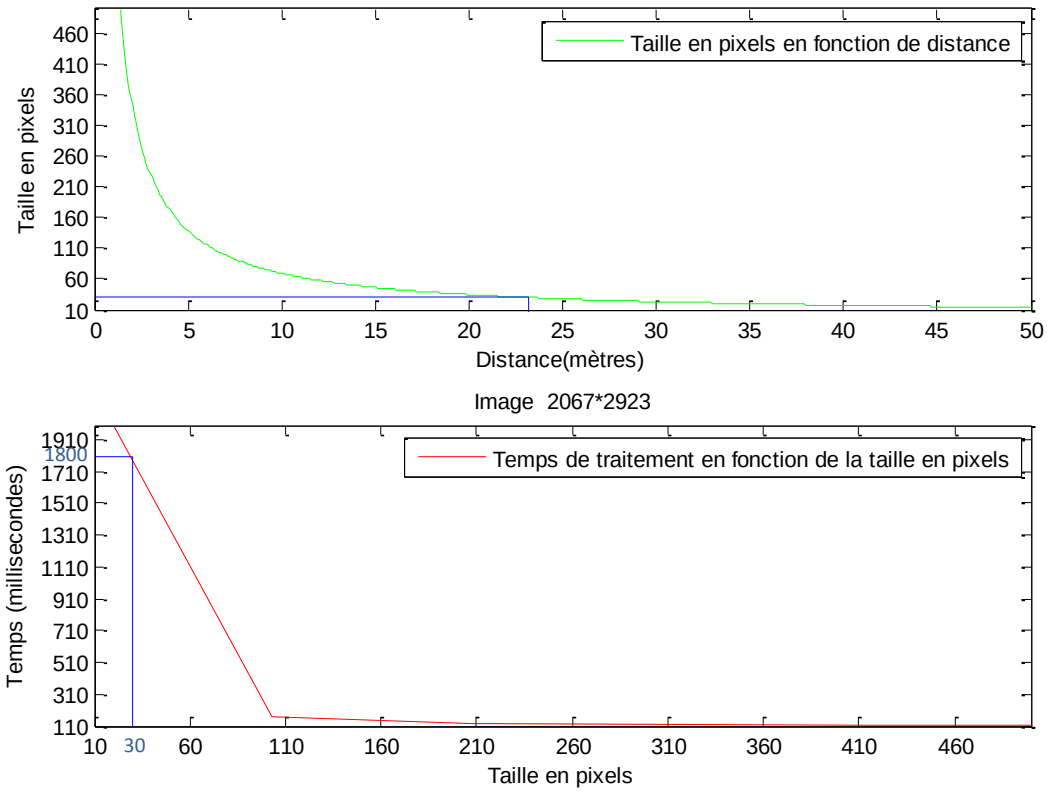


Figure 69: largeur en pixel d'un panneau stop (80cm) en fonction de la distance (en haut); Temps de traitement en fonction de la largeur en pixel du modèle dans l'image de 2067*2923. La ligne bleue correspond dans le graphique du haut à la distance maximale (12m - abscisse) à laquelle le panneau stop pourra être reconnu par Spikenet (>30 pixels de largeur). L'équivalence taille en pixel/temps de reconnaissance est affichée dans le graphique du bas avec la ligne bleue permettant de connaître le temps de traitement nécessaire (1800 ms – abscisse) pour un modèle de 30 pixels.

Nous considérerons en suivant les recommandations de la société Spikenet qu'un modèle dont la taille est inférieure à 30 pixels dans une dimension ne sera pas suffisamment sélectif pour être reconnu et localisé avec précision. La Figure 67 montre qu'avec le dispositif utilisé, un panneau de signalisation (Stop par exemple) atteint donc sa taille minimale pour être reconnu (30*30 pixels) à 4 m avec une résolution de 320*240 pour un champ de vision de 100°. A cette taille minimale, le temps de traitement total inférieur à 40 ms. Avec une résolution de 864*648 pour 100° de champ (Figure 68), le modèle est détectable jusqu'à 8,5m (distance à laquelle le panneau mesurera 30 pixels x 30 pixels). Le temps de traitement nécessaire à sa localisation est de 1210 ms. Avec une résolution de 2067*2923 et toujours 100° de champ, le modèle est détectable jusqu'à 12m, le temps de traitement est alors de 1800 ms.

Résolution de l'image	Distance à laquelle l'objet fait 30x30 pixels	Temps de traitement
320*240	4 m	40 ms
864*648	8,5 m	1210 ms
2067*2923	12 m	1800 ms

Le temps de calcul est un élément très important de l'utilisabilité d'un système de suppléance visuelle. Ces résultats montrent l'importance de la résolution sur le comportement de Spikenet. Il paraît en pratique difficile d'augmenter la résolution au-delà de 320x240 pixels avec une caméra grand angle comme celle utilisée du fait de la très petite taille des objets dans l'image. Il est à noter que les temps présentés ne représentent que les temps de détection d'un seul modèle à une échelle donnée et sans rotation. Le seul moyen d'augmenter la distance de détection est de réduire l'angle de vue pour une même résolution.

Calcul du temps de traitement d'un panneau de signalisation quelconque de 80 cm, ne présentant aucune propriété géométrique particulière dans le motif, entre 2 m et 4 m :

La tolérance au facteur d'échelle est établie pour un modèle quelconque à 5%. Il faudra donc entre 2 et 4 m créer 7 modèles de panneaux, avec une durée totale de détection de 280 ms.

Distance en mètres	Largeur en pixels du modèle	Temps de traitement	Temps de prétraitement	Temps de traitement
2,1053	48,82	22,82	15,72	7,09
2,3269	44,75	26,92	18,8	8,11
2,57	41,30	30,40	21,43	8,9
2,84	37,03	34,71	24,67	10,04
3,14	33,56	38,22	27,31	10,9
3,47	30,68	50,56	36,09	14,47

3,83	27,53	76,37	54,28	22,08

Pour être exhaustif, il faudrait créer tous les modèles de rotation pour chaque échelle. Le temps total doit donc être multiplié par le nombre de modèles nécessaires en rotation. Il paraît ici particulièrement utile de contraindre la reconnaissance en fonction du contexte et supposer que le panneau de signalisation sera par exemple toujours à l'endroit.

Afin d'avoir de meilleures performances, quelques recommandations peuvent ainsi être formulées :

- Un modèle ayant une forme géométrique permettant une tolérance à la rotation diminue grandement le temps de traitement. (Un cercle par exemple)
- Un modèle ayant une forme géométrique permettant une tolérance au rapport d'échelle diminue grandement le temps de traitement. (Une croix par exemple)
- Une meilleure tolérance aux fausses détections permet de diminuer fortement le nombre de modèles puisqu'elle rend Spikenet plus tolérant à la rotation et au facteur d'échelle.
- Le temps de traitement d'un modèle dont la taille dans chaque dimension est inférieure à 20% rend le temps de traitement rédhibitoire.

Discussion

Le logiciel de reconnaissance de formes Spikenet intègre de nombreux paramètres internes d'optimisation de la reconnaissance en fonction du modèle et de la scène dans laquelle le reconnaître. Le paramètre détail règle la complexité du modèle à reconnaître. Un détail élevé diminue le nombre de fausses détections mais augmente le temps de traitement (nombreux détails à tester). Un détail faible diminue le temps de traitement mais augmente le nombre de fausses détections (moins de détails décrivent le modèle). La qualité du modèle est un indice de corrélation (confiance) entre le modèle à détecter et une détection. Le « seuil » correspond à la valeur minimale de qualité pour laquelle la détection est acceptée comme correcte, celui-ci est par défaut à 50 et correspond à un bon compromis sans optimisation. La tolérance au facteur d'échelle est définie par les paramètres sizeMin et sizeMax correspondant à l'intervalle de reconnaissance. Ces deux valeurs sont des coefficients de facteur d'échelle par rapport au modèle initial (ex. 50%). La courbe de qualité

en fonction du rapport d'échelle par pas de 1% montre que dans cet intervalle [sizeMin ; sizeMax], seules les tailles correspondant à 100% + n * 20% sont testées, n étant un nombre entier positif ou négatif. En effet, la qualité des détections dans l'intervalle [sizeMin ; sizeMax] montre un maximum de qualité atteint tous les 20% avec 100% pour origine (la taille initiale du modèle). Un seuil de tolérance trop élevé rendra donc la détection discontinue sur l'intervalle [sizeMin ; sizeMax], la qualité des détections étant plus élevée proche des tailles de modèles réellement testées par Spikenet. Ce pas de test des détections a donc été fixé à 20% pour tous les modèles, vraisemblablement sur la base de tests de tolérance au facteur d'échelle effectués par la société développant le logiciel. Pour la reconnaissance d'une forme ne présentant aucune propriété géométrique particulière, Spikenet est tolérant à la rotation d'environ 4° et au facteur d'échelle d'environ 5% pour un seuil de détection autour de 60. La tolérance peut augmenter en fonction des propriétés géométriques du modèle à reconnaître : un cercle sera tolérant à la rotation sur 360° et une croix sera infiniment tolérante à une augmentation du facteur d'échelle. Ceci étant, un détail et un seuil faibles permettront une meilleure tolérance au changement d'échelle et à la rotation mais engendreront plus de fausses détections. Cette tolérance est directement dépendante du modèle et de la scène dans laquelle la détection sera opérée : il faudra parfois augmenter le seuil pour diminuer le nombre de fausses détections pour un objet dont le coût d'une erreur (gravité d'une erreur de reconnaissance du dispositif de suppléance pour les non-voyants) sera très élevé ou un objet comportant très peu de détail caractéristiques. Dans ce cas, la tolérance diminue et dans le cas d'une reconnaissance dans un intervalle d'échelle, le pas de test devra être faible. Le seuil de 20% choisi par Spikenet est donc arbitraire et convient la plupart du temps mais les paramètres sizeMin et sizeMax deviennent inadaptés pour des réglages fins du logiciel. Ces paramètres ne font rien d'autre que tester les modèles à des tailles différentes par pas de 20% comprises dans l'intervalle. Il sera ainsi préférable de créer l'ensemble des modèles nécessaires dans l'intervalle d'échelle nécessaire avec le pas adapté à chaque modèle pour une reconnaissance continue. Sur le même principe, la tolérance à la rotation est variable et il convient de créer l'ensemble des modèles nécessaires dans l'intervalle de détection voulu.

Le temps de traitement n'est pas constant en fonction de la complexité du modèle et du rapport taille de modèle / taille d'image. Il est difficile d'établir la fonction permettant de prédire le temps de traitement en fonction du modèle et de l'image utilisée. On peut en revanche établir une prédiction moyenne sur la base de relevés pour donner un ordre de

grandeur du temps de traitement. Il est en effet possible de relever le temps de traitement en fonction de la résolution de l'image, de la taille du modèle dans des situations moins contraintes que précédemment : en mesurant le temps de traitement d'un objet réel à différentes distances dans des images réelles. Il est possible d'établir la taille d'un modèle en pixels en fonction de la distance à l'objet, sa taille réelle et le capteur utilisé. A partir de cela et des relevés précédents, il est possible de calculer une approximation du temps de traitement d'un objet en fonction de ses caractéristiques physiques et des situations dans lesquelles il doit être reconnu (intervalle de rotation et d'échelle, coût d'une fausse détection …). En tenant compte du nombre de modèles nécessaires à la reconnaissance, il est possible d'interpoler le temps de traitement total nécessaire. Cette méthode de prédiction a ainsi été appliquée à un panneau de signalisation de 80cm x 80cm à détecter entre 2 m et 4 m (tolérance à l'échelle de 5%) avec une caméra grand angle (100°), une résolution de 320 pixels x 240 pixels et pour toutes les orientations (tolérance à l'orientation de 4°). Le nombre très élevé de modèles nécessaires dans ce cas très général engendre des temps de reconnaissance incompatibles avec une utilisation en temps réel. Dans la réalité, il est possible d'optimiser cette situation. Par exemple, dans le cas d'une application à un système de suppléance pour les non-voyants porté par la tête, une détection pour l'ensemble des orientations n'est pas nécessaire. L'intervalle auquel des modèles devront être créés doit être établi en fonction d'une étude comportementale sur les sujets non-voyants afin d'établir différents profils de mouvements de la tête en mobilité. Il convient donc de paramétrer les modèles de chaque objet indépendamment. Ces résultats montrent qu'il est possible en adaptant et en optimisant l'outil de vision Spikenet de reconnaître des zones d'intérêt en temps réel. Spikenet tire sa force de sa rapidité de reconnaissance pour un modèle en 2 dimensions avec une certaine tolérance à l'échelle et à la rotation. Il est en effet particulièrement adapté pour reconnaître des cibles contraintes comme des panneaux de signalisations par exemple qui ne seront jamais vus à l'envers. D'autres méthodes de reconnaissance d'objets plus tolérantes sont étudiées pour reconnaitre des objets dont le facteur d'échelle et l'orientation varient beaucoup. C'est le cas de SIFT (Lowe, 1999) ou SURF (Bay et al., 2008) dont la reconnaissance est multi-échelle et invariante à la rotation. Ces méthodes sont largement plus coûteuses en termes de temps de calcul que Spikenet dans des environnements contraints comme des objets dont l'orientation varie très peu. Elles peuvent en revanche être un excellent complément dans le cas où le nombre de modèles

nécessaires à la reconnaissance s'envole du fait de la faible tolérance de Spikenet aux changements géométriques présents à faible distance.

En navigation, les problèmes de reconnaissance sont sensiblement différents. Il est possible le long d'un parcours de contraindre fortement la reconnaissance par vision artificielle : les objets ou amers visuels sont souvent dans un même contexte visuel. Il est possible ainsi de ne créer que des modèles de grande taille, qui seront reconnus et localisés très rapidement par le système de vision. Un portage de ce système de vision sur un système embarqué dédié équipé d'une caméra permettrait de réduire grandement les temps de traitement. Une manière de s'intéresser au problème de la taille des objets par rapport à l'angle de vue de la caméra utilisée est d'utiliser une méthode multi-échelle. Il est alors possible d'effectuer un passage à basse résolution sur une image en 320x240 avec 100° d'angle de vue et d'affiner la reconnaissance en effectuant un zoom numérique sur l'image et n'effectuer une reconnaissance que sur une partie de l'image à haute résolution (320x240 mais pour une toute petite zone de l'image). Il faut pour cela que le capteur soit de très haute résolution pour pouvoir effectuer un zoom numérique de bonne qualité dans l'image.

Image entière :

Résolution : 320x240

Champ de vue : 100°

Zoom numérique :

Résolution : 320x240

Champ de vue : adapté

Figure 70 : Illustration du fonctionnement du système de reconnaissance et localisation multi-échelle. L'image du haut représente une capture d'une caméra à basse résolution sur la totalité de son champ visuel ; l'image du bas représente un zoom numérique d'un endroit de l'image à haute résolution.

Ce mode de fonctionnement est adapté à une utilisation en temps réel avec Spikenet puisque chaque modèle recherché est de grande taille dans l'image et un parcours exhaustif de la scène rapidement est alors possible. Une seconde passe sur des zones d'intérêt à haute résolution permettrait ainsi de désambigüiser la reconnaissance. Il est alors possible en cherchant un distributeur de billets dans la rue de reconnaître à basse résolution une façade associée et ensuite, par un zoom numérique, de localiser le distributeur de billets avec précision.

Les différents algorithmes de reconnaissance d'objets sont plus ou moins bons en fonction de la situation dans laquelle ils sont utilisés. C'est à partir de cette idée que nous avons optimisé une utilisation de Spikenet en temps réel dans un seul but : répondre au besoin de manière rapide et fiable. Pour ce faire, nous avons couplé la reconnaissance d'objets en 2D dans une image avec une méthode pour reconstruire les coordonnées 3D de l'objet reconnu par Spikenet. Différentes méthodes permettent d'estimer la position 3D des amers visuels dans une image ; celle que nous avons développée dans le cadre du projet Navig est une estimation par stéréovision de la distance à un objet reconnu par Spikenet.

4) Reconstruction tridimensionnelle

La reconstruction tridimensionnelle stéréoscopique permet à partir d'images provenant de deux caméras de connaître les coordonnées 3D d'un objet présent simultanément dans le champ des deux caméras. Une première phase de calibrage consiste à établir la relation précise entre les images issues des deux caméras. Lors de l'estimation des coordonnées 3D, une mise en correspondance de la position d'un même objet sur chaque image permet de connaître la position tridimensionnelle de l'objet à partir des données de calibrage. Un premier dispositif de capture a été créé à partir de deux caméras miniatures analogiques synchronisées et branchées à une carte d'acquisition. Ce premier prototype nous a permis d'évaluer la précision de localisation du système couplé à Spikenet. Le choix s'est ensuite tourné vers un périphérique de capture stéréoscopique du commerce afin d'augmenter la précision du système et s'abstraire des considérations techniques de l'estimation de la position 3D des objets, notamment celles liées au calibrage.

Première version du dispositif de capture

Ce premier dispositif était réalisé avec deux webcams montées sur une paire de lunettes puis calibrées.

Figure 71 : Première version du dispositif de vision : deux caméras miniatures analogiques sont disposées sur une monture de paire de lunettes. Le sujet est de plus équipé d'un micro et d'un casque audio.

Le calibrage d'une caméra consiste à évaluer les **paramètres intrinsèques,** internes à la caméra et les **paramètres extrinsèques** qui peuvent varier suivant la position de la caméra dans l'espace de travail. Nous avons utilisé un modèle sténopé, couramment utilisé en vision artificielle (Figure 72).

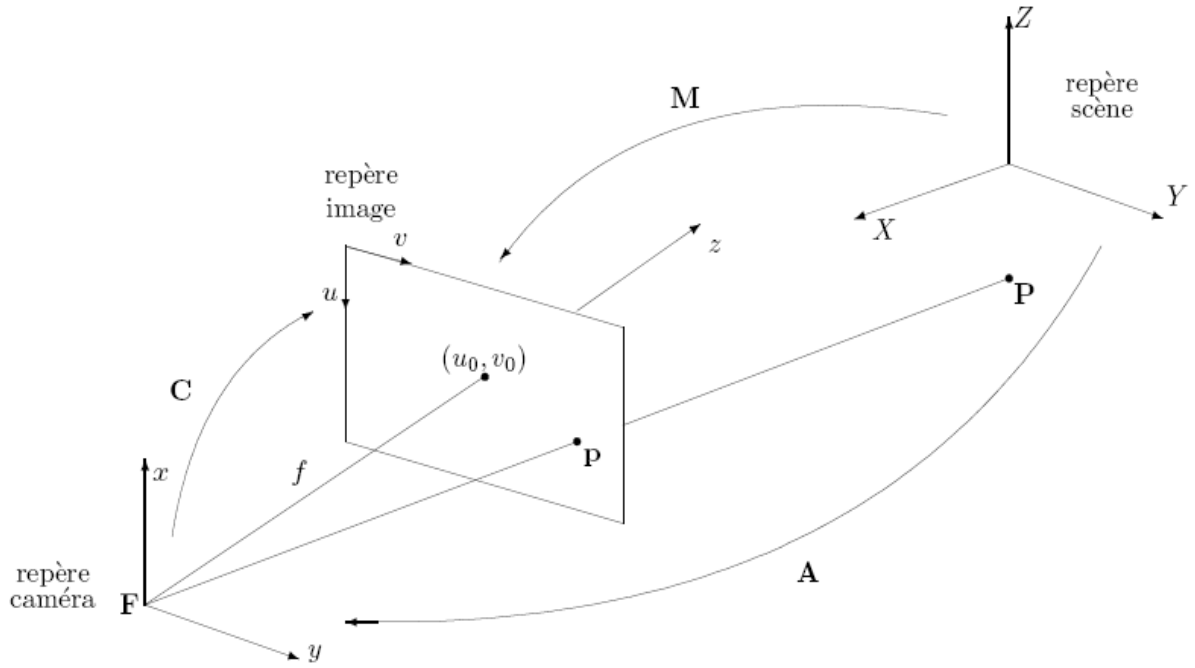


Figure 72 : Modèle géométrique utilisé de la caméra : le modèle sténopé. Le calibrage de la caméra consiste en l'évaluation des paramètres extrinsèques (A) et intrinsèques (C)

La stéréovision consiste à évaluer le relief de la scène en déterminant la relation géométrique qui lie les deux caméras. Cette relation permet de calculer les coordonnées 3D d'un point lorsqu'il est présent dans le champ des deux caméras simultanément (Figure 73). L'ajout des paramètres extrinsèques aux paramètres intrinsèques constitue la matrice fondamentale des deux caméras. Cette matrice est établie par mise en correspondance stéréoscopique de points, soit de manière automatique, soit manuellement.

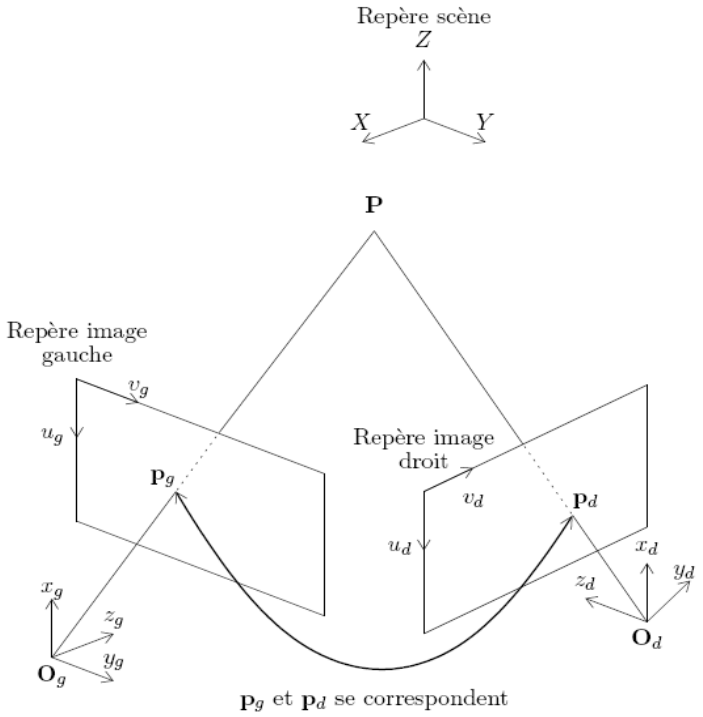


Figure 73 : Modèle géométrique du capteur stéréoscopique : Og et Od sont les centres optiques des deux caméras. On remarque que pour un point P dans la scène réelle, il y a un correspondant Pd et Pg pour chaque caméra. La stéréovision binoculaire consiste en retrouver les coordonnées de P à partir de Pd et Pg.

Cette méthode de calcul de distance est parfaite sur le plan théorique. Elle est cependant tributaire de la qualité du calibrage et de l'évaluation de la matrice fondamentale liant les deux caméras. Elle dépend également de la précision de la mise en correspondance des points entre les deux images. Cette dernière a été effectuée en effectuant la reconnaissance d'objet dans chaque flux vidéo. Pour chaque objet reconnu simultanément dans les 2 images, la position 2D dans chaque image est utilisée pour la mise en correspondance. La précision de cette mise en correspondance est donc directement tributaire de la précision de localisation en 2D de l'objet dans l'image. Notre évaluation a permis d'obtenir des résultats acceptables avec des erreurs d'estimation de la distance comprises entre 4 cm et 20 cm (Figure 74) lorsque la localisation de l'objet par Spikenet était correcte.

Distances réelles :

Bouteille = 80cm

Souris = 80cm

Tasse= 100cm

Distances calculées :

Bouteille = 101cm

Souris = 77cm

Tasse= 96cm

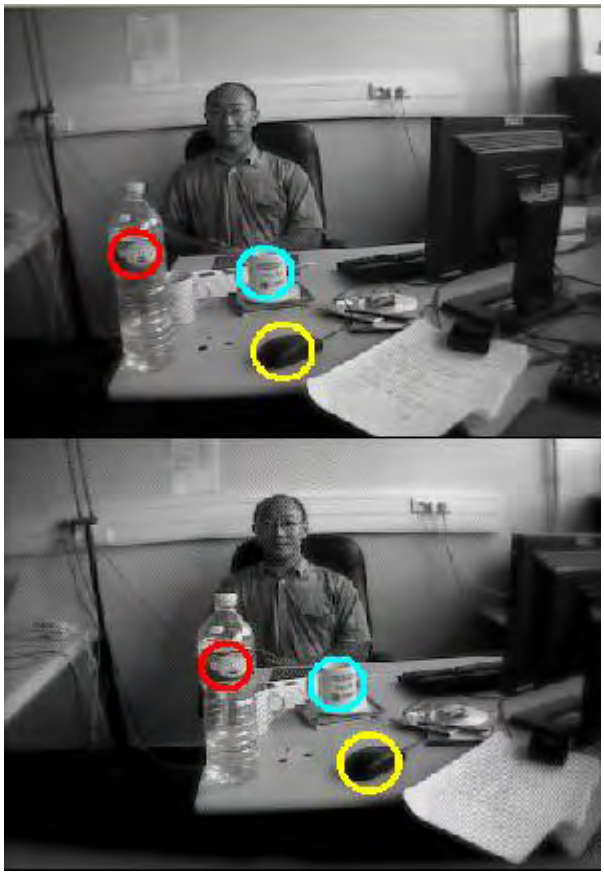


Figure 74 : Évaluation de la distance réelle par rapport à la distance calculée : les erreurs n'excèdent pas 20 cm à un mètre de distance.

Une estimation de la distance plus précise mais plus coûteuse peut être obtenue en utilisant des algorithmes de mise en correspondance à l'échelle du pixel (ex. Détecteur de Harris) dans la zone où un objet a été reconnu par Spikenet. Cette dernière méthode fait intervenir un algorithme d'analyse d'image supplémentaire pour mettre en correspondance des points d'intérêts dans les images. Cette méthode permet de connaître les coordonnées 3D d'un ensemble de points d'intérêts dans l'image au prix d'une étape d'analyse supplémentaire. Ce calcul supplémentaire est toutefois en partie compensé par la suppression d'une instance de Spikenet sur l'un des deux flux vidéo. C'est la méthode que nous avons choisie dans une deuxième version du prototype en choisissant un dispositif de capture stéréoscopique du commerce (BumbleBee, PointGrey Research).

Deuxième version du dispositif de capture

Le choix du capteur de reconstruction 3D s'est porté sur le BumbleBee, produit de la société PointGrey Research. Le système repose sur la reconstruction 3D de la scène visuelle basée

sur la mise en correspondance de tous les pixels dans les images des deux caméras. Les spécifications auxquelles devait répondre le système était : (1) une reconstruction 3D efficace, une résolution minimale de 640x480 et un grand angle de vue, supérieur à 60°. Le BumbleBee que nous avons acheté comporte donc deux caméras SONY en niveaux de gris et une résolution de 640x480 et 48 images par seconde (FPS) pour chacune d'elle (Figure 75). Une lentille de 2,5 mm de focale est placée devant la caméra pour assurer un angle de vue de 100°.

Specification	BB2-03S2	BB2-08S2	BBX3
Image Sensor Type	Sony® 1/3" progressive scan CCD		
	ICX424 (648x488 max pixels) 7.4µm square pixels	ICX204 (1032x776 max pixels) 4.65µm square pixels	ICX445 (1280x960 max pixels) 3.75µm square pixels
Baseline	12 cm		12 cm and 24 cm
Focal Lengths	2.5mm with 97° HFOV (BB2 only) or 3.8mm with 66° HFOV or 6mm with 43° HFOV		
A/D Converter	12-bit analog-to-digital converter		
White Balance	Automatic / Manual (Color model)		Manual (Color model)
Frame Rates	48 FPS	20 FPS	16 FPS
Interfaces	6-pin IEEE-1394a for camera control and video data transmission 4 general-purpose digital input/output (GPIO) pins		2 x 9-pin IEEE-1394b for camera control and video data transmit 4 general-purpose digital input/output (GPIO) pins
Voltage Requirements	8-30V via IEEE-1394 interface or GPIO connector		
Power Consumption	2.5W at 12V		4W at 12V
Gain	Automatic/Manual		
Shutter	Automatic/Manual, 0.01ms to 66.63ms at 15 FPS		
Trigger Modes	DCAM v1.31 Trigger Modes 0, 1, 3, and 14		DCAM v1.31 Trigger Modes 0, 1, 3, and 14
Signal To Noise Ratio	60dB		54dB
Dimensions	157 x 36 x 47.4mm		277 x 37 x 41.8mm
Mass	342 grams		505 grams
Camera Specification	IIDC 1394-based Digital Camera Specification v1.31		
Lens mount	2 x M12 microlens mount		3 x M12 microlens mount
Emissions Compliance	Complies with CE rules and Part 15 Class A of FCC Rules		
Operating Temperature	Commercial grade electronics rated from 0° to 45°C		
Storage Temperature	-30° to 60°C		

Figure 75 : Image du BumbleBee et ses spécifications techniques

Temps de traitement et précision

Les temps de traitement observés sur un ordinateur de bureau équipé d'un processeur Intel Xeon 3GHz et de 1 Go de mémoire vive, sous Windows XP SP2 sont autour de 8 ms (14 FPS) pour le seul calcul des coordonnées 3D en utilisant 20% du microprocesseur.

La précision de localisation en distance dépend directement de la distance à l'objet. Le graphique suivant présente les données de reconstruction pour une résolution de 320* 240, avec un champ de vue de 100° dans le champ proche et le champ lointain. La Figure 76 affiche la précision de la reconstruction 3D théorique donnée par le constructeur pour le capteur que nous utilisons (caméra SONY, 320x240@48FPS, 100° d'angle de vue).

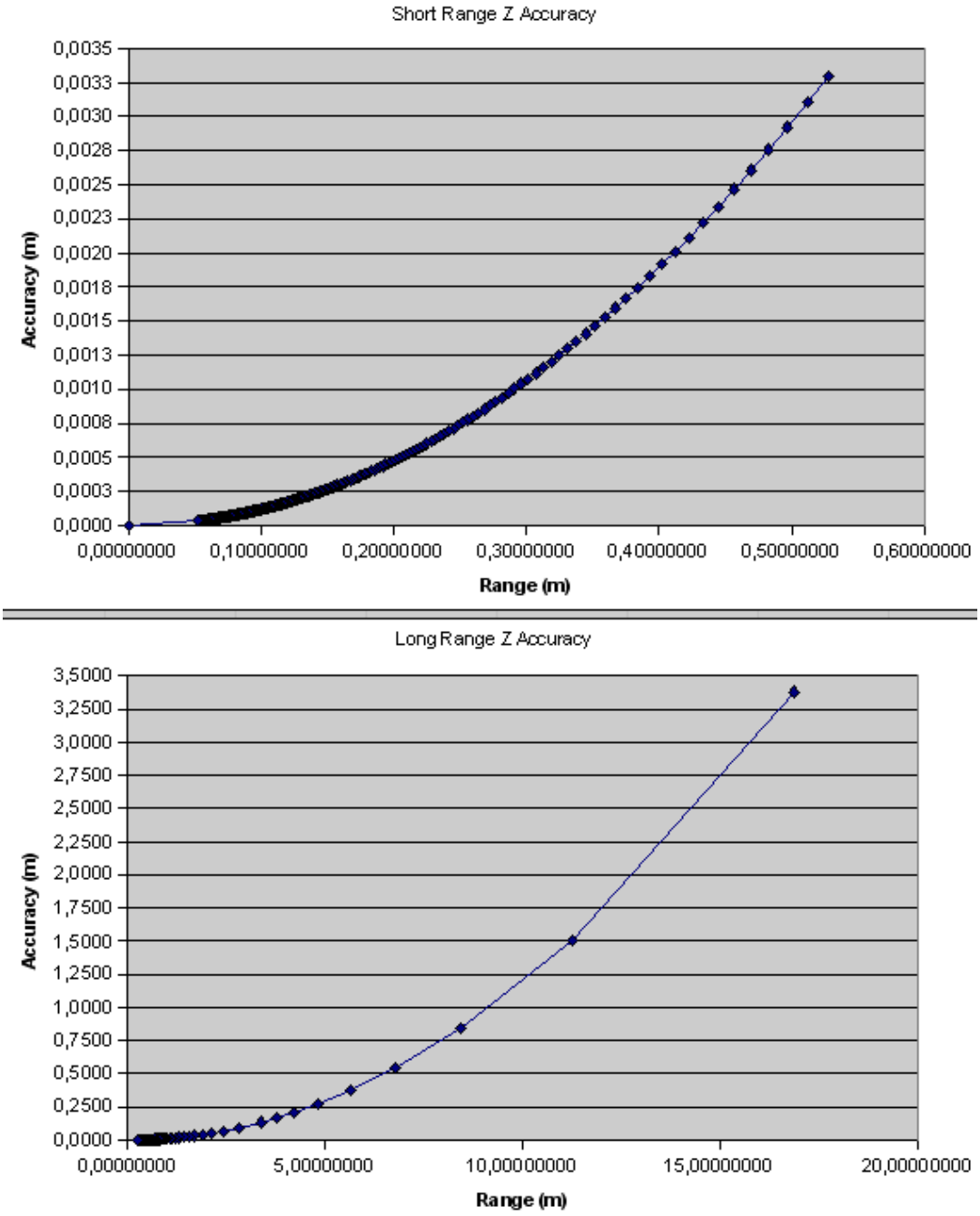


Figure 76: Données constructeur pour une résolution de 320*240, avec un champ de vue de 100° entre 0 et 0,5 m (en haut) et entre 0 et 17 m (en bas).

Le BumbleBee est interfacé avec une bibliothèque pour la programmation en C++ permettant de connaître les coordonnées 3D des pixels des images de chaque caméra. Il est à noter que la mise en correspondance des pixels étant faite en amont par du logiciel du BumbleBee et l'estimation des coordonnées 3D des pixels est effectuée sur la totalité de l'image. A l'usage, le pilote logiciel se base sur des algorithmes de mise en correspondance de points peu robustes quand les images sont à une distance très proche, donc avec un contenu assez différent entre les deux images. En pratique, avec le matériel utilisé, nous ne

pouvons pas calculer les coordonnées d'objets dont la distance est inférieure à 24 cm. Il est bien entendu que cette limitation est due à l'algorithme de mise en correspondance et du capteur utilisé et qu'il n'existe aucune restriction de distance si la mise en correspondance peut être faite (les points doivent juste être présents dans les deux images simultanément).

5) Conclusion

Un premier prototype de capture stéréoscopique a été créé à partir de deux caméras miniatures analogiques montées sur une paire de lunettes. Celles-ci présentaient l'avantage d'être petites, légères et synchronisables. L'utilisation de Spikenet pour effectuer la mise en correspondance de la position des objets reconnue sur chaque caméra s'est avérée peu efficace du fait de deux problèmes :

- Il fallait tout d'abord exécuter une instance de Spikenet pour chaque caméra, ce qui était coûteux en temps de calcul. L
- La précision de localisation 2D dans l'image n'était pas assez bonne pour estimer de manière fiable la distance à l'objet. Un deuxième prototype de capture pour la reconnaissance et la localisation 3D d'objets a donc été développé à partir d'un dispositif matériel et logiciel du commerce.

Ce module utilise la librairie de reconstruction 3D livré avec la caméra et la librairie Spikenet. L'interface graphique a été programmée avec la bibliothèque QT en C++ et permet de suivre en temps réel le fonctionnement du dispositif. Les modèles sont chargés à la main et des messages IVY sont envoyés lorsqu'un objet est reconnu. Les modèles sont de la forme

SN <nom_modèle> <X><Y><Z>

Lorsque les coordonnées 3D ne peuvent pas être identifiées, la troisième coordonnée est rapportée à 0. La Figure 77 montre le fonctionnement du module de vision artificielle. Spikenet ne fonctionne que sur une seule caméra. Quand un objet est reconnu, sa position 2D en pixels dans l'image est calculée. La position 3D correspondante est ensuite établie par le pilote logiciel du BumbleBee et envoyée sur le bus Réseau IVY accompagnée du nom du modèle reconnu à cet emplacement. Le système de vision reçoit aussi des requêtes pour le chargement d'objets par le bus IVY.

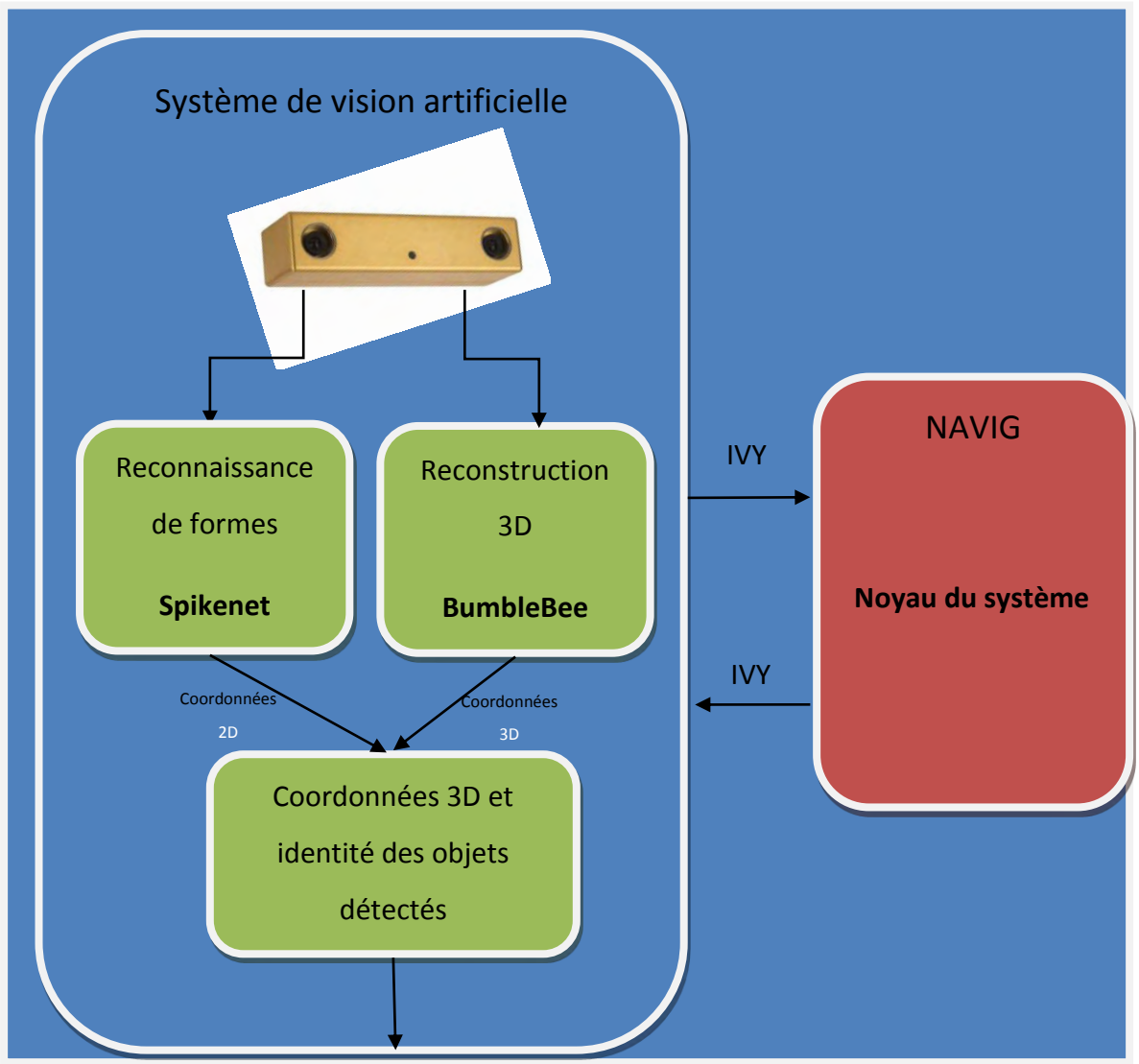


Figure 77: Capteur de vision artificielle du projet Navig : les coordonnées et l'identité des objets reconnus sont envoyées par le bus IVY. Les messages entrants dans le module de vision par le bus IVY correspondent à des requêtes de chargement de modèle.

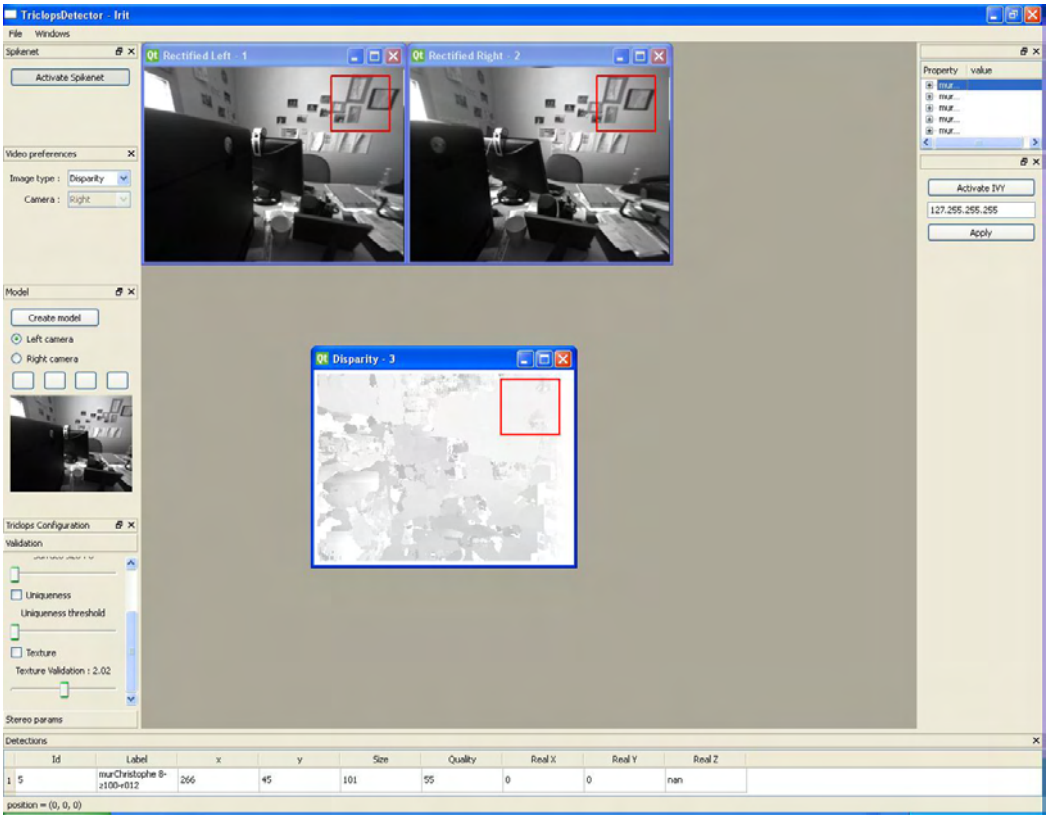


Figure 78: Capture d'écran de l'interface du module de vision. L'image de gauche, l'image de droite ainsi que la carte de disparités sont disponibles, la dernière permettant l'estimation de la distance à l'objet. Le rectangle rouge correspond à une reconnaissance par Spikenet sur l'image de droite.

L'objectif de ce module de vision pour aider les personnes non-voyantes à se représenter l'espace qui les entoure est de restaurer une fonction très utile à la navigation et à la préhension d'objets : la reconnaissance et la localisation d'objets en 3 dimensions. Ce capteur intelligent permet d'effectuer une telle tâche et constitue la brique principale de notre système d'aide à la navigation et à la localisation d'objets. Il diffère des approches habituellement utilisées dans les systèmes de suppléance basés sur la vision artificielle par l'analyse des objets présents dans la scène visuelle. L'ensemble des informations restituées repose sur la qualité de l'analyse de cette scène visuelle et une attention particulière devra être portée sur les réglages fins du système en cas de fausse détection par rapport au coût humain d'une telle erreur.

Restitution par synthèse binaurale

Le choix de l'analyse de la scène visuelle par vision artificielle met à disposition du module de restitution deux informations principales : la position d'un objet et son identité (Figure 79). L'objectif de l'étude présentée dans ce paragraphe est de déterminer et manipuler la précision de la localisation binaurale chez l'homme ainsi que les indices acoustiques qui en sont la base. Finalement cette étude servira à concevoir une interface sonore permettant d'évaluer l'utilisation de sons 3D dans la construction d'une représentation spatiale auditive.

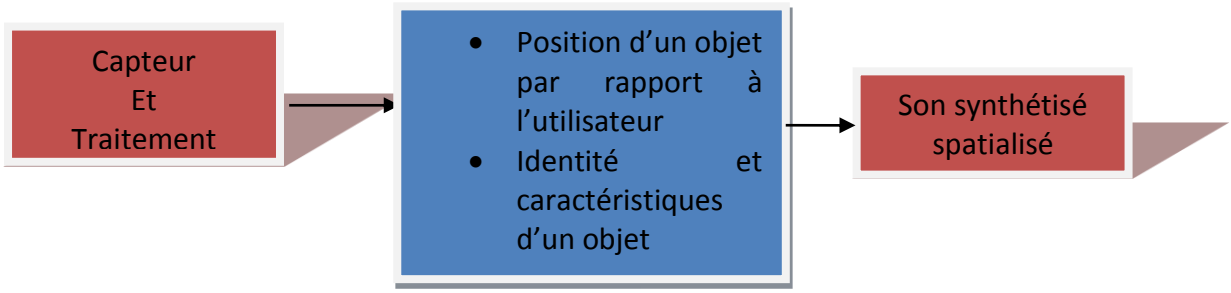


Figure 79: Informations disponibles à l'étape de restitution dans le dispositif de localisation de cibles par vision artificielle

1) Principes de la synthèse binaurale

La synthèse binaurale consiste à générer des sons différents dans chaque oreille afin de produire une sensation sonore unique localisée précisément dans l'espace. Ces sons peuvent être produits par des enceintes disposées devant l'utilisateur ou des écouteurs placés sur ses oreilles. De nombreux jeux vidéo utilisent aujourd'hui cette méthode pour augmenter l'immersion du joueur. Les algorithmes de synthèse embarqués dans ce genre d'applications sont cependant simplistes et se basent le plus souvent sur une différence d'intensité de niveau sonore gauche/droite. Les indices précis de localisation d'une source sonore dans l'espace sont connus et nécessitent un réglage minutieux des filtres de restitution des sons dans les oreilles pour les rendre opérants. Nous allons présenter ici la synthèse binaurale comme une méthode permettant de restituer la position d'un objet réel qui ne serait pas perçu visuellement. La tâche est ici plus critique que pour des jeux par exemple où la position précise de la source a moins d'importance.

Le son qui parvient aux tympans diffère par bien des aspects de ce qu'il est à sa source. D'une part, la distance interaurale entraîne une différence de temps (ITD : Interaural Time Difference) et d'intensité du son (IID : Interaural Intensity Difference) entre les deux oreilles ;

ces différences étant fonction de la position de la source par rapport à la tête de l'auditeur. D'autre part, la réflexion du son par l'oreille externe (le pavillon), la tête, les cheveux et le tronc et son acheminement vers l'oreille interne modifient le spectre du son. Ces différences, liées à la morphologie de l'individu, font qu'il existe une véritable « signature » de chaque auditeur sur le plan de la réception globale du son. Cette signature est unique et propre à chaque individu. On peut la modéliser par une fonction de transfert du signal acoustique : les HRTF (Head Related Transfer Function).

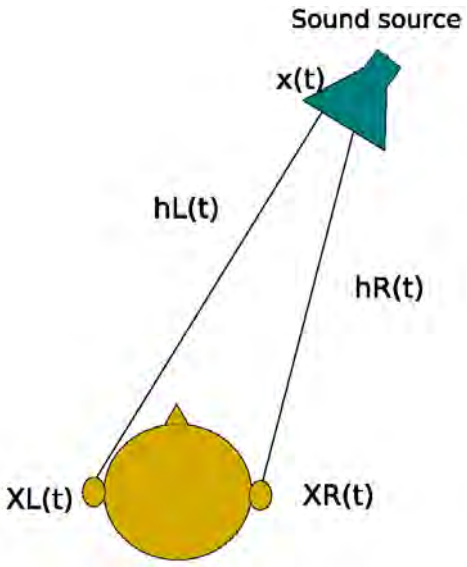


Figure 80 : Une fonction de transfert (HRTF : Head Related Transfer Function) peut être modélisée pour chaque individu. Pour cela, la réponse impulsionnelle à un son x(t) est enregistrée à différentes positions et arrive sur l'oreille de gauche (XL(t)) et sur l'oreille de droite (XR(t)) en suivant les fonctions de transfert hL(t) et hR(t). Trois principales variables caractérisent une fonction de transfert : l'ITD (le son arrive d'abord sur l'oreille droite puis sur l'oreille de gauche), l'IID (le son arrive plus atténué sur l'oreille de gauche par sa propagation dans l'air plus longue et le masquage de la tête) et le spectre. Toutes ces variables sont directement dépendantes de la morphologie du sujet.

La simulation d'une source sonore spatialisée par synthèse binaurale intègre les données de transformation (HRTF) issues de l'enregistrement stéréo du son obtenu en plaçant un microphone dans chacune des deux oreilles d'un auditeur donné et en analysant les modifications entre le son émis et le son enregistré dans les oreilles (Kulkarni and Colburn, 1998a). Il devient alors possible de modéliser un filtre imitant la signature auditive de l'utilisateur.

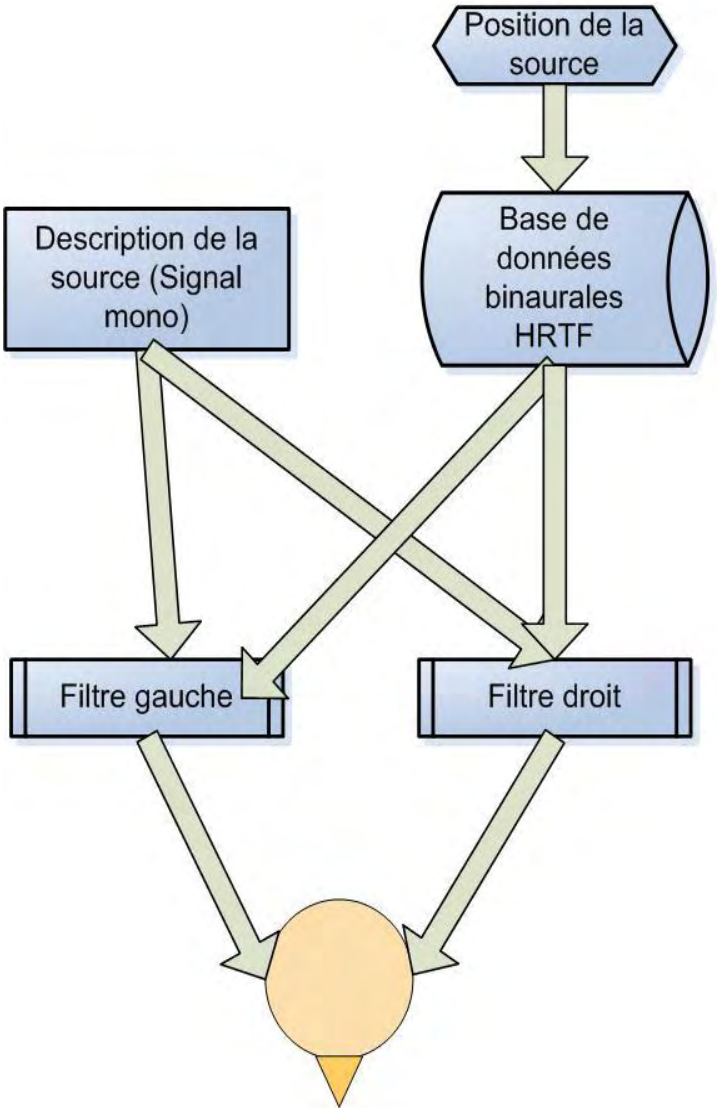


Figure 81 : Principe de la synthèse binaurale

Afin de localiser une source sonore, le système auditif a recours au traitement d'indices binauraux, monauraux et à l'analyse des réverbérations. Les mécanismes cérébraux, impliqués dans l'analyse des réverbérations restent aujourd'hui encore à déterminer. Un son émis dans un environnement réverbérant se réfléchit sur les surfaces situées à proximité de la source sonore. Dans une telle situation, une personne perçoit un ensemble de vibrations acoustiques, qui proviennent soit directement de la source, soit des surfaces de réflexions. Le cortex auditif peut, à partir de ces différentes informations acoustiques, déterminer la localisation de la source sonore, ainsi que certaines informations sur la surface de réverbération.

2) La localisation auditive spatiale chez l'homme et son objectivation

Il est possible de localiser une source sonore en champ lointain (>1 m) chez un sujet voyant (Blauert, 1997) ou non-voyant en élévation et azimut mais la précision en distance est très

faible (Fukuda et al., 2003). Aller saisir des objets proches est théoriquement possible mais des différences dans la localisation de cibles auditives dans le champ proche et dans le champ lointain ont été montrées (Brungart et al., 1999;Farne and Ladavas, 2002). Contrairement au champ lointain, la précision de localisation en champ proche (espace péri-personnel) a très peu été étudiée. Nous souhaitons dans notre système évaluer la faisabilité d'un système de localisation d'objets par synthèse binaurale en champ proche et en champ lointain. La comparaison de la précision de localisation auditive des sujets voyants et non-voyants dans l'espace lointain (>1m) est controversée avec des résultats très différents selon les études (Doucet et al., 2005;Voss et al., 2004). Nous nous limiterons ici à une étude bibliographique de la localisation dans l'espace lointain et évaluerons la précision de localisation humaine dans l'espace proche. Les résultats des sujets voyants et non-voyants seront étudiés et comparés.

Méthodes de mesure

Évaluer de manière objective la perception auditive est une tâche très complexe puisque très difficile à exprimer. Il existe de nombreuses méthodes permettant d'établir une mesure qualitative ou quantitative d'un percept auditif mais elles passent toujours par son expression dans un espace de description (Blauert, 1997).

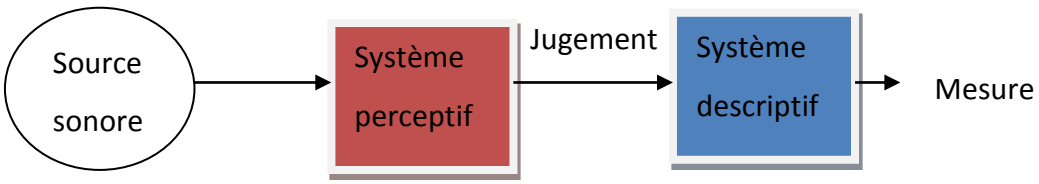


Figure 82 : Mesure de la perception auditive d'un son

La mesure la plus simple consiste en une description du son en une dimension : s'il est présent ou non. Il est évident que le fait de passer par un espace descriptif risque de créer des biais qui ne sont pas représentatifs de la perception réelle du sujet. Il est possible d'utiliser un système de description relatif à un autre percept (égal, différent, plus/moins fort, plus/moins au centre, appartient à tel groupe de sons, …). Le sujet décrit ainsi la source qu'il entend en faisant référence à la nature du son dans le même espace perceptif que le percept de référence. Les méthodes de jugement absolues s'avèrent plus complexes puisqu'il s'agit de décrire la source dans un autre système de mesure. Dans le cas de la localisation de sources, il faut transformer la perception spatiale de la source vers un repère

physique où l'on va généralement pointer. La plupart du temps, le pointage se fait à l'aide d'un faisceau lumineux permettant d'indiquer la position angulaire de la source (azimut et élévation). Cette méthode présente le désavantage de ne pas pouvoir évaluer la distance de la source sonore. Il est alors possible de pointer avec une baguette en bois (Brungart et al., 1999), ou d'aller atteindre la cible avec son doigt (Dramas et al., 2008). Deux notions sont introduites par Blauert (Blauert, 1997). Pour caractériser l'espace auditif humain : la localisation (perception de la position / direction de la source sonore) et le flou de localisation qui correspond à la plus grande valeur de déplacement d'une source sonore pour laquelle l'homme ne perçoit pas de changement de position. Cette dernière valeur caractérise finalement la distribution des réponses intra-sujets dans la perception d'une source sonore.

Flou de localisation

La notion de flou de localisation vient du fait que l'espace n'a pas la même résolution spatiale que l'espace auditif. Comme décrit dans la section précédente, l'objectivation d'une position sonore passe par une transformation de la position dans un espace auditif vers un espace descriptif ou un autre espace sensoriel. C'est en combinant les résultats des différentes méthodes d'objectivation que nous nous approcherons d'un résultat fiable. Le flou de localisation est défini comme le déplacement minimal de la position d'une source sonore qui est reconnue par 50% des sujets de l'expérimentation comme un changement de l'évènement sonore. Cette mesure permet d'établir la plus petite mesure de discrétisation de l'espace auditif pour laquelle l'oreille humaine perçoit une différence de position. Elle représente la résolution maximale de l'espace auditif. Se basant sur 12 études entre 1920 et 1970 avec différents types de stimuli et méthodes d'objectivation en émettant un son devant le sujet suivi d'un son d'un coté ou de l'autre par rapport au premier, Blauert (Blauert, 1997) a montré que la plus petite valeur de déplacement perceptible pour la position droit devant et dans le plan horizontal est de 1°. Le flou angulaire est minimum pour les plus petites valeurs angulaires d'élévation et d'azimut (devant le sujet). Cette notion permet de discrétiser l'espace en établissant l'erreur moyenne de localisation en fonction de la position de la source sonore. La Figure 83 montre le flou de localisation et la perception de localisation en fonction de l'azimut dans le plan horizontal dans deux études sur respectivement 600 et 900 sujets. Les sujets devaient aligner une source sonore avec la position perçue du stimulus. Cette méthode d'objectivation est particulièrement adaptée à l'expression d'une position derrière le sujet, tête fixe. Avec cette méthode, le flou de

localisation obtenu dans les deux études présentées ci-dessous est de +-3,6° devant le sujet et augmente avec l'azimut à gauche et à droite (respectivement +-9,2° et +-10° à 90°). Sa valeur redescend ensuite jusqu'à 5,5° à +-180° d'azimut (derrière).

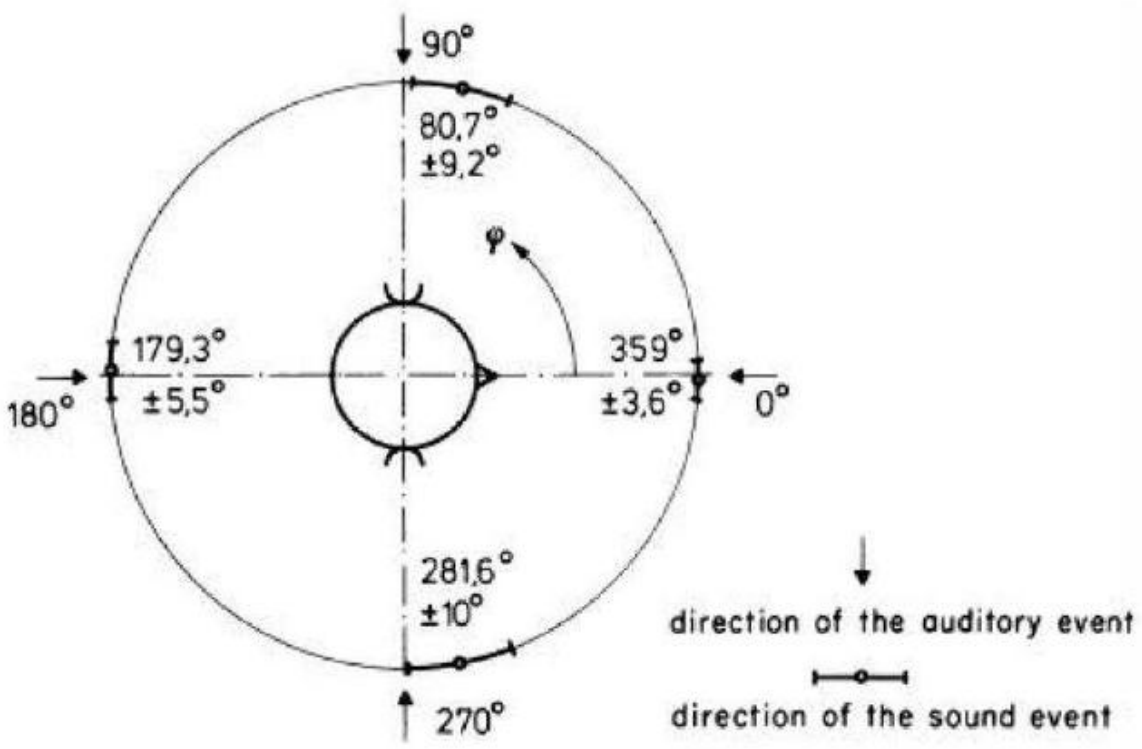


Figure 83: Précision de la localisation d'une source sonore dans le plan horizontal et flou de localisation. (Tiré de Blauert 1997).

Le flou de localisation a aussi été rapporté par Blauert dans son livre décrivant deux études effectuées dans le plan vertical sur la localisation et le flou de localisation en élévation.

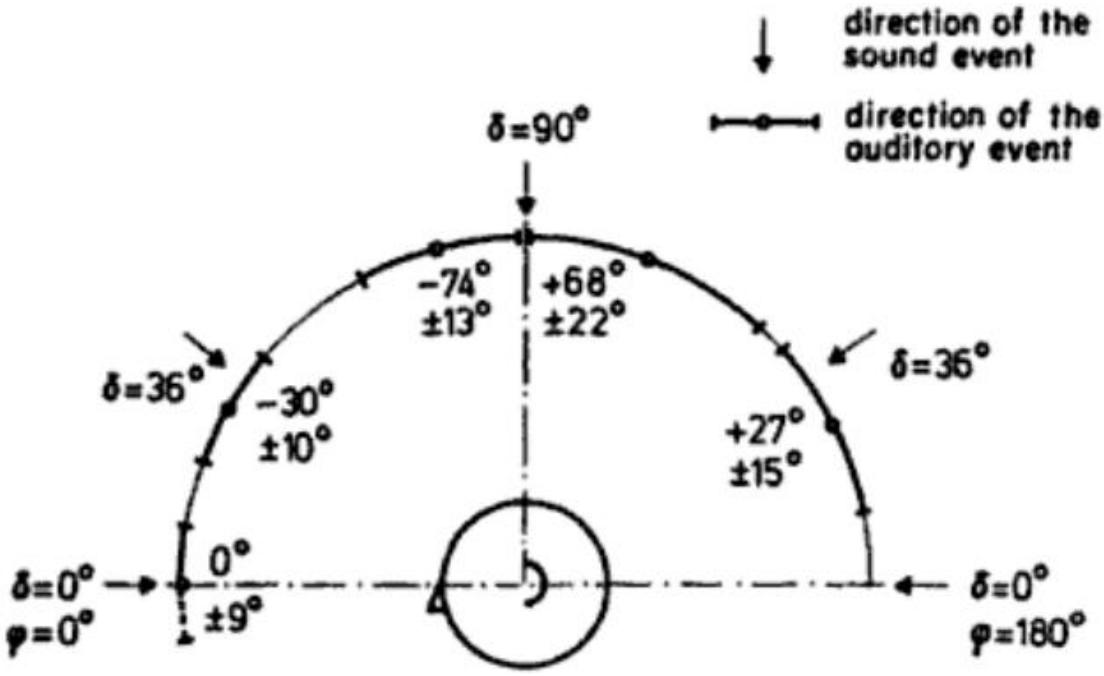


Figure 84: Localisation et flou de localisation de la parole continue d'une personne familière dans le plan médian.

La Figure 84 montre dans un cas particulier reporté par Blauert, de localisation de la parole continue dans le plan médian à partir de données provenant d'une étude menée par Damaske et Wagener en 1969 auprès de 7 sujets, tête immobilisée. Le flou de localisation est de 9° pour une élévation de 0° puis augmente jusqu'à atteindre 22° au dessus de la personne (élévation de 90°). Le flou de localisation est donc un une erreur inhérente au système de perception auditif. L'espace auditif n'est pas échantillonné de la même façon dans tout l'espace.

Précision de localisation

Il a été montré que les capacités des humains à localiser une source sonore dans l'espace étaient bonnes, aussi bien chez les voyants (Middlebrooks and Green, 1991) que chez les non-voyants (Lessard et al., 1998). Selon certaines études, les non-voyants auraient des facilités (Doucet et al., 2005;Voss et al., 2004) à utiliser le spectre sonore qui se réverbère sur l'objet pour déceler un objet, sa distance, sa taille, ainsi que des variations de revêtement (Hausfeld et al., 1982;Rice and Feinstein, 1965). La précision de localisation d'un son dans l'espace extra-personnel (>1 m) a été largement étudiée, Les propriétés des fonctions de transfert dans l'espace lointain sont très différentes pour l'espace proche (Brungart and Rabinowitz, 1999). Dans ce cas particulier, les variations de distance sont

proportionnellement plus importantes entre les deux oreilles que dans l'espace extra-personnel. La morphologie de l'individu joue un rôle accru en champ proche du fait des interactions plus marquées entre le son et les structures anatomiques.

Les indices de perception spatiale auditive chez l'homme ont été étudiés en condition monaurale (Brungart, 1999) ou binaurale (Blauert, 1982;Blauert, 1997;Brungart et al., 2003;Fukuda et al., 2003;Middlebrooks and Green, 1991). Comme attendu, en condition monaurale le spectre du son semble avoir un effet plus important sur la localisation du son qu'en condition bilatérale. La perception spatiale binaurale permet en plus d'utiliser les différences du son arrivant à chaque oreille pour localiser la cible auditive. Ces indices binauraux sont présentés dans le Tableau 3.

Tableau 3: Incidence des principaux indices de localisation binauraux sur la perception de l'espace.

	Azimut	Devant / Derrière	Élévation	Distances	Taille de la pièce	Objets proches
IID	X			X		
ITD	X					
Réflexions proches					X	X
Réverbération globale du son				X	X	
Spectre	X	X	X	X		

Les résultats sur la perception d'une source sonore (Figure 83 et la Figure 84) montrent que la précision de localisation diminue de la même façon que le flou auditif augmente avec la position angulaire de la source sonore. Les sujets sous-estiment les positions angulaires avec une erreur moyenne de 10° pour une source sonore à 90° d'azimut et de 1° devant le sujet. Le même phénomène peut être observé en élévation où l'erreur moyenne devant le sujet est de 0° et de 26° au dessus du sujet. Ces résultats montrent que la précision angulaire maximale est atteinte pour des sources placées droit devant et se dégrade avec l'angle.

Pointage dans l'espace péri-personnel

La localisation d'une source sonore (azimut, élévation, distance) est calibrée par la boucle sensori-motrice (Blum et al., 2004). Plusieurs publications (Brungart et al., 1999;Brungart et

al., 2003;Brungart and Rabinowitz, 1999;Doucet et al., 2005) montrent que les humains sont capables de localiser des sources sonores dans l'espace proche. Brungart (Brungart et al., 1999) a notamment montré que la précision de localisation dépend fortement de la nature du son à localiser. Dans cette étude, Brungart compare la localisation de cinq sources sonores avec des contenus spectraux différents. Dans une première condition, un filtre passe-haut appliqué sur la source diminuait la précision de sa localisation en azimut, en élévation et en distance. Dans une autre condition, un filtre passe-bas diminuait la précision de localisation angulaire (azimut et élévation) et faisait augmenter le nombre d'erreurs de localisation devant/derrière. Ces résultats montrent qu'une source proche peut être localisée avec précision seulement si elle couvre une large bande de fréquences audible (basses et hautes fréquences).

Localisation sonore chez les non-voyants

De nombreux travaux ont mis en évidence des relations entre différentes modalités sensorielles, notamment entre le système visuel et le système auditif, grâce à l'étude de certaines illusions perceptives. Par exemple, dans un environnement bruyant, nous pouvons aisément suivre le discours d'une personne, si nous regardons les mouvements de ses lèvres. Au cinéma on a l'impression que les voix viennent de la bouche de l'acteur alors que les enceintes sont sur le coté : le système visuel modifie donc la perception auditive de la position de la source sonore. La relation forte qui existe entre ces deux sens peut laisser penser que leur développement est intimement lié.

Les capacités de localisation de sources sonores seront donc probablement différentes en fonction de l'âge de la perte de vision, tardif ou précoce. Les travaux de thèse d'Olivier Deprés (Deprés, 2006) montrent l'incidence de l'expérience visuelle sur les capacités d'un sujet à se représenter spatialement son environnement ou à situer une source sonore dans l'espace. Les premières données révélaient que le cortex visuel des aveugles de naissance ne présente aucune caractéristique structurelle particulière, mais que son activité métabolique est supérieure à celles de sujets voyants fermant les yeux (Wanet-Defalque et al., 1988). D'autres travaux mesurant l'activité du cerveau au cours de tâches de discrimination tactile ont montré que les aires visuelles sont impliquées dans le traitement de signaux non-visuels chez les aveugles de naissance. Ces résultats sont confirmés par Sadato et ses collègues (Sadato *et al.*, 1996) qui ont observé que lors d'une tâche de lecture en braille, l'activité électrique corticale est distribuée plus postérieurement chez des sujets non-voyants de

naissance que chez des sujets voyant ayant appris le braille. Cette même distribution d'activités est constatée lorsqu'il est demandé aux sujets d'imaginer les sensations tactiles produites par des textures (Uhl *et al.*, 1994) ou lors d'une tâche de rotation mentale de formes présentées dans la modalité tactile (Roder *et al.*, 1997). Ces résultats confirment que le cortex visuel n'est pas endommagé, même chez les non-voyants de naissance et qu'il continue de fonctionner tout au long de la vie chez les non-voyants. Il est donc possible de stimuler ce cortex « visuel » par le biais d'autres modalités sensorielles (tactiles, auditives…). Ces résultats montrent à la fois les relations très étroites entre la vision et ces autres modalités sensorielles ainsi que l'importance de la plasticité corticale. Il est donc nécessaire de connaître les expériences visuelles passées du sujet pour répondre à son besoin.

Les données sur les différences inter-population entre des voyants et des non-voyants sont controversées. Les comparaisons entre voyants et non-voyants en termes de précision de localisation des sons ne sont à l'heure actuelle pas concluante. En effet, des études montrent que la précision de localisation d'une source sonore est aussi bonne chez les sujets voyants que chez les sujets non-voyants (Lessard et al., 1998;Middlebrooks and Green, 1991) alors que certaines montrent des performances meilleurs chez les sujets non-voyants (Voss et al., 2004). Cette faculté pourrait s'expliquer par une interprétation plus fine du contenu spectral du son arrivant à chaque oreille (Doucet et al., 2005). L'analyse du spectre du son joue un rôle déterminant dans l'estimation de l'élévation, de la distance et pour désambigüiser les erreurs avant-arrière. Les indices spectraux étant plus marqués dans l'espace proche que dans l'espace lointain, la précision de localisation de sources proches de l'utilisateur serait ainsi un bon indice des différences d'utilisation du spectre du son pour la localisation de sources chez ces deux populations.

3) Étude expérimentale du pointage vers des cibles sonores situées dans l'espace péripersonnel

Aller saisir un objet visuel dans l'espace proche de soi est une tâche que les personnes voyantes effectuent très souvent. Pour restaurer cette faculté aux personnes non-voyantes, nous voulons émettre un son dans l'espace proche comme si celui-ci venait de l'objet. Nous voulons ainsi savoir si le fait de substituer la localisation visuelle par une localisation auditive permet de restaurer la faculté d'aller saisir des objets. Pour cela, deux expérimentations ont été menées dans l'espace proche pour évaluer la précision de localisation d'une source sonore auprès de sujets voyants et non-voyants.

Matériels et méthodes

Un dispositif expérimental a été conçu pour tester la précision de localisation dans le plan horizontal dans l'espace péri-personnel (<1 m). Ce dispositif (Figure 89) comporte 35 enceintes (réf. :CB990,8O hm, 3 W) disposées sur un demi-disque de 80 cm de rayon, sous une grille acoustiquement transparente à hauteur de table. Un système de couettes acoustiques a été créé tout autour du dispositif expérimental permettant de réduire le temps de réverbération de 500 ms à 350 ms. La salle acoustiquement traitée avait ainsi une surface de 2 m*2 m et 2 m d'élévation. Le bruit ambiant pendant l'expérimentation était principalement dû aux appareils présents dans la salle et était de 37,5dB.

Enregistrement de la position du doigt

Le sujet devait positionner son doigt au centre du demi-disque (plateau) et attendre de percevoir un stimulus provenant d'une enceinte. Il devait alors toucher le grillage en surface du plateau avec l'index de la main droite, à l'endroit d'où le son semblait provenir puis revenir à la position initiale pour qu'un nouveau stimulus soit présenté. Les sujets n'avaient aucun retour d'information lorsqu'ils allaient pointer et ne pouvaient pas corriger leurs erreurs au fur et à mesure de l'expérience. La position du doigt était suivie en temps réel à l'aide d'un outil de suivi optique : l'outil COFT (Cheap Optic Finger Tracker). Cet outil a été conçu pour l'expérimentation et comporte une caméra SONY EVI D70 d'une résolution de 768 x 494 pixels disposée au dessus du plateau et un ordinateur cadencé à 3 GHz.

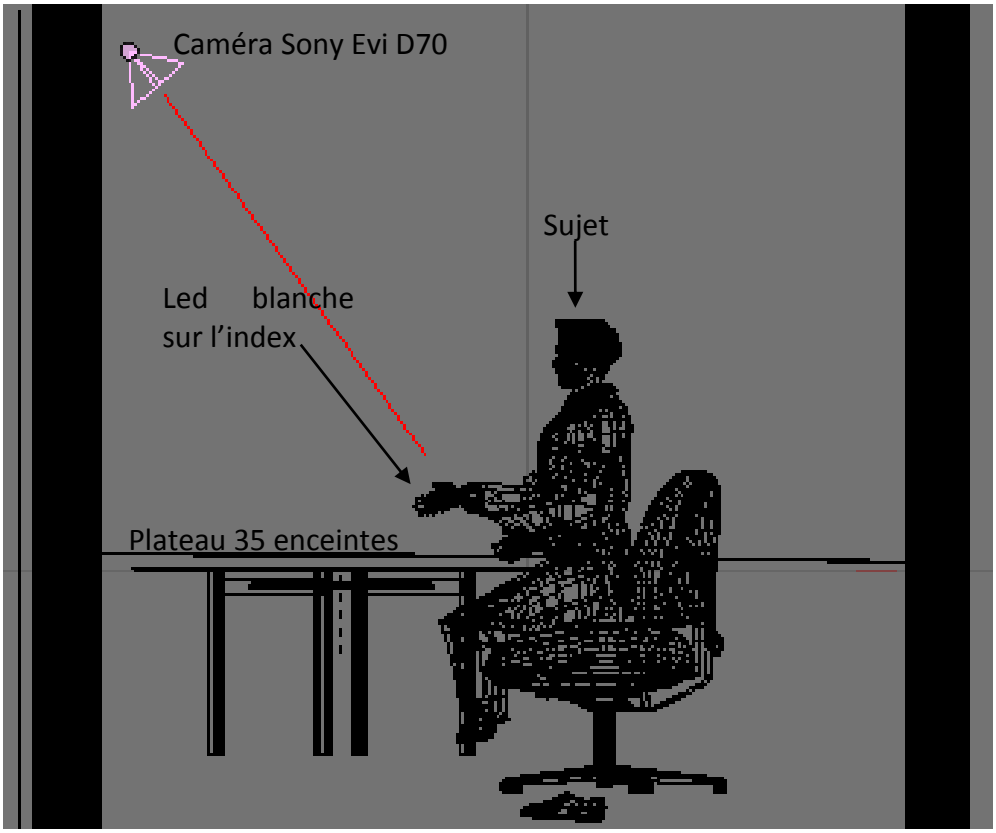


Figure 85: Vue de profil du dispositif utilisé dans une expérience de localisation des sons. Le sujet est assis devant une table sur laquelle est disposé un plateau munis d'enceintes. Une caméra (Sony Evi D70) filme l'expérimentation.

L'utilisateur porte une petite diode électroluminescente accrochée sur le dessus de l'index. Cette diode est suivie en temps réel par un traitement d'image en 5 phases (Figure 86) :

1) L'image est seuillée puis binarisée pour ne garder que les zones blanches de l'image, tout le reste est alors noir. La valeur de seuil est paramétrable en fonction de la luminosité de la pièce.

2) Une Erosion d'une valeur x, paramétrable, est effectuée sur l'image pour éliminer les petites zones blanches (reflets) : seule la zone blanche ayant la plus grande aire est conservée.

3) Une dilatation de la même valeur x est alors effectuée pour retrouver la taille initiale de la zone d'intérêt ayant été érodée.

4) Un filtre de Canny est appliqué sur l'image pour délimiter le contour de la zone d'intérêt qui est à présent unique. L'algorithme renvoie une liste de coordonnées de pixels définissant le contour de la zone.

5) Le centre de localisation est alors calculé en faisant la moyenne des coordonnées de l'ensemble de ces points de contour issus du filtre de Canny.

Une phase de calibrage est effectuée en début d'expérimentation. Pour cela, une diode électroluminescente est positionnée sur la projection de chaque enceinte sur le grillage pour définir la position de pointage correcte pour chaque enceinte. Un fichier est ainsi enregistré comportant l'ensemble des positions de pointage pour toutes les enceintes en pixels dans l'image. Les coordonnées réelles métriques ont été enregistrées avec une précision supérieure au millimètre (0,1 mm d'erreur RMS, données constructeur). L'équivalence entre les données réelles mesurées en millimètre et les données en pixel interviennent dans l'établissement de la matrice d'homographie (matrice de conversion des données du repère image vers le repère métrique en tenant compte des distorsions éventuelles de l'image et de l'angle de la caméra avec le plan horizontal). Dans notre cas, la matrice d'homographie et la conversion des données dans le repère métrique sont effectuées au moment de l'analyse de données sous Matlab.

Soit Ai le vecteur des coordonnées de l'enceinte i dans le repère image, Bi le vecteur des coordonnées de l'enceinte i dans le repère métrique. L'objectif est de trouver une matrice H telle que quel que soit i, Ai*H=Bi. Nous optimisons les paramètres de la matrice d'homographie en minimisant la somme des distances au carré. L'erreur de positionnement est d'au maximum 2 cm. La position réelle du doigt dans le repère métrique n'est valide que lorsque le doigt touche le grillage (sur le plan horizontal du grillage).

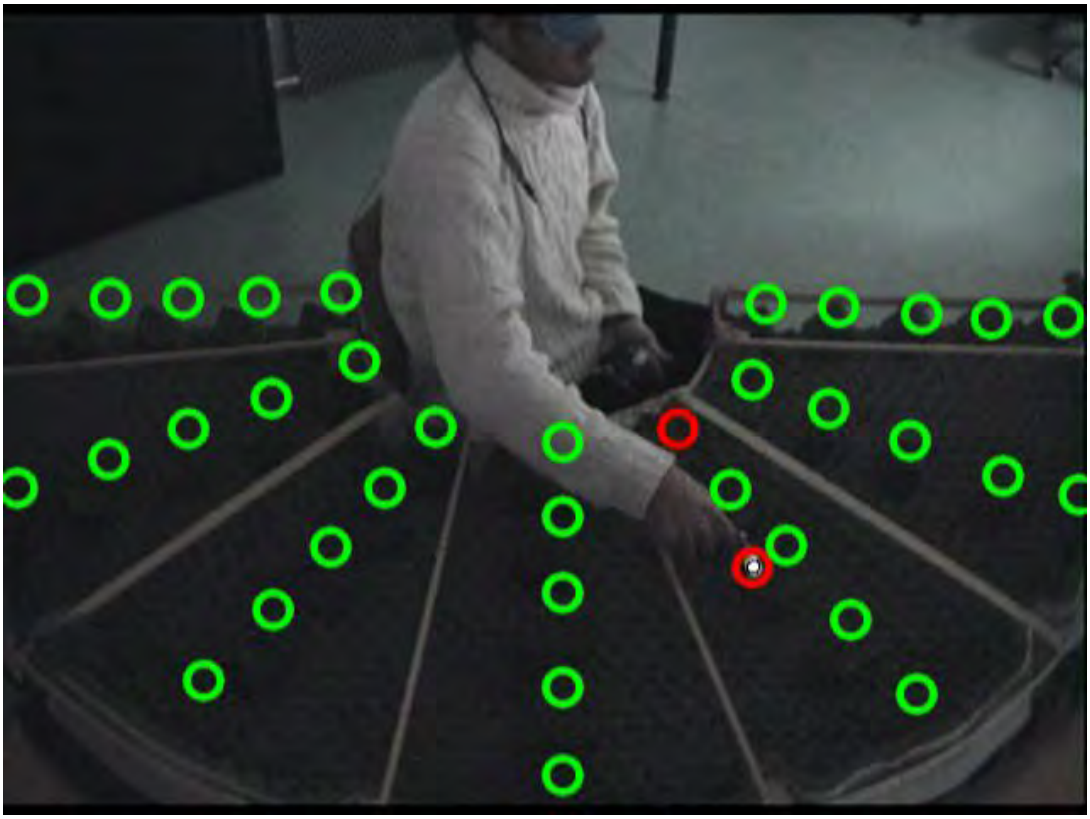


Figure 86: Utilisation de l'outil COFT sur une vue de la caméra utilisée. Les cercles verts correspondent à la position de pointage correcte de chaque enceinte, dans le repère image. Le cercle rouge entourant la lumière provenant de l'index de l'utilisateur suit le doigt. La position du doigt est enregistrée en temps réel. Le deuxième cercle rouge correspond à l'enceinte de laquelle provient le stimulus.

Extraction des données de pointage

L'ensemble de la trace du doigt dans le repère image est enregistrée à une fréquence minimale de 10 Hz. La position d'atteinte de la cible est établie en prenant le point de vitesse nulle de la trace. Un algorithme de détection automatique permet d'établir ce point qui est systématiquement validé manuellement par l'expérimentateur (Figure 87).

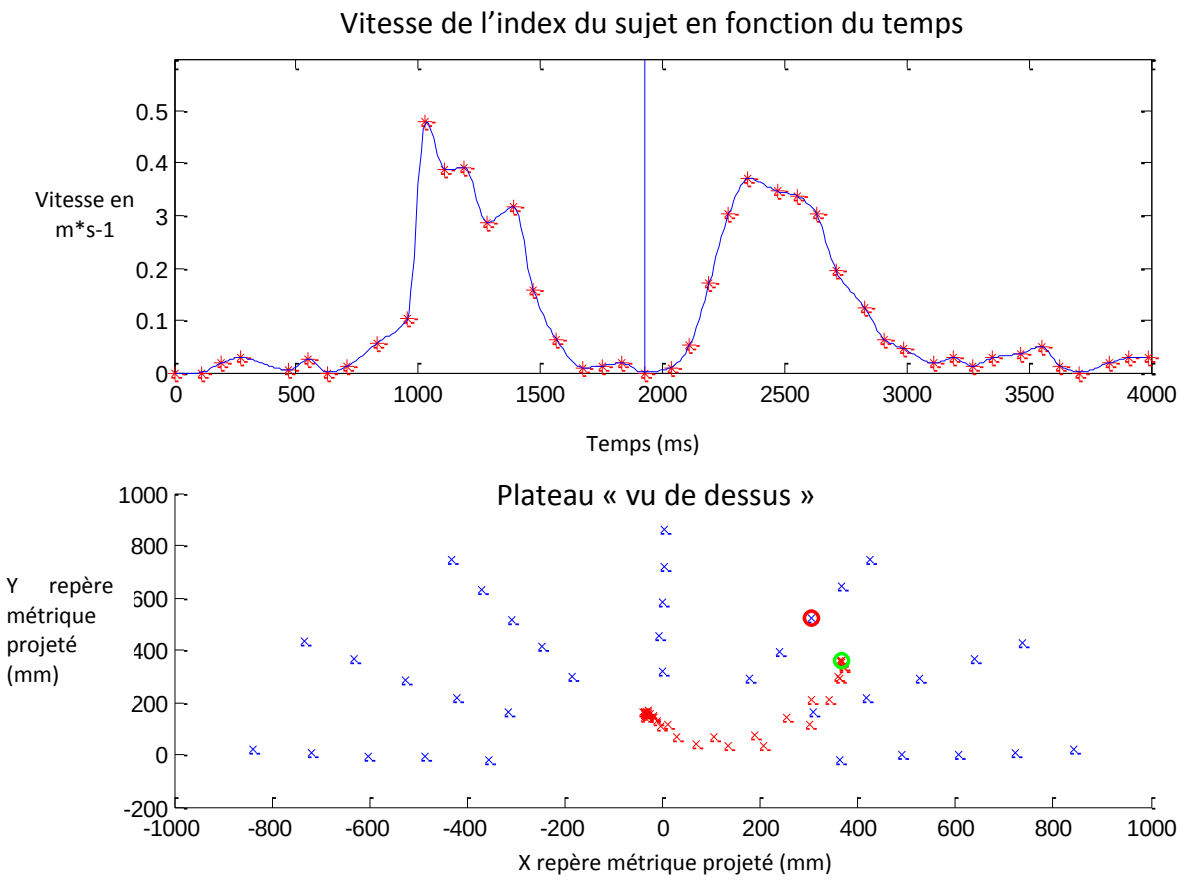


Figure 87: Extraction de la position de pointage de l'utilisateur pour un essai d'une durée de 4 secondes. Le graphique du haut affiche la vitesse du doigt de l'utilisateur en fonction du temps. Le graphique du bas affiche la position des enceintes dans le repère métrique (croix bleues), la trace (croix rouges) dans le repère métrique du doigt, projeté sur le plan horizontal du grillage (vue de haut du dispositif). La position en deux dimensions n'est alors valable que lorsque l'utilisateur touche le grillage, ce qui correspond également au moment où la vitesse de son doigt est nulle. La valeur de temps pour laquelle le doigt a une vitesse nulle est affichée par une ligne verticale sur le graphe du haut et sa position par une croix verte sur le repère du bas. L'enceinte d'où provenait le stimulus est entourée d'un cercle rouge.

Données étudiées

Une fois l'ensemble des coordonnées 3D des points atteints extraites, des statistiques sont effectuées sur les précisions moyennes en azimut et en distance atteintes en fonction des sujets, s'ils sont voyants ou non-voyants, des conditions acoustiques et des stimuli testés.

La précision moyenne en azimut, en distance et en distance absolue (distance métrique entre l'enceinte et le point atteint) est étudiée suivant les 5 variables évoquées ci-dessus par une ANOVA en utilisant le logiciel de statistiques Statistica. La précision de pointage est évaluée selon 5 valeurs :

- L’erreur de distance est la valeur absolue de la différence entre la distance de l’endroit pointé et la distance de l’enceinte.
- L’erreur d’azimut est la valeur absolue de l’angle formé entre l’endroit pointé et l’enceinte.
- L’erreur de distance absolue est la distance euclidienne entre l’endroit pointé et l’enceinte.
- Le nombre d’erreurs « Avant-Arrière » (front-back) qui correspond aux erreurs de localisation symétriques par rapport à l’axe interaural. Une erreur avant-arrière n’est possible que pour les enceintes se trouvant hors l’axe interaural (il n’y aura donc pas d’erreur à 90°). Ne sont détectées comme erreur avant-arrière que les pointages qui sont au-delà de 100° alors que la position angulaire de la source sonore est inférieure à 80° ou inversement.

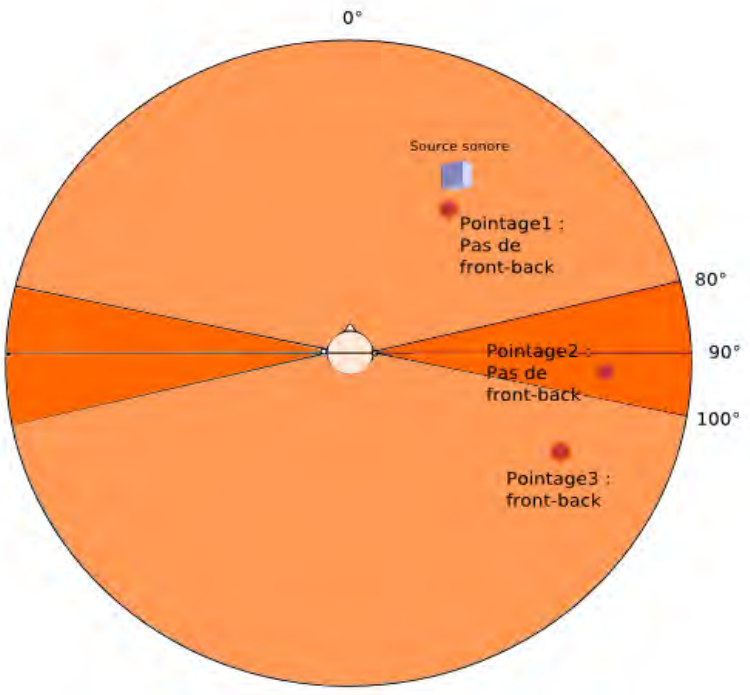


Figure 88: Une erreur avant-arrière (front-back) est définie par une erreur de localisation symétrique par rapport à l’axe interaural. L’algorithme de détection des erreurs avant-arrière détecte ces erreurs quand la source sonore a un azimut inférieur à 80° ou un azimut supérieur à 100° et que la zone de pointage a respectivement un azimut supérieur à 100° / inférieur à 80°.

Expérience 1 : Étude préliminaire sur les mouvements auditivement guidés dans l'espace proche

Une première étude chez des sujets voyants a été menée pour étudier la précision de localisation d'une source sonore suivant trois variables : 1/ Les stimuli : des bruits gaussiens dont la durée et le nombre varient, 2/ la présence ou non d'un équipement diminuant les réverbérations du son, et 3/ la position de la source dans le plan (azimut et distance).

Protocole expérimental

10 sujets voyants âgés de 20 à 60 ans ayant les yeux bandés ont participé à l'expérimentation. Un audiogramme a été établi pour vérifier qu'aucun d'entre eux ne présentait de déficit auditif supérieur à 20 dB pour des fréquences entre 125 et 4000 Hz. Les stimuli étaient dérivés de bruits gaussiens avec 4 conditions :

1. 200 ms * 1
2. 40 ms * 1
3. 40 ms * 3 avec 30 ms de pause entre chaque répétition
4. 40 ms * 8 avec 30 ms de pause entre chaque répétition

Seuls les 20 haut-parleurs de l'hémisphère droit ont été utilisés pour diminuer le nombre de stimuli et la durée de l'expérience qui était déjà d'environ une heure. Nous avons supposé pour cette expérimentation que la précision de localisation serait la même à droite ou à gauche. Des couettes acoustiques ont été installées pour atténuer la réverbération du son sur les murs de la pièce, les stimuli ont été testés avec et sans les couettes acoustiques (conditions 1b, 2b, 3b, 4b).

Les sujets étaient assis au centre de l'hémi-disque, la main appuyée sur un bouton juste devant eux. Lorsqu'un bref stimulus auditif (conditions 1, 2, 3 ou 4) était présenté sur un haut-parleur, le sujet devait aller atteindre la position estimée du son avec son index de la main droite. La position du doigt était enregistrée par l'outil de suivi optique COFT.

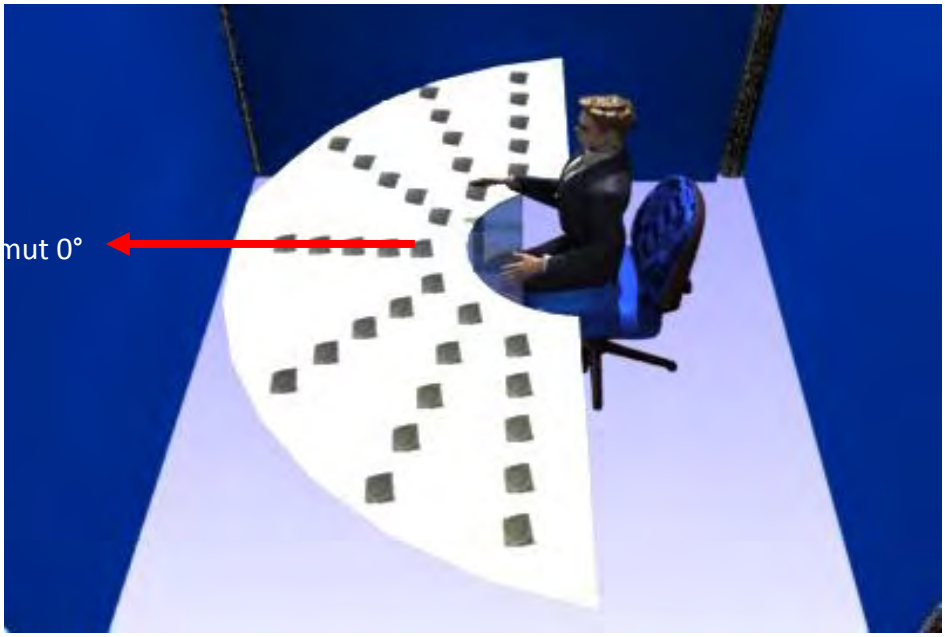


Figure 89: Dispositif expérimental comportant 35 enceintes. Dans cette première expérimentation, l'utilisateur est positionné au centre de l'hémi-disque. 15 haut-parleurs sont positionnés à sa droite (azimuts de +30 à +90°, distance de 20 à 80 cm), 15 à sa gauche (azimuts de -30° à -90°, distance de 20 à 80 cm) et 5 devant lui entre 20 et 80 cm de distance.

Résultats

Précision de pointage en fonction de la position de la source

L'erreur de précision en distance est définie par la valeur absolue de la différence entre la distance de l'utilisateur à la source et la distance de l'utilisateur à la position pointée. Nous avons étudié cette mesure suivant 5 valeurs de distances (380 mm, 500 mm, 620 mm, 740 mm, 860 mm) et 4 valeurs d'azimut (0°, 30°, 60°, 90°). La Figure 90 montre que l'erreur de distance diminue quand la source se rapproche de l'utilisateur jusqu'à 38 cm, quel que soit l'azimut de la source (d'une erreur de 250 +/- 5 mm à 860 mm de distance à une erreur de 99 +/- 5 mm à 380 mm de distance, en face du sujet). De plus, à distance égale, l'erreur en distance de pointage diminue quand la source se déplace sur le coté du sujet (à 86 cm, l'erreur de distance pointée est de 250 + 5 mm en face du sujet et 145 + 5 mm sur le coté du sujet).

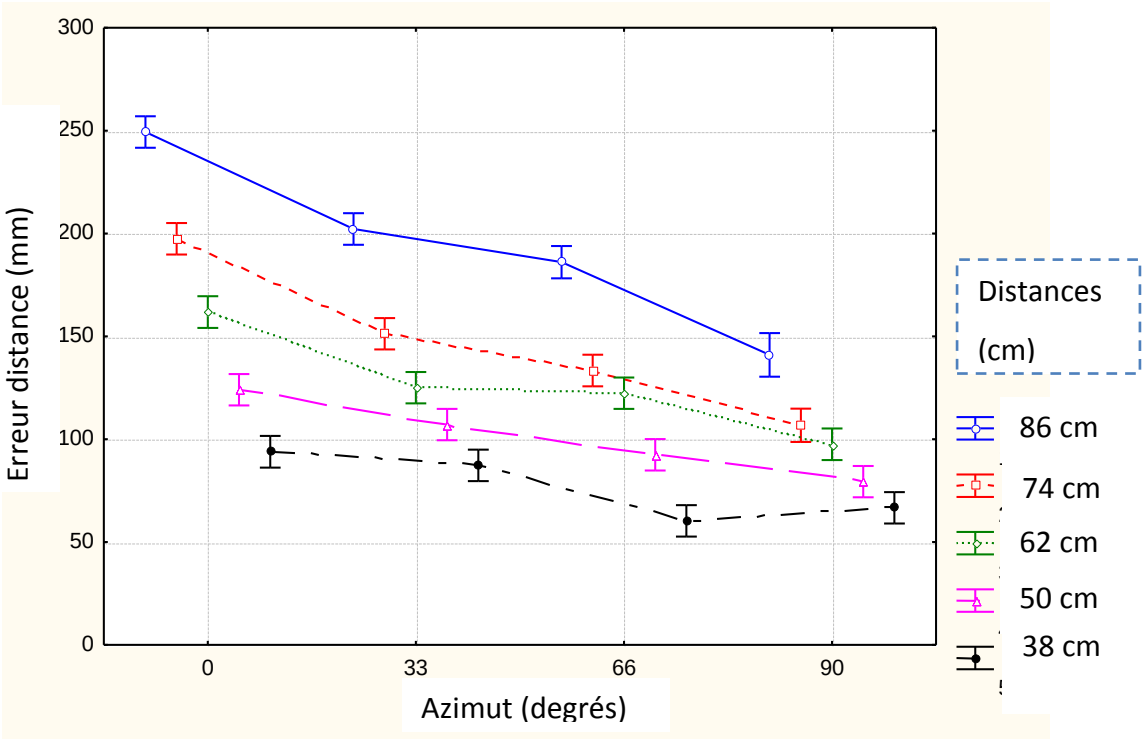


Figure 90: Erreur en distance en fonction de la distance et de l'azimut de la source. L'erreur est directement dépendante de l'azimut de la source. L'erreur moyenne d'atteinte de la cible sur les 10 sujets est moins importante en face du sujet (90 + 5 mm) que sur le coté (153 + 5 mm). (F(12,7255)=11,277 et p=0,0000)

La Figure 91 montre l'erreur d'azimut pointé en fonction de la distance et de l'azimut de la source auditive. On observe, quelle que soit la distance de la source, une courbe en forme de cloche : l'erreur d'azimut est de 4 +/- 0,3° en face du sujet et augmente jusqu'à l'azimut 66° (erreur de 10 +/- 0,3°) puis diminue jusqu'à la précision initiale pour 90° d'azimut. Ce résultat n'est pas cohérent avec la littérature et une explication sera proposée ci-dessous pour des cibles latérales.

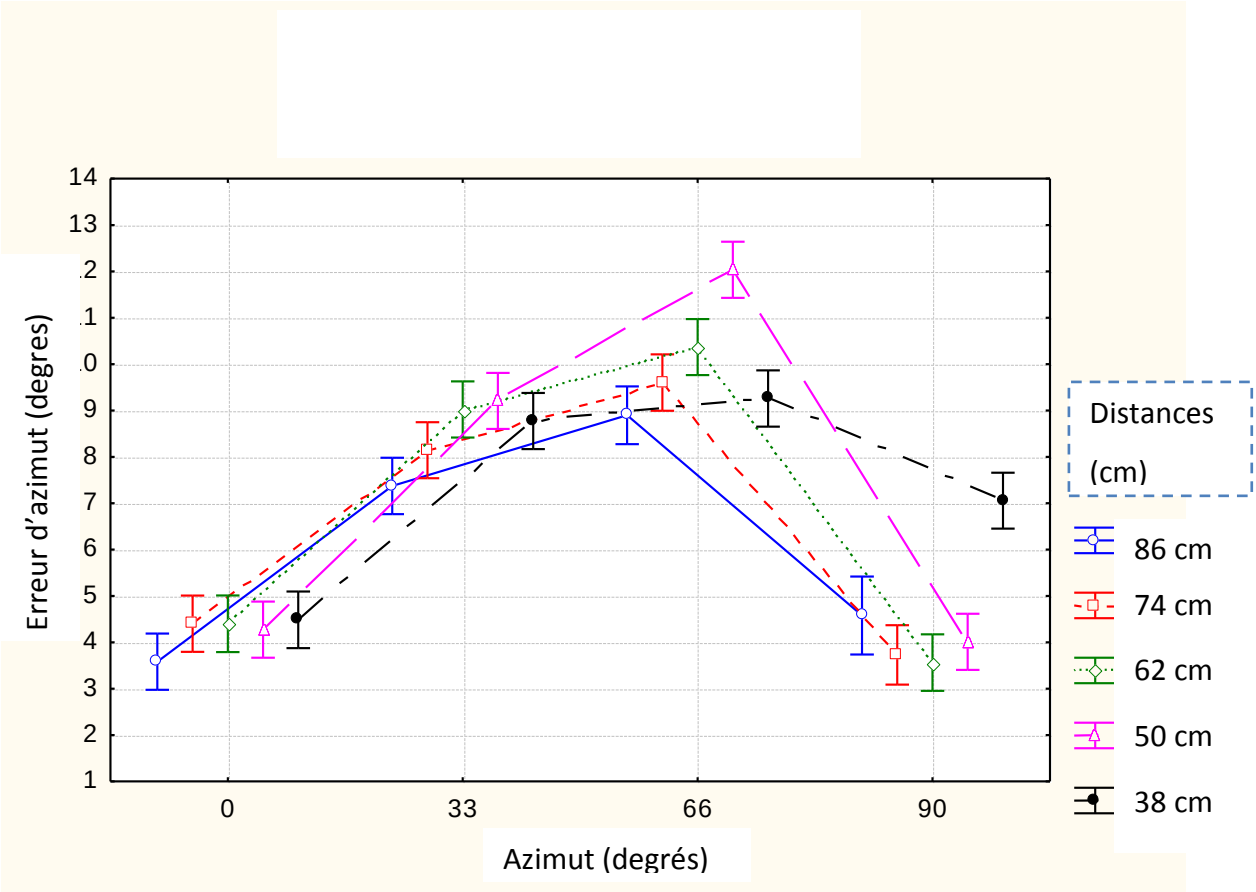


Figure 91: Erreur d'azimut en fonction de la distance et de l'azimut de la source auditive. La moyenne de l'erreur d'azimut sur les 10 sujets est de 4+/-0,3° devant et sur le coté du sujet. La précision diminue entre ces deux paliers, atteignant près de 10° en moyenne pour un azimut de 66°. (Anova, F(12,7255)=10,283 ; p=0,0000)

Discussion

Les résultats illustrés dans les deux figures ci-dessus, peuvent être résumés dans la Figure 92 : dans cette étude, la précision de la distance pointée est meilleure sur le coté que devant le sujet et diminue quand la distance du sujet à la cible auditive augmente. L'erreur d'azimut, en revanche, augmente presque linéairement jusqu'à 66° puis diminue à nouveau pour le dernier azimut testé, en bord de plateau (azimut 90°). Ce résultat n'est pas congruent avec la littérature et peut être expliqué par le fait que les sujets apprenaient rapidement où était le bord du plateau et n'allaient donc jamais pointer dans le vide à droite du dispositif. La faible erreur d'azimut pointé pour l'azimut 90° est donc probablement un biais dû à l'apprentissage du dispositif par les sujets et ne reflète pas les capacités de localisation des sons pour cet azimut. Nous faisons l'hypothèse que sans ce biais, le résultat pour l'erreur d'azimut à la position 90° serait en accord avec la littérature et la précision de pointage continuerait de décroître linéairement.

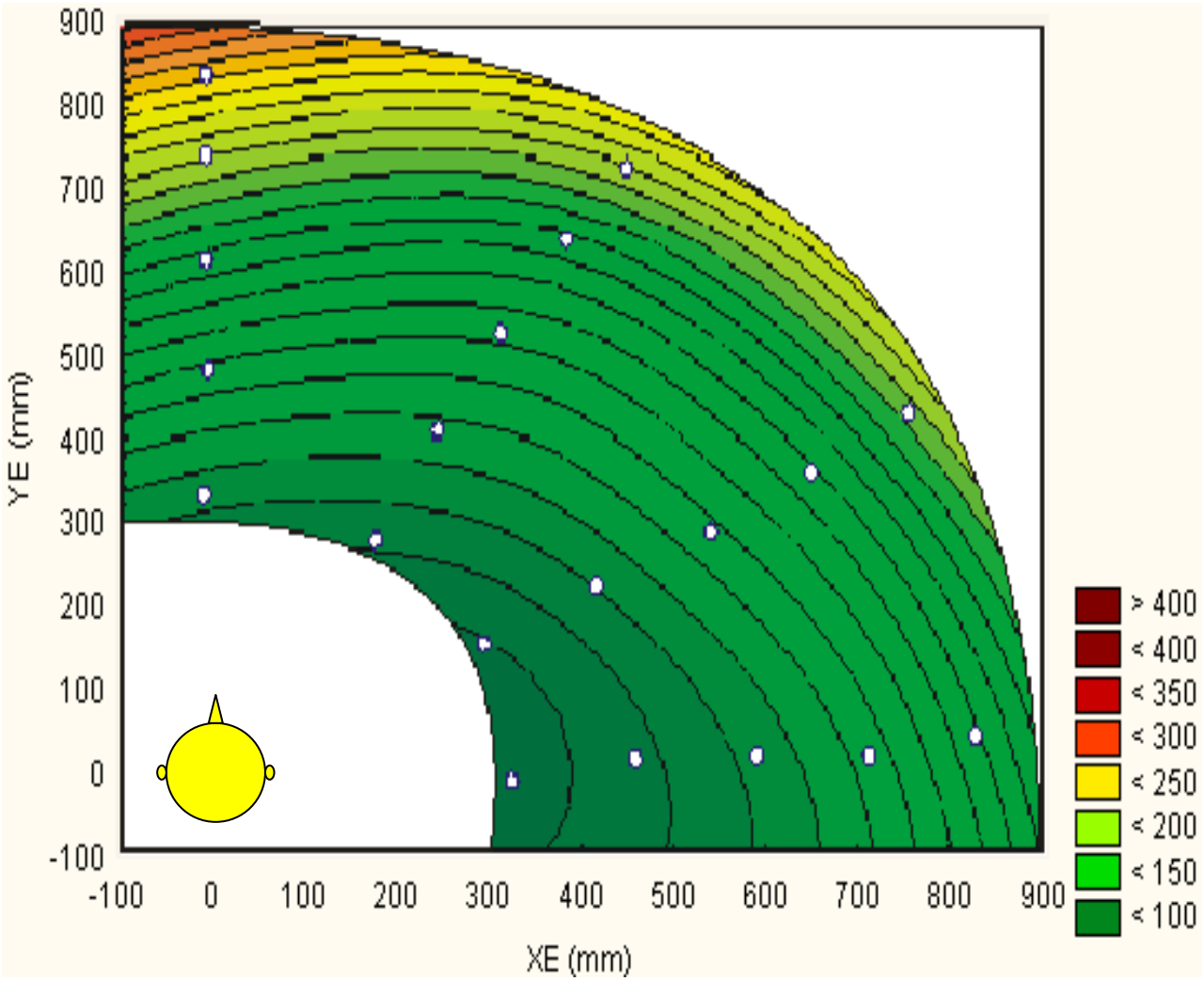


Figure 92: Gauche : précision en distance de pointage dans l'hémi plan droit, pour un même stimulus moyenné sur les 10 sujets. La précision décroît avec la distance et augmente avec l'azimut. Droite : précision en azimut lors d'une tâche de pointage dans le quart de plan droit devant le sujet pour un même stimulus moyenné sur les 10 sujets. Nous observons sur la figure ci-dessus que la précision en azimut décroît jusqu'à 45° puis croît jusqu'à 90°.

Précision en distance en fonction du stimulus

La Figure 93 montre la moyenne de l'erreur d'azimut de pointage chez les 10 sujets en fonction des 4 conditions étudiées (1 :1*200 ms ; 2 : 1*40 ms ; 3 : 3*40 ms ; 4 : 8*40 ms). La moyenne des erreurs angulaires en azimut est faible, quel que soit le stimulus utilisé (entre 6,5 +/- 0,3° et 7,3 +/- 0,3°), avec moins d'un degré de différence entre les conditions ayant les erreurs les plus extrêmes. Il apparaît toutefois que les conditions 3 et 4 ne sont pas significativement différentes : la valeur de l'erreur angulaire pour ces deux conditions étant la même : 6,3 +/- 0,3°. La condition 3, constituée d'un stimulus composé de 3 répétitions d'un son de 40 ms semble ici être le meilleur compromis entre la précision et la durée du son. Il est en effet significativement meilleur que dans les conditions 1 et 2 avec une durée

de son effective de 120 ms et une durée globale de 180 ms, inférieure à la condition 1. La durée d'un son continu semble ainsi influer sur l'erreur d'azimut : il y a une légère différence entre un son de 200 ms (précision de 7,0 +/- 0,3°) et un son de 40 ms (précision de 7,3 +/- 0,3°). Pour une même durée de son, la précision de l'azimut de pointage est plus précise si le son est découpé en plusieurs segments.

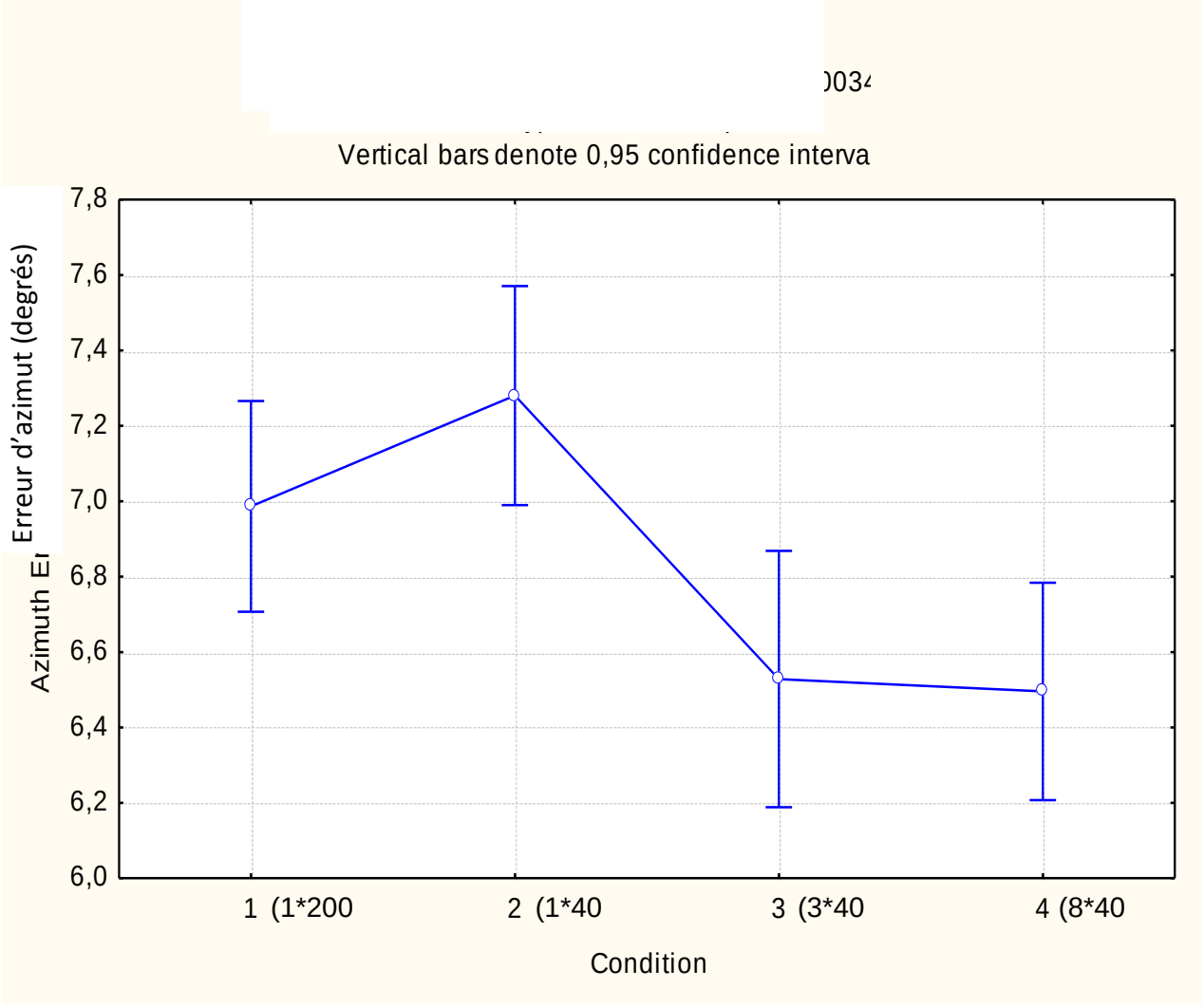


Figure 93: Moyenne des erreurs d'azimut de pointage en fonction des stimuli pour les 10 sujets. Les données avec et sans couettes acoustiques sont groupées. La précision de pointage en azimut est identique (7,1 +/- 0,3°) pour les conditions 1 (1*200 ms) et 2 (1*40 ms) et croit (6,5 +/-0,3°) ensuite pour les 2 autres conditions (3 : 3*40 ms et 4 : 8*40 ms). (Anova, F(3,6045)=6,1996, p=0,00034)

La Figure 94 montre la moyenne des erreurs de distance sur les 10 sujets en fonction des différents stimuli. Les conclusions concernant l'erreur de distance en fonction des différents stimuli sont comparables à celles des erreurs d'azimut. Un son de 40 ms (condition 2) ne semble pas suffisant pour localiser avec la même précision que pour les autres conditions. La précision de la distance de pointage est la même pour les conditions 1,3 et 4, confirmant

qu'il est possible de réduire la durée du stimulus en préservant la précision de localisation en découpant le son.

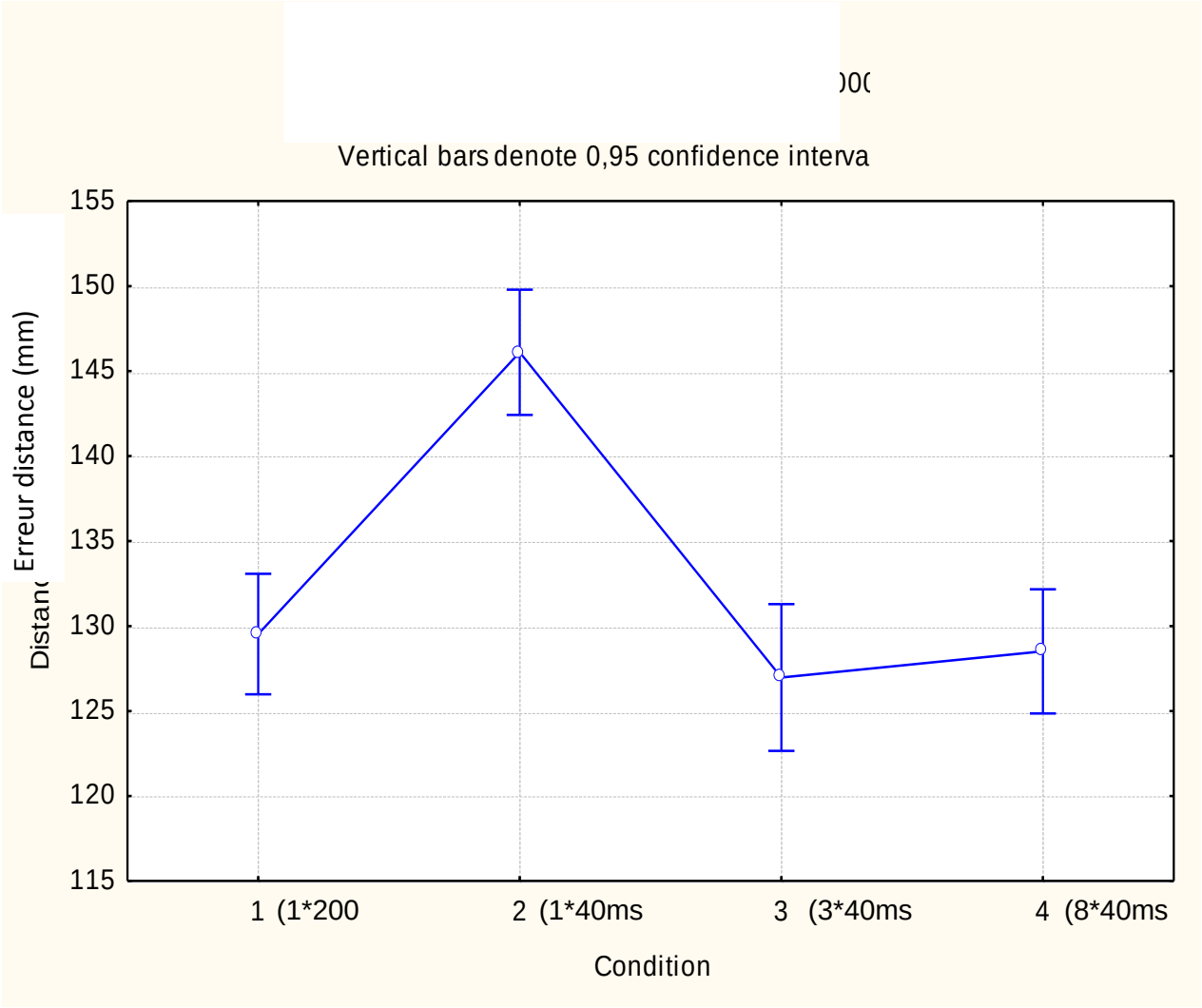


Figure 94 : Moyenne des erreurs de distance de pointage en fonction des stimuli pour les 10 sujets, les données sont groupées avec et sans couettes acoustiques. La précision de pointage est identique (129 mm) pour les conditions 1 (1*200 ms), 3 (1*40 ms) et 4 (3*40 ms) et inférieure pour la condition 2 (1*40 ms). (Anova, F(3,6045)=21,868 ; p=0,00000)

Précision de la distance et de l'azimut du pointage en fonction de l'environnement acoustique de la pièce

Nous avons voulu ici étudier et comparer la précision de localisation avec et sans privation des indices de réverbération des échos. Nous avons ainsi répété l'expérience précédente dans une salle équipée de couettes acoustiques. Dans un environnement où la réverbération est atténuée, la précision du pointage mesurée en azimut et en distance varie comme précédemment en fonction du stimulus. Cependant, on observe que la précision a globalement baissé dans toutes les conditions (Figure 95). La moyenne des erreurs de

distance était de 112 + 2 mm dans la condition sans les couettes acoustiques (avec réverbérations) et atteignait 145 + 2 mm pour la condition avec les couettes acoustiques. Par contre, l'absence de réverbérations n'a eu aucune incidence sur la précision angulaire de la localisation (azimut).

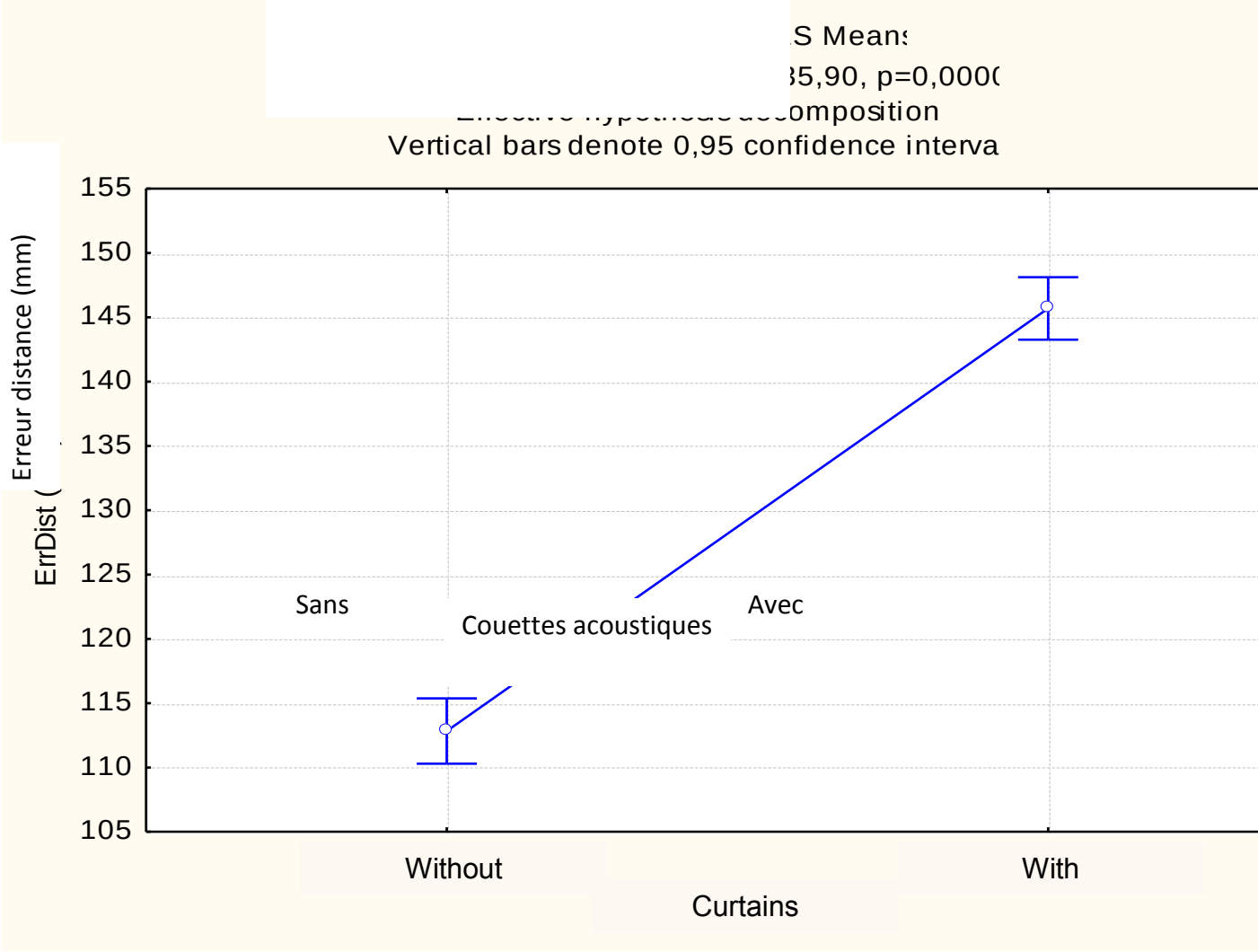


Figure 95: Influence de l'environnement acoustique de la salle d'expérimentation : Erreur de la distance de pointage en fonction des paramètres acoustiques de la pièce. La précision sans couettes acoustiques est de 112 +/- 2 mm, elle atteint 145 +/- 0,2 mm pour une condition avec les couettes acoustiques. (F(1,7255)=335,90 ; p=0,0000)

Discussion sur les résultats de l'expérience 1

Nous avons étudié par cette expérimentation la précision de localisation d'une source sonore en fonction de l'azimut, de la distance, du stimulus et de l'environnement acoustique de la salle. L'erreur de pointage en distance diminue quand l'azimut de la source augmente

de 0° à 90° et augmente avec la distance de la source. Ce résultat est cohérent avec la littérature sur la localisation de sons en champ proche (Brungart et al., 1999) :

- La précision en distance est meilleure sur le coté (sur l'axe interaural) que devant l'utilisateur,
- La précision en distance diminue quand la source s'éloigne.

Une étude antérieure sur l'influence des mouvements de la tête sur la perception de la distance dans le champ proche (Simpson and Stanton, 1973) contredit le résultat sur l'influence de l'azimut sur la perception de la distance. En effet, selon cette étude, la rotation de la tête n'influerait pas sur la perception de la distance.

- La précision en azimut ne semble pas varier en fonction de la distance à la source.
- La précision en azimut diminue jusqu'à un azimut de 60° puis augmente ensuite pour un azimut de 90°

L'erreur d'azimut en fonction de la position de la source sonore n'est pas en accord avec la littérature : cette erreur doit croître avec l'azimut de la source, et ce jusqu'à 90°. Une explication de ce biais pour l'azimut 90 (bord du plateau) est que les sujets ont pu apprendre à détecter ce bord de plateau pour pointer plus juste. Ce biais de pointage est à prendre en considération mais ne modifie pas les résultats relatifs en fonction de la condition. En effet, les valeurs absolues de l'erreur d'azimut pour un azimut de source de 90° sont vraisemblablement faussées. En revanche, cette valeur exprime de manière pertinente l'erreur d'azimut en fonction de la distance mais aussi de l'erreur d'azimut relativement aux conditions. La précision de pointage en azimut de 0 à 60° décroît et rejoint les résultats de la littérature.

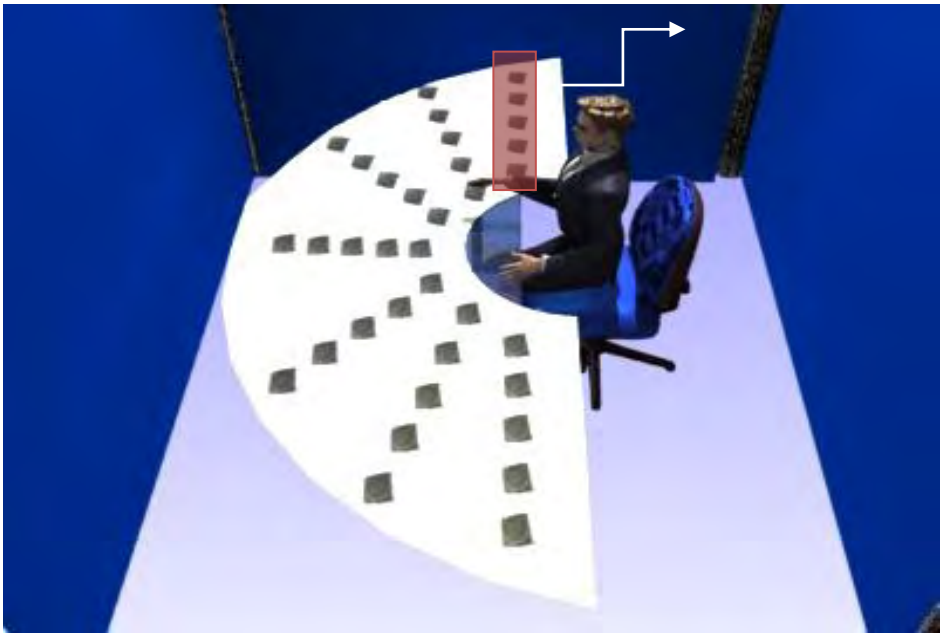


Figure 96: Possible explication du biais de localisation en azimut sur le plateau pour l'azimut 90° : les sujets ont pu apprendre à localiser le bord du plateau puisqu'il y a un vide au-delà de 90° (rectangle rouge). Il est probable que les sujets se soient aidés de cet indice pour localiser plus précisément la source en azimut.

Les bruits blancs utilisés ont été conçus pour couvrir une large bande de fréquence et permettre au sujet d'utiliser le spectre du son. La précision de localisation en azimut et distance semble meilleure quand on augmente la durée du stimulus ou que l'on augmente le nombre de répétitions d'un son court par rapport à un son de même durée effective. L'ensemble des stimuli utilisés étaient suffisamment courts pour ne pas laisser la possibilité aux sujets de tourner la tête pendant la présence du stimulus.

L'erreur d'azimut atteint un minimum pour les deux conditions 3*40 ms (180 ms de stimulus) et 8*40 ms (530 ms de stimulus). La précision angulaire semble ici atteindre un maximum avec 3 répétitions d'un son de 40 ms avec 30 ms de pause entre chaque alors que la durée effective du stimulus est inférieure à la condition 1 : 1*200 ms. L'erreur de distance atteint un minimum pour les conditions 1*200 ms, 3*40 ms et 8*40 ms. Seule la condition 1*40 ms présente une précision dégradée par rapport aux 3 autres. Il semble donc qu'il y ait un seuil de précision atteint entre 1*40 ms et 3*40 ms que nous ne pouvons déterminer ici. De plus, pour une même durée de stimulus (1*200 ms), le fait d'y incorporer des pauses semble donc avoir une influence positive (3*40 ms) sur la précision de pointage en distance.

Expérience 2 : Optimisation de la localisation de sons pour l'interaction non-visuelle

Nous avons voulu dans un deuxième temps tester de manière plus approfondie la précision de localisation d'un son en fonction du son utilisé, dans des conditions non-acoustiquement

traitées. Nous avons ainsi mené une deuxième étude auprès de sujets voyants et non-voyants avec une disposition ne permettant pas de s'aider des bords du plateau.

Matériel et méthodes

Nous avons modifié l'orientation du sujet devant le plateau pour qu'aucune des positions des sources auditives testées ne soit sur un bord du plateau.

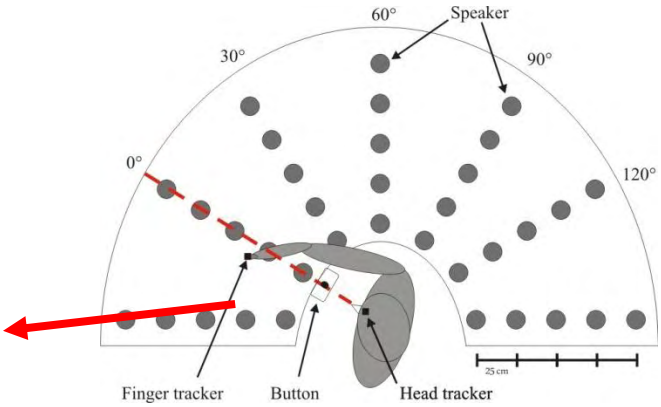


Figure 97: Le sujet était orienté sur l'azimut 0° d'un dispositif expérimental comportant 35 enceintes sous une grille acoustiquement transparente. Le sujet devait poser sa main sur un bouton situé au centre du demi-disque puis aller atteindre le son quand celui-ci apparaissait et revenir ensuite au centre. La position du doigt était mesurée par un outil optique de suivi de mouvement.

L'index de la main droite des sujets était suivi par l'outil COFT présenté précédemment. Une phase de calibrage est nécessaire en début d'expérimentation. La fréquence de l'outil était de 10 Hz et 2 cm d'erreur maximum (sur les bords du plateau). Sept conditions ont été testées en faisant varier le nombre et la durée des stimuli (bruits blancs).

1. 1*10 ms
2. 1*25 ms
3. 1*50 ms
4. 1*200 ms
5. 2*25 ms / 30 ms (50 ms de son effectif / 80 ms de stimulus au total)
6. 3*25 ms / 30 ms (75 ms de son effectif / 135 ms de stimulus au total)
7. 4*25 ms / 30 ms (100 ms de son effectif / 190 ms de stimulus au total)

Seules 27 enceintes ont réellement été utilisées pour éviter les effets liés aux bords du plateau.

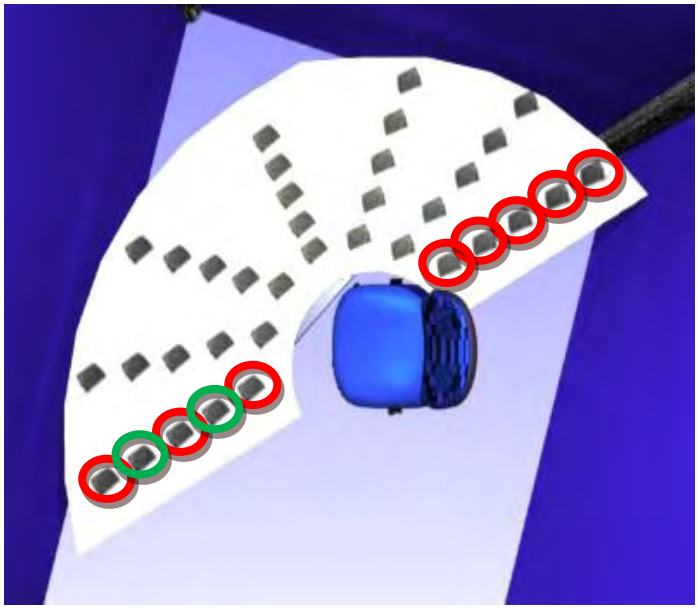


Figure 98: Dispositif expérimental : les enceintes encerclées en rouges étaient désactivées; celles en vert et celles qui ne sont pas entourées étaient actives. Seules les enceintes qui ne sont pas entourées sont prises en compte dans l'analyse des résultats.

L'orientation de la tête était contrôlée par un outil magnétique de capture de mouvement (Flock of Bird, Ascension) avec une précision de 0,5°RMS et une fréquence d'acquisition de 144 Hz. Ce capteur permettait de contrôler l'expérience en ne présentant le son que lorsque l'utilisateur regardait droit devant lui.

Figure 99 : Dispositif expérimental avec le capteur de tête (fixé sur un casque) et la lumière sur le doigt du sujet pour suivre les déplacements du doigt.

Chaque sujet entendait 3 répétitions de chaque stimulus sur chacun des 27 haut-parleurs testés, pour un total de 756 stimuli. 19 sujets ont été testés : 11 sujets voyants et 8 sujets non-voyants. Nous commencerons par présenter les résultats sur l'ensemble des 19 sujets puis comparerons les performances des sujets voyants et non-voyants.

Résultats sur l'ensemble des sujets

Comme pour l'expérimentation précédente, la position de pointage est ensuite extraite des données du suivi de l'index. Les résultats obtenus sont classés comme pour l'expérimentation précédente, selon deux principales variables : l'erreur d'azimut et l'erreur de distance pointées. L'écart moyen entre la source sonore et la position pointée est stable dans le temps sur toute la durée de l'expérimentation (Figure 100) pour chaque sujet.

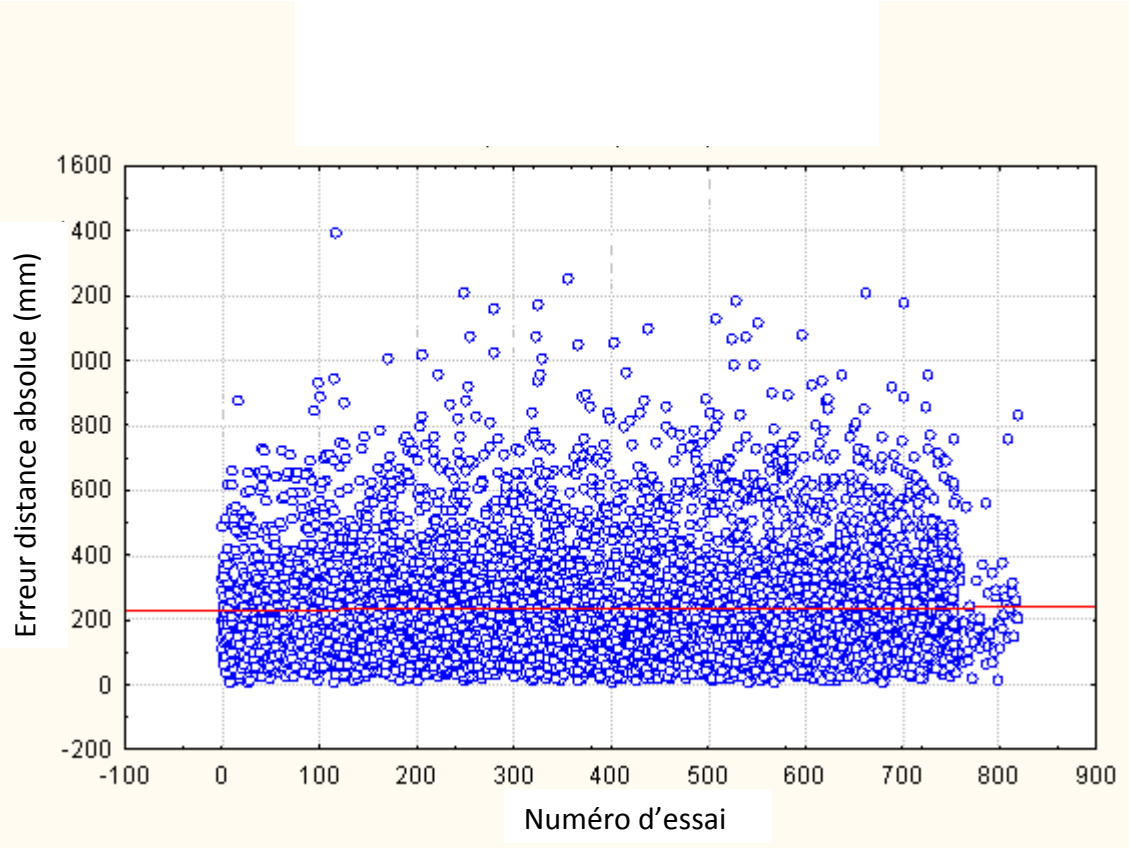


Figure 100 : l'erreur de distance absolue est calculée par la distance entre la position pointée et la position de la source sonore. La figure ci-dessus affiche l'erreur de distance absolue en fonction du numéro d'essai (classé chronologiquement). L'erreur est donc très stable au cours de l'expérience et nous avons obtenu ce même résultat pour chaque sujet.

Ce résultat est le même pour l'erreur d'azimut et l'erreur de distance pointées et montre que la précision de localisation reste stable même si l'attention baisse au cours de l'expérimentation.

Erreur « devant-derrière » (front-back)

Nous avons décrit précédemment les trois principaux indices permettant de localiser une source sonore en condition binaurale : ITD, ILD et le spectre du son. Il a été montré (Brungart, 1999) que le spectre du son était très important pour désambigüiser si la source se trouve devant ou derrière le sujet. En effet, la déformation fréquentielle du son arrivé jusqu'au tympan et altéré principalement par le pavillon de l'oreille est un indice important de la position de la source. Nous avons voulu savoir dans le plan horizontal, quelles étaient les conditions qui engendraient le plus d'erreurs avant-arrière.

Le pourcentage d'erreur avant-arrière semble être très dépendant de l'azimut de la source. En effet, ce phénomène peut être largement observé dans les zones proches de 90° ou plus largement pour les sources derrière le sujet.

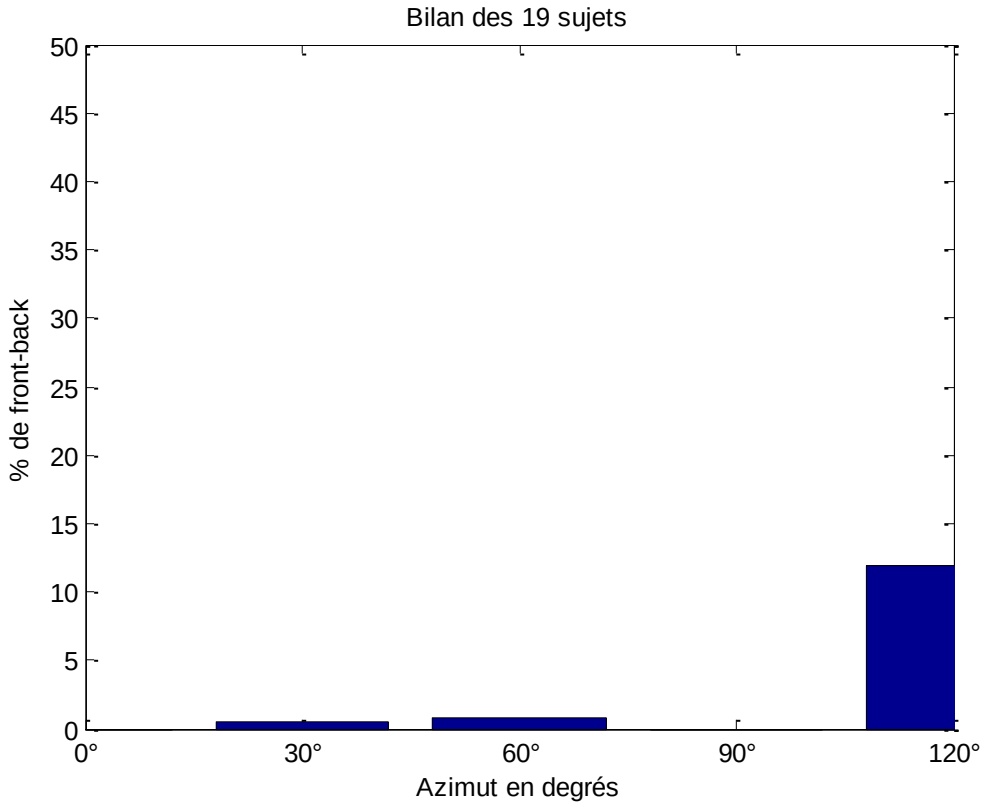


Figure 101: Pourcentage d'erreurs avant-arrière en fonction de l'azimut de la source. Le pourcentage d'erreur augmente avec la latéralité de la source. Le nombre d'erreurs avant-arrière est très élevé quand la source est derrière le sujet. Aucune erreur de symétrie autour de 90+-15° n'a été considérée comme une erreur devant-derrière.

La Figure 102 présente le pourcentage d'erreurs avant-arrière chez tous les sujets confondus en fonction des conditions. Le pourcentage d'erreur est maximal pour la condition 1 (1*10 ms ; 3,2% d'erreur), puis elle atteint un palier à environ 2,5% pour les conditions 2 (1*25 ms), 3 (1*50 ms), et 4 (1*200 ms). Le pourcentage d'erreur diminue ensuite jusqu'à la condition 6 (3*25 ms, 1,9% d'erreurs) pour remonter à 2,4% pour la condition 7 (4*25 ms). Il semble donc y avoir une incidence de la durée du son et du nombre de répétitions sur le pourcentage d'erreurs avant-arrière.

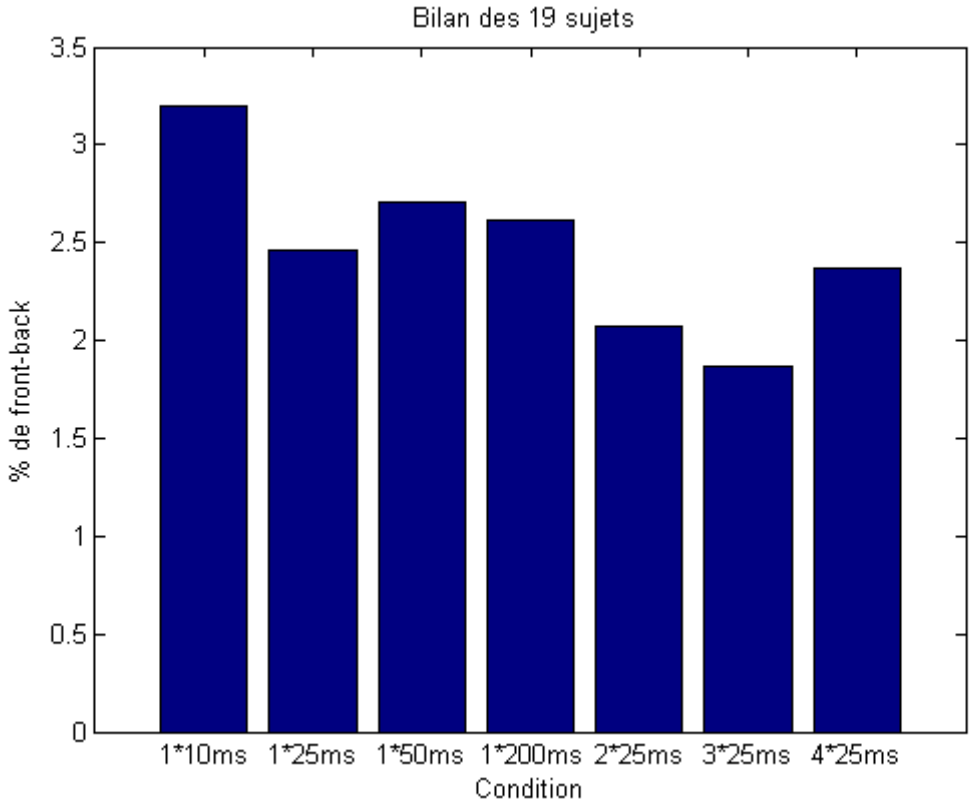


Figure 102: Pourcentage d'erreurs avant-arrière en fonction de la condition. Afin de comparer le pourcentage d'erreurs en fonction des conditions, l'échelle des pourcentages a un maximum de 3,5% d'erreur.

Ellipses de confiance

Les erreurs d'azimut et de distance sont des bons indicateurs de la performance des sujets pendant une tâche de pointage. Ils ne donnent en revanche pas d'informations sur la distribution spatiale des pointages et leur variabilité. Ces données peuvent être étudiées par le biais d'ellipses de confiance englobant par exemple 95% des pointages (Figure 103).

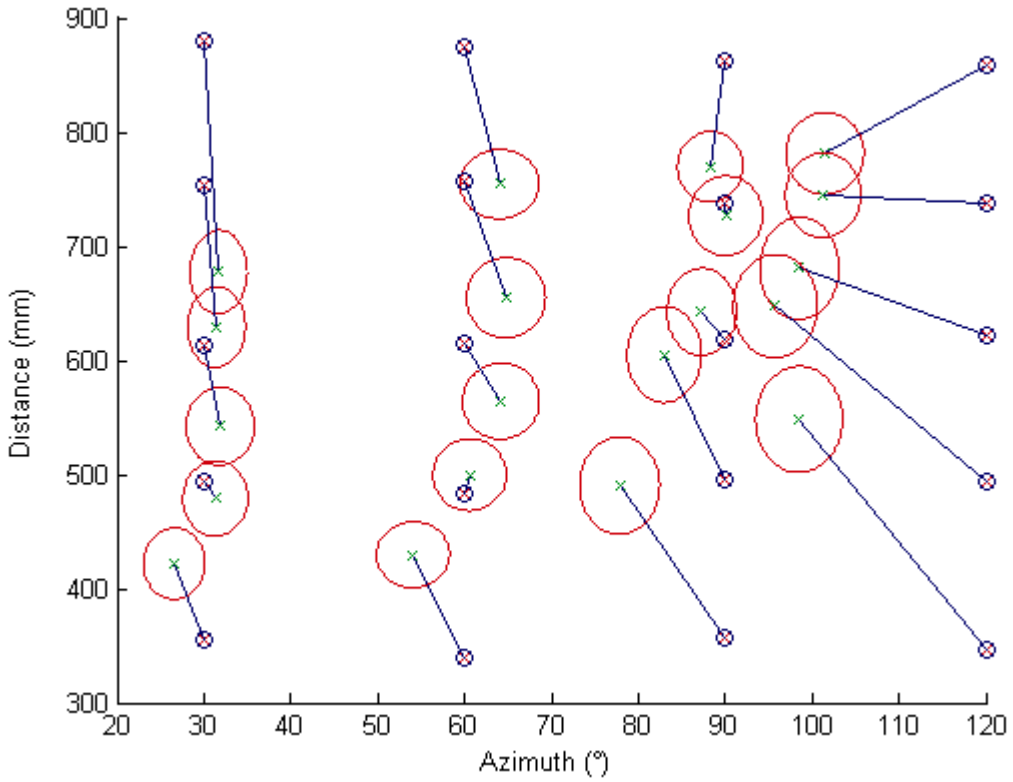


Figure 103: Les ellipses rouges correspondent aux ellipses de confiance (95% des pointages) pour l'ensemble des sujets pour la condition 2*40 ms. Les ellipses sont liées par leur centre (croix vertes) à l'enceinte associée (croix rouges). La taille des ellipses est proportionnelle à l'erreur de pointage.

Les sujets ont une tendance très claire à sous-estimer la distance aux cibles auditives lointaines et surestimer les cibles auditives très proches. L'azimut pointé est de plus en plus sous-estimé au fur et à mesure que l'azimut de la cible augmente. Cette tendance peut être expliquée par le fait que la ligne d'épaule cache une partie du stimulus auditif pour les sources proches et sur le coté. De nombreux autres facteurs influent sur la précision de pointage et nous allons étudier ici l'influence des stimuli et de la position de la source sonore sur la précision de pointage en azimut et en distance.

Erreur de l'azimut pointé en fonction de la condition et de la position de la source sonore

L'erreur d'azimut pointé est calculée en enlevant les essais considérés comme des erreurs avant-arrière. Ce phénomène a été observé uniformément sur toutes les conditions et ne change pas les résultats.

L'erreur d'azimut en fonction des conditions montre que cette précision de localisation angulaire est dépendante de la durée du stimulus et du nombre de répétitions du son

(Figure 104). En effet, les erreurs d'azimut dans les conditions avec un son continu de 10 ms (condition 1) jusqu'à 200 ms (condition 4) décroissent (de 11,8° d'erreur à 10,5°), montrant ainsi que la précision augmente significativement ($p<1,10^{-5}$) avec la durée du stimulus. Les conditions 1, 2 ,3 et 4 sont significativement différentes les unes des autres d'après un test post-hoc (Tukey, $p<0.0005$). La précision des conditions 2, 5, 6 et 7 (1*25ms ; 2*25ms ; 3*25 ms ; 4*25 ms) ne sont pas significativement différentes entre elles concernant l'erreur de l'azimut pointé. Ce résultat montre que la précision dans cette tâche ne s'améliore pas indéfiniment avec le nombre de sons et la durée des sons. La précision semble atteindre un plateau à partir de la condition 5 : pour deux sons de 25 ms entrecoupés de 30 ms de pause. Cette condition 5 a une durée totale de stimulus de 80 ms et la même précision que la condition 4 (1*200ms). Ce résultat montre la faculté des sujets pour localiser des sources, avec de meilleures performances pour des sons entrecoupés de silence que pour des sons continus, à durée de son égale.

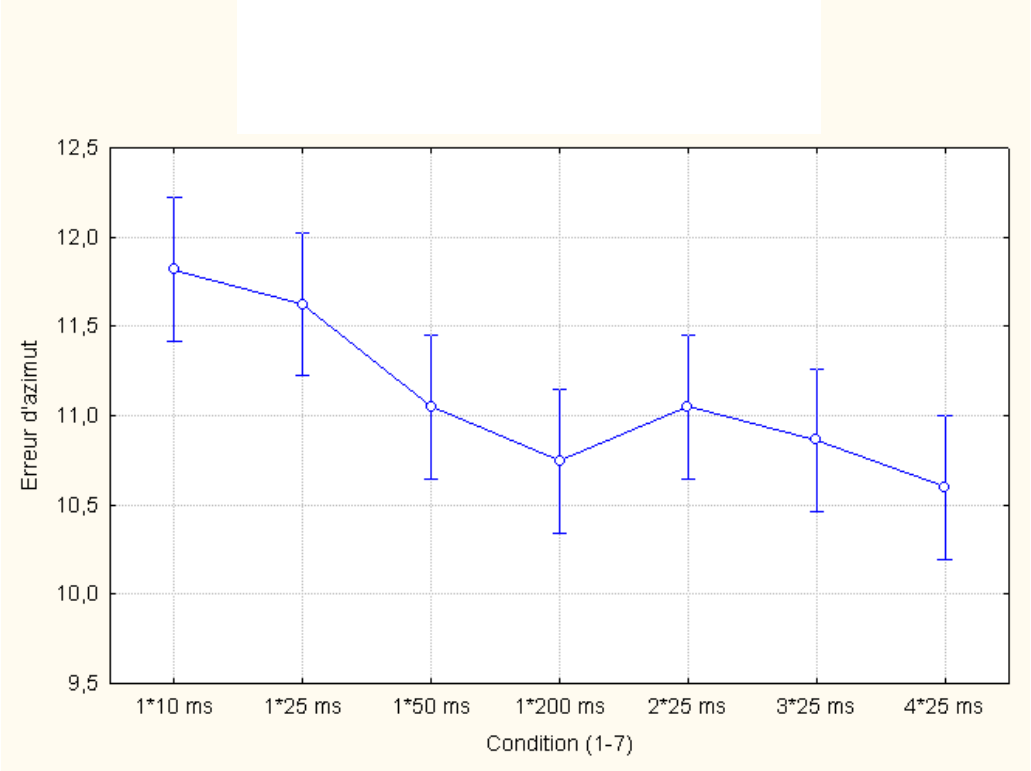


Figure 104 : Moyenne de l'erreur d'azimut pointé en degrés sur l'ensemble des 19 sujets en fonction des 7 conditions. Entre les conditions 1 et 4, l'erreur d'azimut décroît avec la durée du stimulus. Les autres conditions montrent qu'il est préférable pour une même durée de stimulus, d'avoir un son discontinu. (Anva, F(12,25696)=8,5814 ; p=0,00000)

L'erreur d'azimut en fonction de la position de la source nous permet de catégoriser l'espace péripersonnel en fonction de la précision de localisation. Nous allons dans un premier temps

étudier la précision de localisation angulaire en fonction de l'azimut de la source. Mais à cause de la méthode de calcul de cette composante, l'erreur d'azimut en fonction de la distance est triviale et se trouve être beaucoup plus grande à des distances proches de l'utilisateur. C'est pour cela que cette variable n'a pas été étudiée en fonction de la distance de la source.

En accord avec la littérature (Blauert, 1997;Brungart et al., 1999), les résultats de la Figure 105 montrent que l'erreur d'azimut est très dépendante de l'azimut de la source sonore. En effet, la précision devant le sujet (erreur d'azimut de 6,1+/-0,1° à 0° d'azimut) est très supérieure à celle sur le coté (erreur de 11+/-0,1° à 90°). La précision à 120° (erreur de 21+/-1°) est certainement très dépendante de la tâche car il était difficile de produire le geste pour pointer derrière soi.

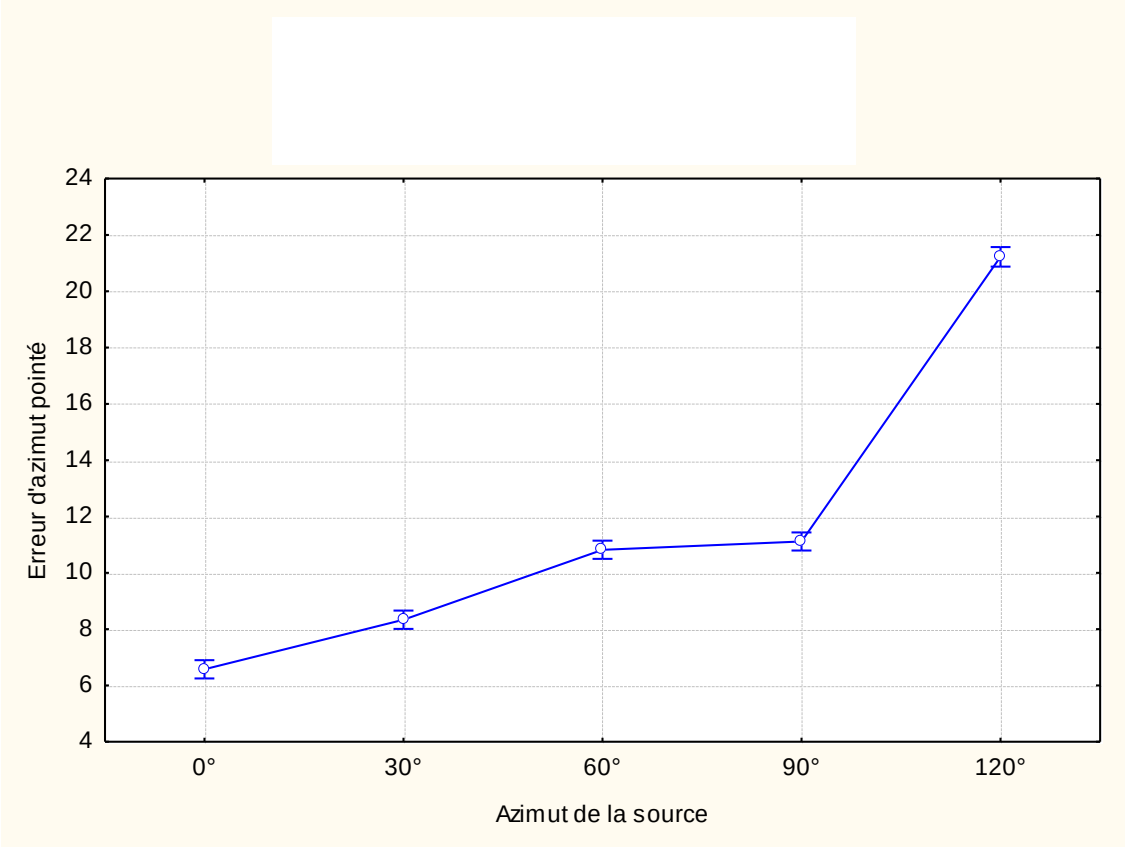


Figure 105 : Moyenne de l'erreur d'azimut des 19 sujets en fonction de l'azimut de la source sonore. L'erreur augmente avec l'azimut de la source. (Anova, F(4,12823)=1071,4 ; p=0,0000)

Erreur de distance en fonction de la condition et de la position de la source sonore

L'erreur de distance est la valeur absolue de la différence entre la distance de l'utilisateur à la source et la distance de l'utilisateur à la position pointée. Nous allons dans un premier temps étudier la précision de pointage en fonction de l'azimut de la source sonore, puis de sa distance et enfin des différentes conditions. La Figure 106 montre que l'erreur de distance dans la tâche de pointage est dépendante de l'azimut de la source. L'erreur diminue de l'azimut 0° (127 +/-4 mm d'erreur) jusqu'à 60° (103 +/-4 mm d'erreur) puis remonte jusqu'à 120° d'azimut (145 +/-4 mm d'erreur).

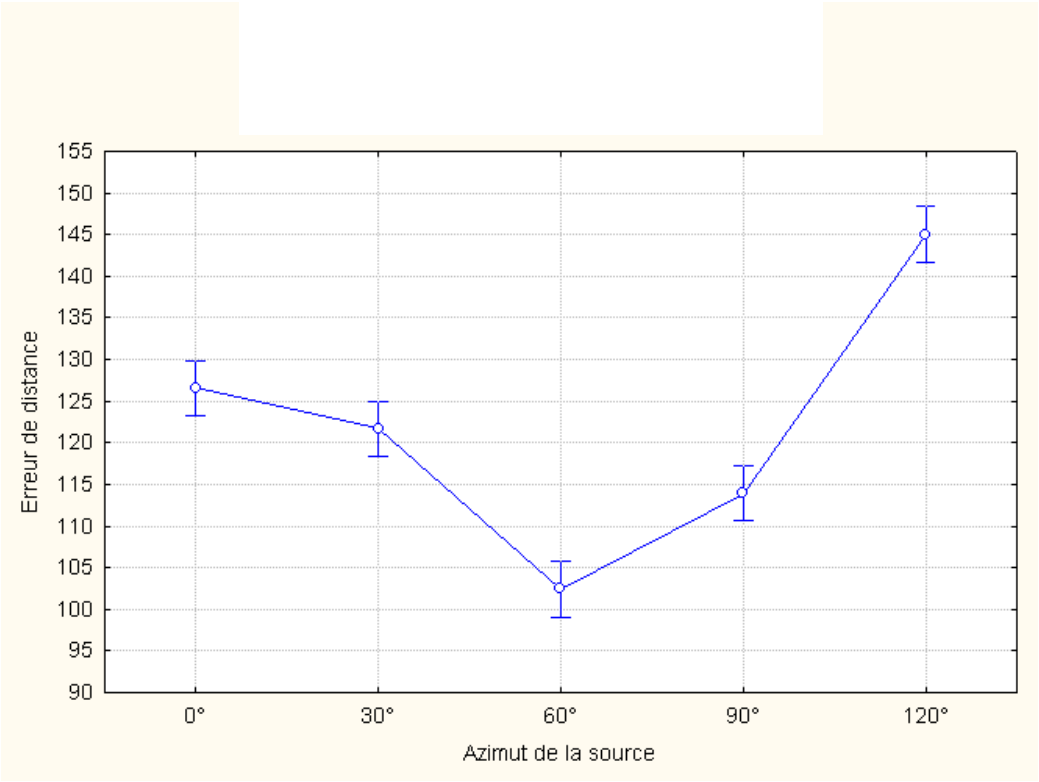


Figure 106: Moyenne de l'erreur de distance sur tous les sujets par rapport à l'azimut de la source sonore. La précision augmente avec l'azimut de la source jusqu'à 60° et diminue ensuite. (F(8,25696)=413,25, p=0,0000)

Nous obtenons une courbe similaire à la précédente pour l'erreur de distance par rapport à la distance de la source (Figure 107) : la précision augmente de 345 mm de distance (128+/-4 mm d'erreur) à 595 mm (109+/-4 mm d'erreur) et diminue ensuite jusqu'à 835 mm de distance (145+/-4 mm d'erreur).

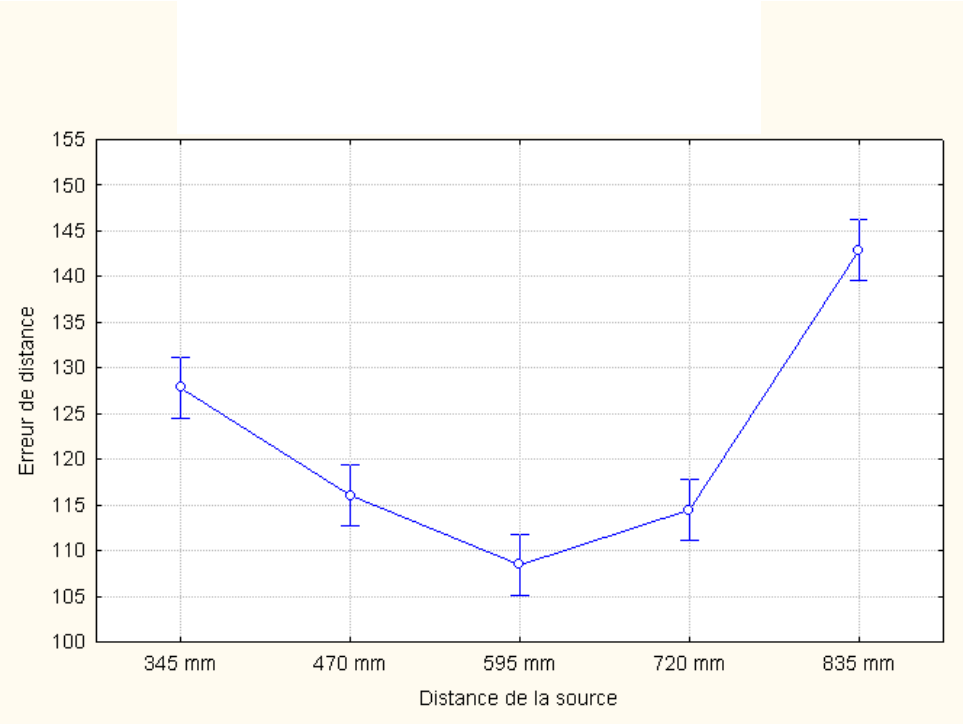


Figure 107: Moyenne de l'erreur de distance pointée en fonction de la distance de la source sonore. (F(8,25696)=49,318 ; p=0,0000)

La moyenne de l'erreur de distance pointée est très dépendante de la condition. L'erreur diminue linéairement de la condition 1 à la condition 4 avec la durée du son qui augmente (de 133+/-4 mm d'erreur à 115+/-4 mm). Les conditions 4, 5,6 et 7 présentent une erreur moyenne inférieure à 120+/-4 mm avec une précision minimale pour 2 sons de 25ms (123+/-4 mm d'erreur), qui augmente ensuite avec le nombre de répétitions du son jusqu'à 4 sons de 25 ms (115+/-4 mm d'erreur)

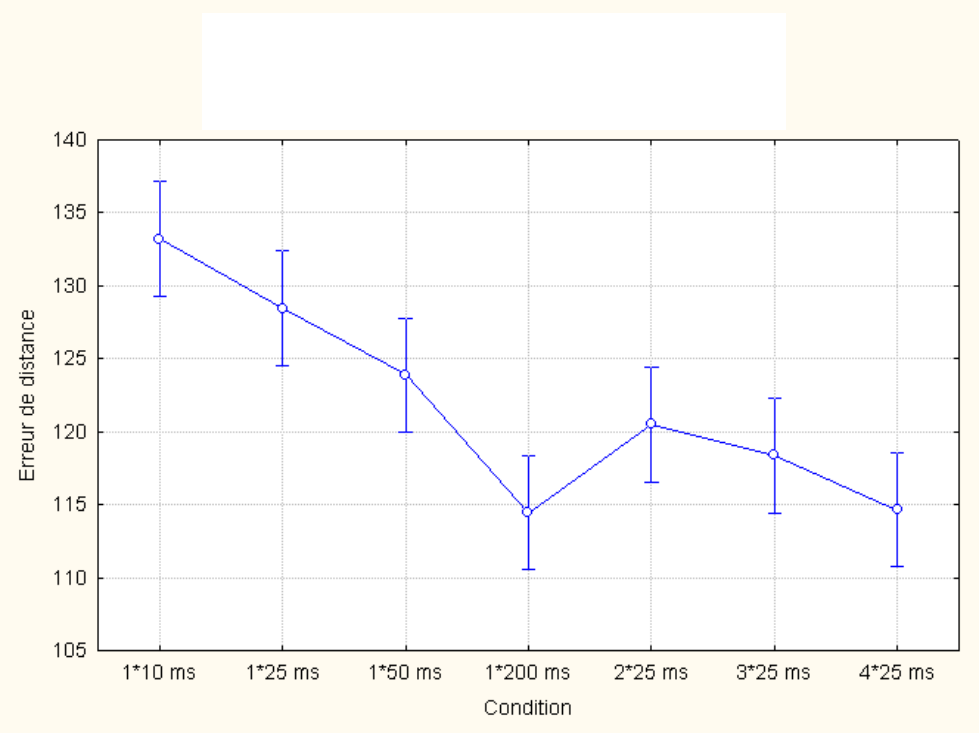


Figure 108: Moyenne de l'erreur de distance pointée en fonction de la condition. (Anova, F(12,25696)=8,5814, p=0,00000)

Discussion

En moyenne chez tous les sujets, les résultats de cette étude confirment ceux obtenus lors de l'expérience préliminaire : la précision de pointage est dépendante de la durée du son et du nombre de répétitions. Les résultats obtenus concernant l'erreur d'azimut sont conformes avec la littérature : l'erreur de l'azimut pointé augmente avec l'azimut de la source. L'erreur angulaire dans cette tâche est restée faible pour l'ensemble des sujets, quelle que soit la condition (<12°), depuis l'azimut 0° (devant) jusqu'à l'azimut 90°. Cette erreur est minimale pour les conditions 4, 5, 6 et 7. L'erreur de distance pointée est comprise entre 133 et 115 mm en moyenne et atteint un minimum (115 mm) pour les conditions 4 (1*200 ms) et 7 (4*25 ms). Les conditions 4, 5, 6 et 7 obtiennent toutes des erreurs moyennes inférieures à 120 mm

L'erreur de distance absolue représente la distance entre la source sonore et la position de pointage en millimètres. En étudiant cette mesure, nous retrouvons certains résultats décrits pour la précision azimutale et de distance. La précision absolue est la meilleure devant le sujet et proche du sujet (moyenne < 120 mm), diminue avec l'azimut et la distance de la source. Il existe une zone dans le plan horizontal (plateau) où la précision reste toujours

inférieure à 180 mm : elle se situe sur une ligne entre les azimuts 0° et 90°, à une distance de 835 mm. Ce résultat peut avoir plusieurs explications. Il apparaît que l'erreur de la distance de pointage est la plus faible pour une distance de la source à 595 mm en moyenne. Les sujets ont pu en cas d'incertitude sur la distance, développer des stratégies de pointage (aller pointer le plus loin, le plus proche, à distance de bras …) et cela pourrait expliquer cette disparité inter-sujet des erreurs de distance. L'erreur d'azimut et de distance pour un pointage derrière les sujets peut être expliquée par des difficultés biomécaniques pour atteindre la localisation des cibles perçues.

Très peu d'erreurs d'avant-arrière ont été effectuées en moyenne chez les sujets, mais une minorité d'entre eux ont concentré la majorité des ces erreurs. Les résultats obtenus montrent qu'il existe une grande différence inter-sujet dans la perception binaurale. Nous avons ensuite étudié les différences de pointage entre des sujets voyants et des sujets non-voyants qui pourraient expliquer cette hétérogénéité des résultats.

Résultats : Comparaison des performances de localisation entre les sujets voyants et non-voyants

Il a été montré que les voyants et les non-voyants n'utilisaient pas exactement les mêmes indices binauraux pour localiser une source sonore (Dramas et al., 2008;Lessard et al., 1998). Les résultats seront ici présentés comme pour le paragraphe précédent : après avoir présenté les erreurs avant-arrière chez les sujets voyants et non-voyants, nous nous intéresserons aux erreurs de localisation en termes d'erreur d'azimut pointé et d'erreur de distance pointée. Il a été montré que le spectre du son était particulièrement utile pour localiser une source (Kulkarni and Colburn, 1998b) quand l'ILD et l'ITD ne sont pas efficients. Il semble que ces modulations fréquentielles du son d'une position à l'autre permettent en particulier d'en estimer l'élévation et de désambigüiser les erreurs avant-arrière (Brungart, 1999;Middlebrooks and Green, 1991). Non seulement les non-voyants semblent développer une plasticité cérébrale particulière pour la perception auditive (Roder et al., 1999) mais aussi développer des capacités de localisation plus fines dans l'espace extra-personnel (Voss et al., 2004).

Erreur « avant-arrière » (Front-back)

Nous nous sommes intéressés à ces erreurs pour les sujets voyants et non-voyants. Le pourcentage des erreurs avant-arrière chez les non-voyants est beaucoup plus important

(3,2%) que chez les sujets voyants (1,9%). Ce résultat s'explique plus particulièrement par l'un des sujets qui faisait trois fois plus d'erreurs avant-arrière que la moyenne des autres sujets non-voyants. La moyenne sans ce sujet est de 1,9% d'erreurs avant-arrière pour les non-voyants, au même niveau que les voyants.

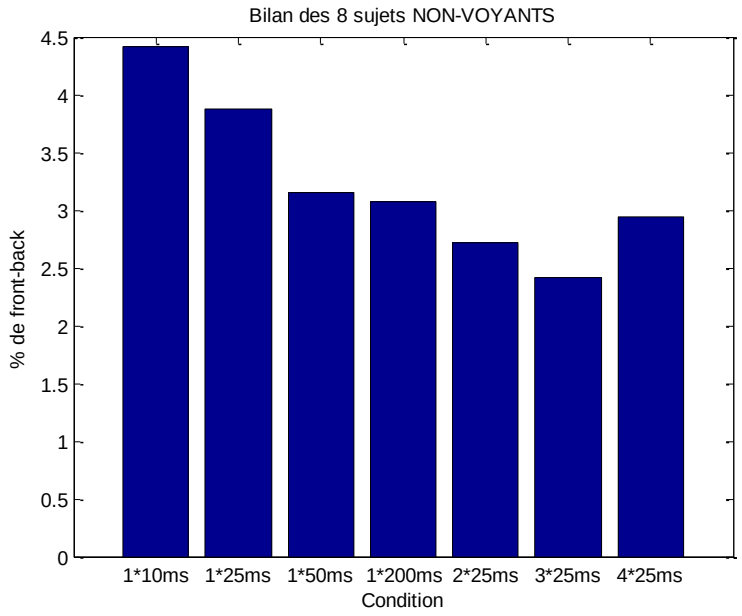


Figure 109: Moyenne des pourcentages d'erreurs avant-arrière en fonction de la condition pour les sujets non-voyants. Le pourcentage décroît avec l'augmentation de durée des sons et le nombre de répétitions jusqu'à la dernière condition où les sujets semblent faire plus d'erreur. Une grande échelle a été choisie pour l'affichage des résultats afin de comparer les conditions entre elles.

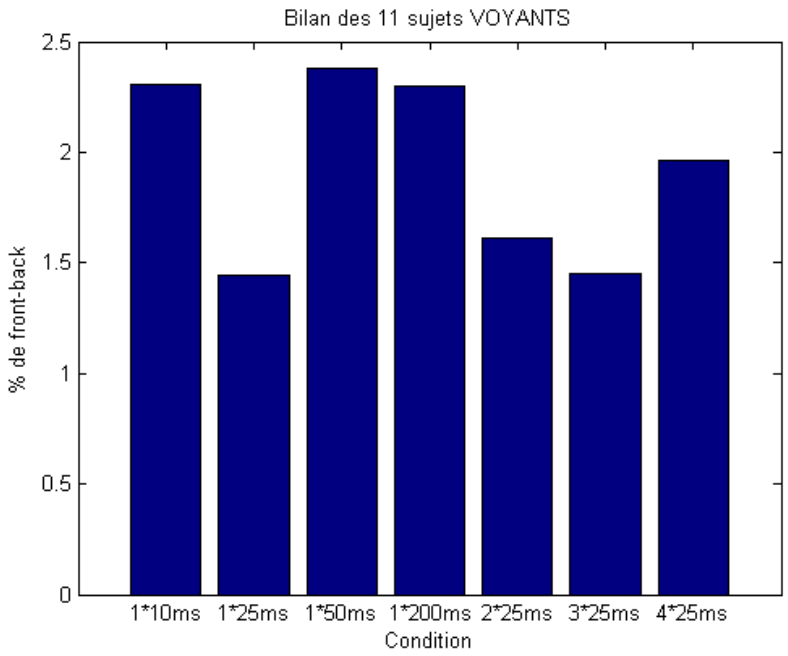


Figure 110: Moyenne des pourcentages d'erreurs avant-arrière en fonction de la condition pour les sujets voyants. Les trois conditions pour lesquelles il y a eu le plus d'erreurs sont les conditions 1(1*25ms), 3 (1*50ms), 4 (1*200ms) et 7 (4*25ms), où le pourcentage d'erreurs est supérieur à 2%. Les conditions 2 (1*25ms), 5 (2*25ms) et 6 (3*25ms) ont un pourcentage d'erreurs d'environ 1,5%

Les Figure 109 et Figure 110 présentent le pourcentage erreurs avant-arrière en fonction des 7 conditions. Les résultats séparés des voyants et des non-voyants montrent de nombreuses similarités dans le pourcentage d'erreurs avant-arrière en fonction de la condition. Pour les voyants et les non-voyants, la condition 6 (3*25ms) engendre le pourcentage d'erreur le plus faible (2,5% pour les non-voyants et 1,5 pour les voyants). Les conditions 1 (1*10ms), 3(1*50ms), 4(1*200ms) et 7 (4*25ms) engendrent pour les voyants un pourcentage d'erreurs élevé (>2%). Les résultats des non-voyants montrent que le pourcentage d'erreur diminue de manière continue de la condition 1 jusqu'à la condition 6 où ce pourcentage remonte ensuite pour la condition 7 de 2,5% à 3%. Dans les deux cas, les conditions engendrant le pourcentage le plus faible d'erreur sont les conditions 5 (2*25ms) et 6 (3*25ms).

Erreur de l'azimut pointé en fonction de la condition et de la position de la source sonore

La précision en azimut est très bonne aussi bien chez les sujets voyants que chez les sujets non-voyants. La moyenne des erreurs d'azimut est la même pour les deux groupes de sujets (9° entre 0 et 90° et 11° sur l'ensemble des azimuts entre 0° et 120°). L'erreur de l'azimut pointé en fonction de l'azimut de la source (Figure 111) évolue de la même façon pour les sujets voyants et les sujets non-voyants, avec une erreur qui augmente avec l'azimut jusqu'à

90° (11+/-0,5°) puis augmente très fortement pour doubler pour l'azimut 120°, derrière le sujet (21+/-0,5° pour les sujets voyants et 17+/-0,5° pour les sujets non-voyants). L'erreur d'azimut pointé en fonction de la condition évolue chez les deux groupes de sujets de la même manière que les résultats présentés dans la Figure 104 : un minimum d'erreur est atteint pour les conditions 4 (1*200ms), 5 (2*25ms), 6 (3*25ms) et 7 (4*25ms) pour une moyenne de 10,4+/-5° d'erreur d'azimut pointé et 11,3+/-4° d'erreur d'azimut pointé respectivement pour le groupe de sujets non-voyants et voyants. L'erreur pour ces conditions est donc plus élevée (1°) pour les sujets voyants.

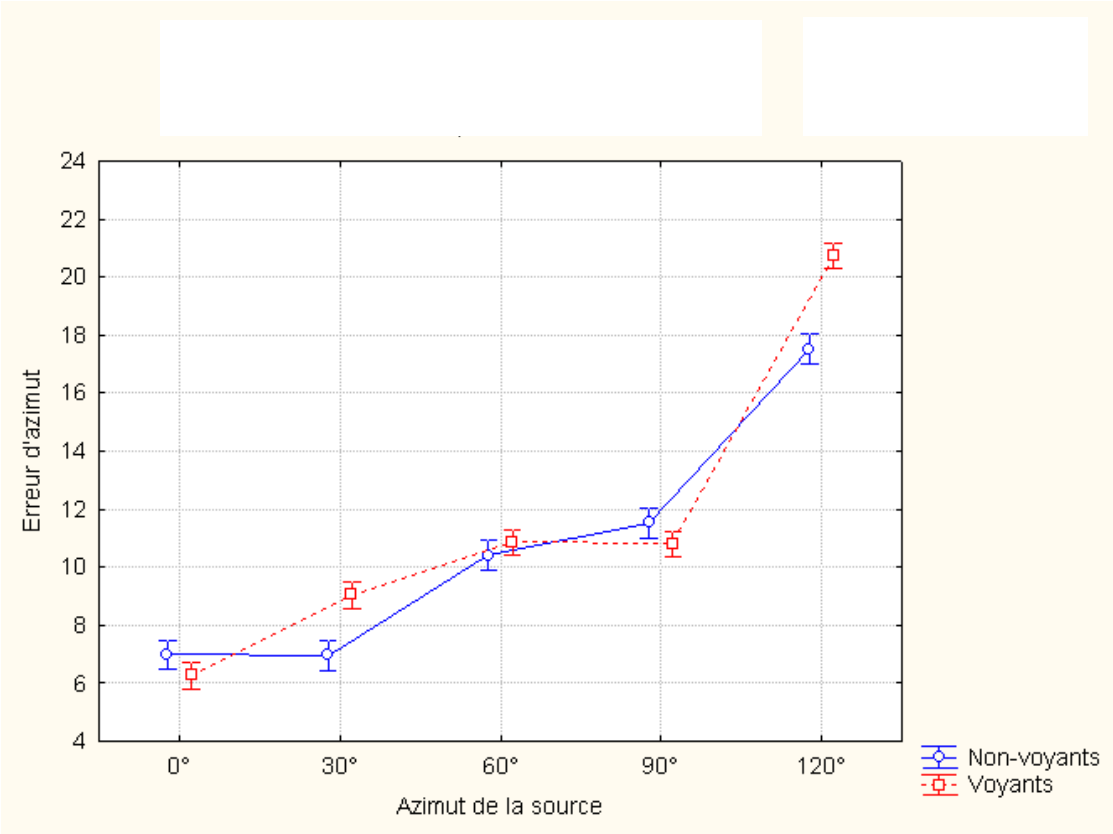


Figure 111: Erreur d'azimut pointé en fonction de l'azimut de la source pour les deux groupes de sujets : voyants (rouge) et non-voyants (bleu). (Anova, F(8,25696)=20,619 ; p=0,0000)

Erreur de distance pointée en fonction de la condition et de la position de la source sonore

L'erreur de distance pointée en fonction de la position de la source (Figure 112) révèle des différences de précision entre les groupes de sujets voyants et non-voyants. L'erreur de distance moyenne pour le groupe de sujets non-voyants est de 128+-2 mm et de 116+-2 mm pour les sujets voyants avec une significativité très élevée (p<10^-5). L'erreur de distance pointée est donc beaucoup plus élevée dans le groupe des sujets non-voyants.

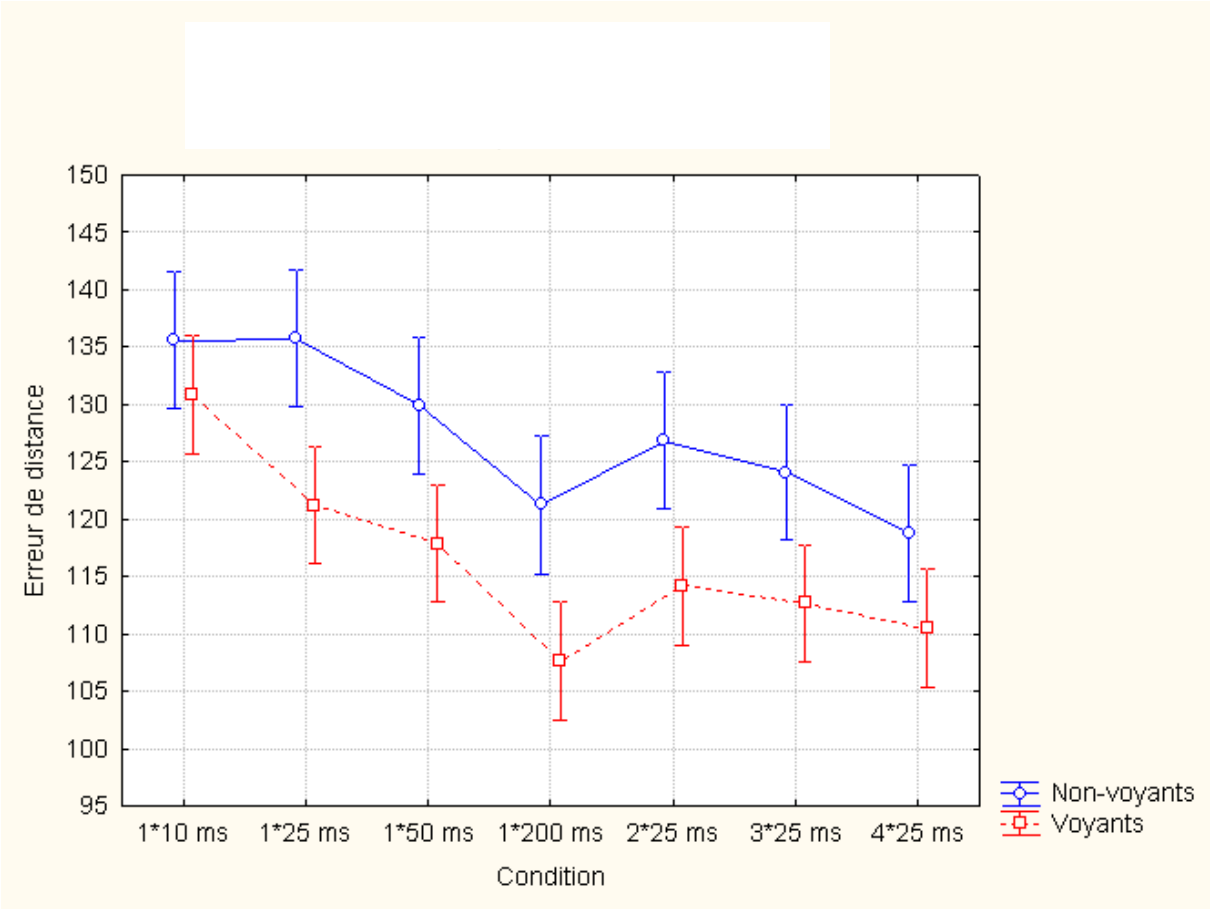


Figure 112: Moyenne des erreurs de distance en fonction de la condition (1-7) et du groupe de sujets : voyants (rouge) et non-voyants (bleu). (F(12,25696)=0,64576 ; p=0,80439)

Les résultats sur les erreurs de distance en fonction de la condition montrent que quelle que soit la condition, les non-voyants allaient pointer avec une erreur de distance bien supérieure au groupe des sujets voyants en moyenne. Cette différence n'est pas statistiquement significative quand on considère les conditions séparément, vraisemblablement à cause d'une trop grande variabilité inter-sujet. L'effet de la condition sur chaque groupe est en revanche très significatif (<10^-4). Les deux groupes ont une erreur maximale pour la condition 1 (1*10ms) (Erreur de 135+/-12 mm pour les sujets non-voyants et 131+/-12 mm pour les sujets voyants). L'erreur décroît ensuite jusqu'à la condition 4 (1*200 ms). L'erreur pour les conditions 5, 6 et 7 décroît faiblement en moyenne et est inférieure à 127 mm pour les non-voyants et 115 mm pour les sujets voyants.

L'étude de la moyenne des erreurs de distances pointées en fonction de l'azimut de la source (Figure 113) montre qu'elle évolue différemment dans les deux groupes de sujets avec une significativité élevée (p<10^-4). L'erreur pour les deux groupes de sujets est

d'environ 130 mm pour les azimuts 0° et 30°. La différence augmente ensuite pour les azimuts 60°, 90° et 120° où le groupe de sujets voyants a en moyenne une erreur de distance d'environ 95 mm, 105 mm et 134 mm respectivement et un écart-type d'environ 5 mm. Les non-voyants ont une erreur de pointage plus élevée dans la composante distance à ces 3 azimuts avec respectivement 111 mm, 123 mm et 156 mm et un écart-type d'environ 7 mm. La précision en distance est donc égale pour les deux groupes de sujets quand la source se trouve devant (entre 0° et 30° d'azimut). La différence augmente pour les autres valeurs d'azimut avec une différence d'erreur entre les deux groupes comprise entre 16 mm (azimut 60°) et 32 mm (azimut120°).

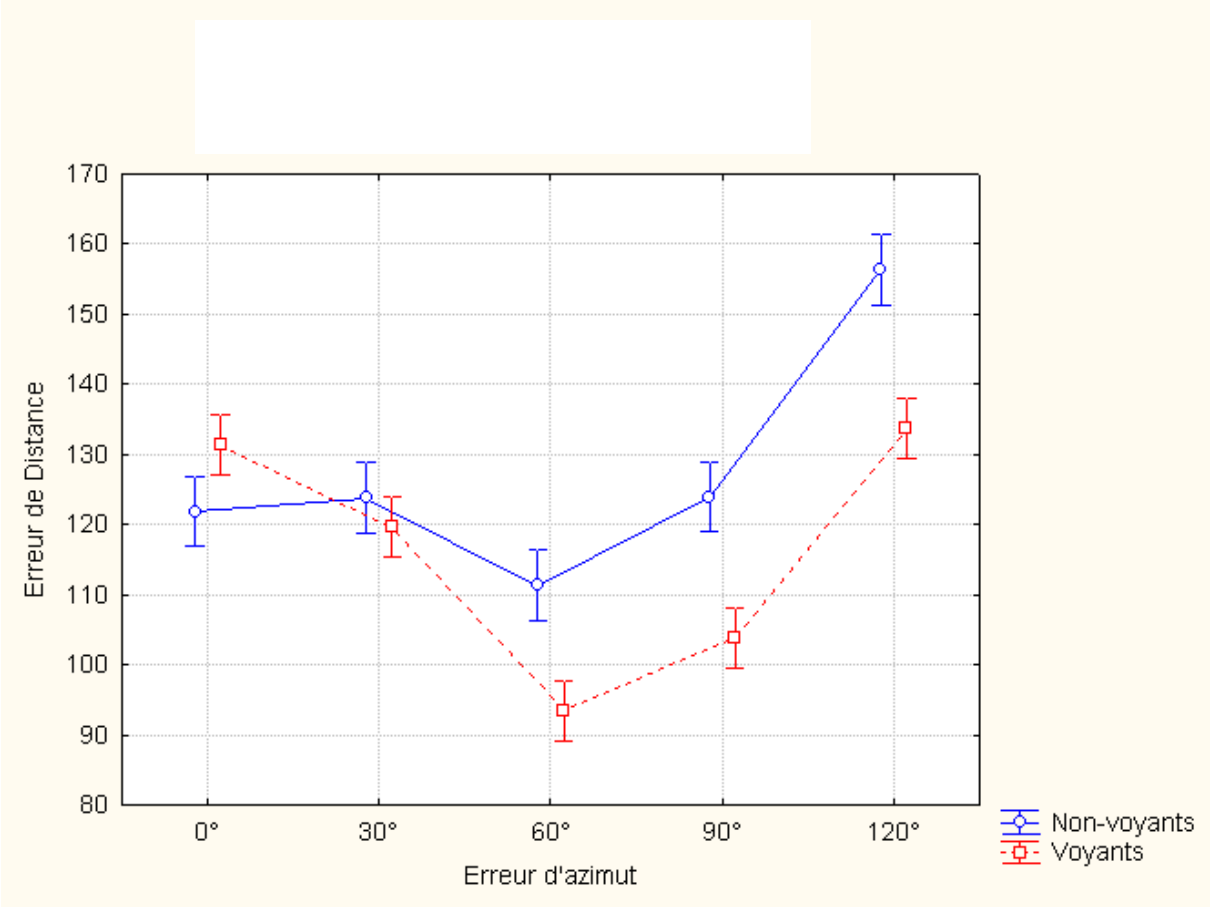


Figure 113: Moyenne de l'erreur de distance pointée en fonction de l'azimut de la source pour les deux groupes de sujets : voyants et non-voyants. (F(8,25696)=20,619 ; p=0,0000)

Nous retrouvons des différences dans la composante distance du pointage entre les deux groupes de sujets. En effet, après avoir comparé la différence d'erreur de distance en fonction de l'azimut de la source, nous allons comparer cette même mesure en fonction de la distance de la source sonore (Figure 114). L'erreur des sujets voyants et non-voyants est maximale pour la distance la plus proche (345 mm de distance, respectivement 107+-5 mm

et 147+-4 mm d'erreur). L'erreur de distance entre les deux groupes de sujets n'est pas significativement différente à 595 mm (109 mm d'erreur pour les deux groupes, un écart type de 5 mm pour les sujets non-voyants et 4 mm pour les sujets voyants). Entre ces deux distances, à 470 mm, l'écart entre les deux groupes se ressert linéairement. Pour les distances supérieurs à 595 mm, l'erreur de distance évolue de la même façon en augmentant jusqu'à 835 mm (147+-5 mm d'erreur pour les sujets non-voyants et 139+-5 mm d'erreur pour les sujets voyants). La moyenne de l'erreur de distance pointée pour chacun des groupes de sujets voyants et non-voyants est respectivement de 107+-4 mm et 125+-5 mm à 720 mm de distance et de 109+/-4 mm et 147+/-5 mm à 835 mm de distance.

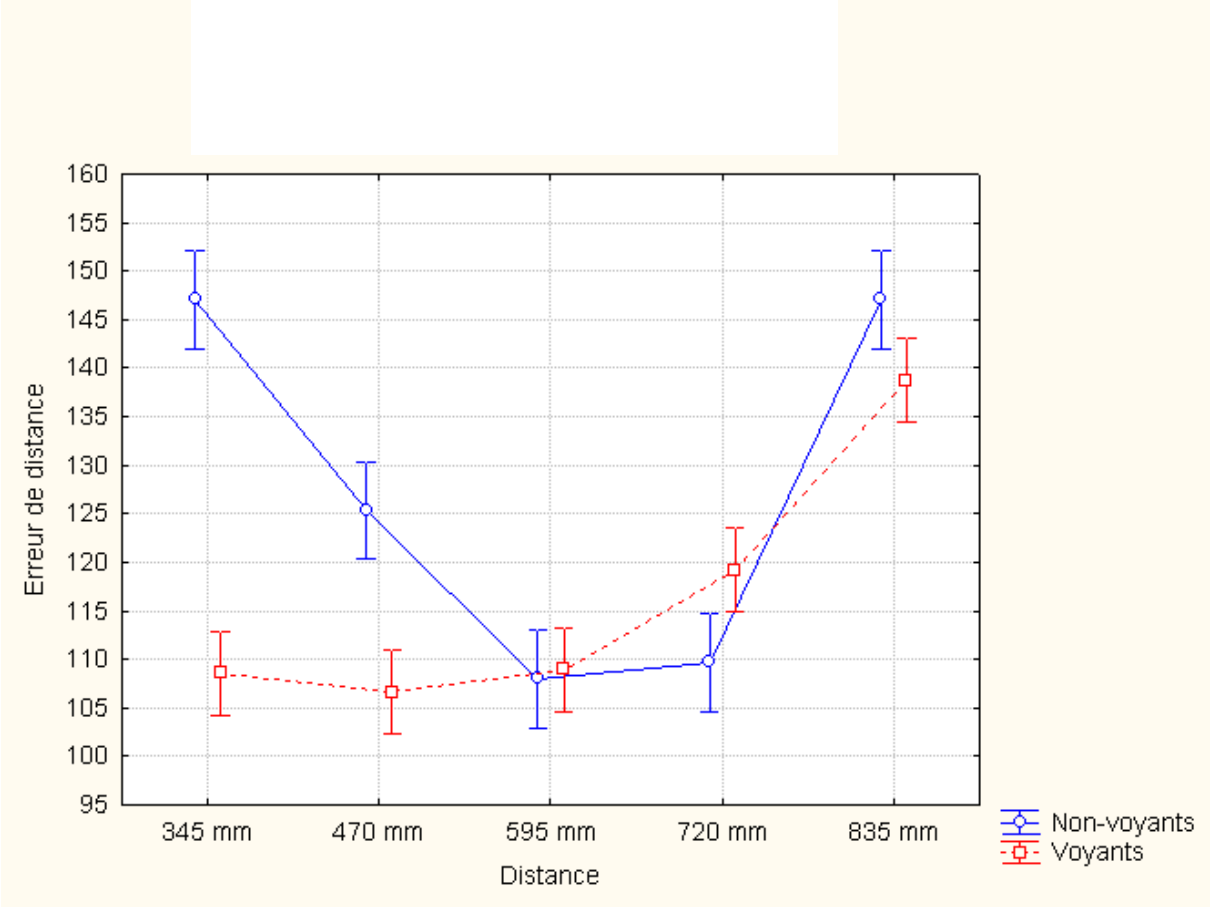


Figure 114: Moyenne des erreurs de pointage dans la composante distance de la source sonore pour les deux groupes de sujets voyants et non-voyants. (F(8,25696)=25,602 ; p=0,0000)

Les résultats précédents sur l'erreur de distance pointée en fonction de la position de la source sonore suivant les deux dimensions (azimut et distance) sont présentés dans la Figure 115. La précision de la distance de pointage est la meilleure autour d'un axe ayant pour extrémités les points (az=0° ; dist=345 mm) et (az= 90°;dist=720 mm) pour les non-voyants. Pour les voyants, le premier point se situe un peu plus loin (à une distance de 470 mm) pour

un même azimut. La zone du plan horizontal où l'erreur est inférieure à 200 mm se situe autour de cet axe et est plus large pour le groupe de voyants que le groupe de non-voyants. Ce résultat confirme la précision plus faible dans la distance de pointage du groupe de sujets non-voyants.

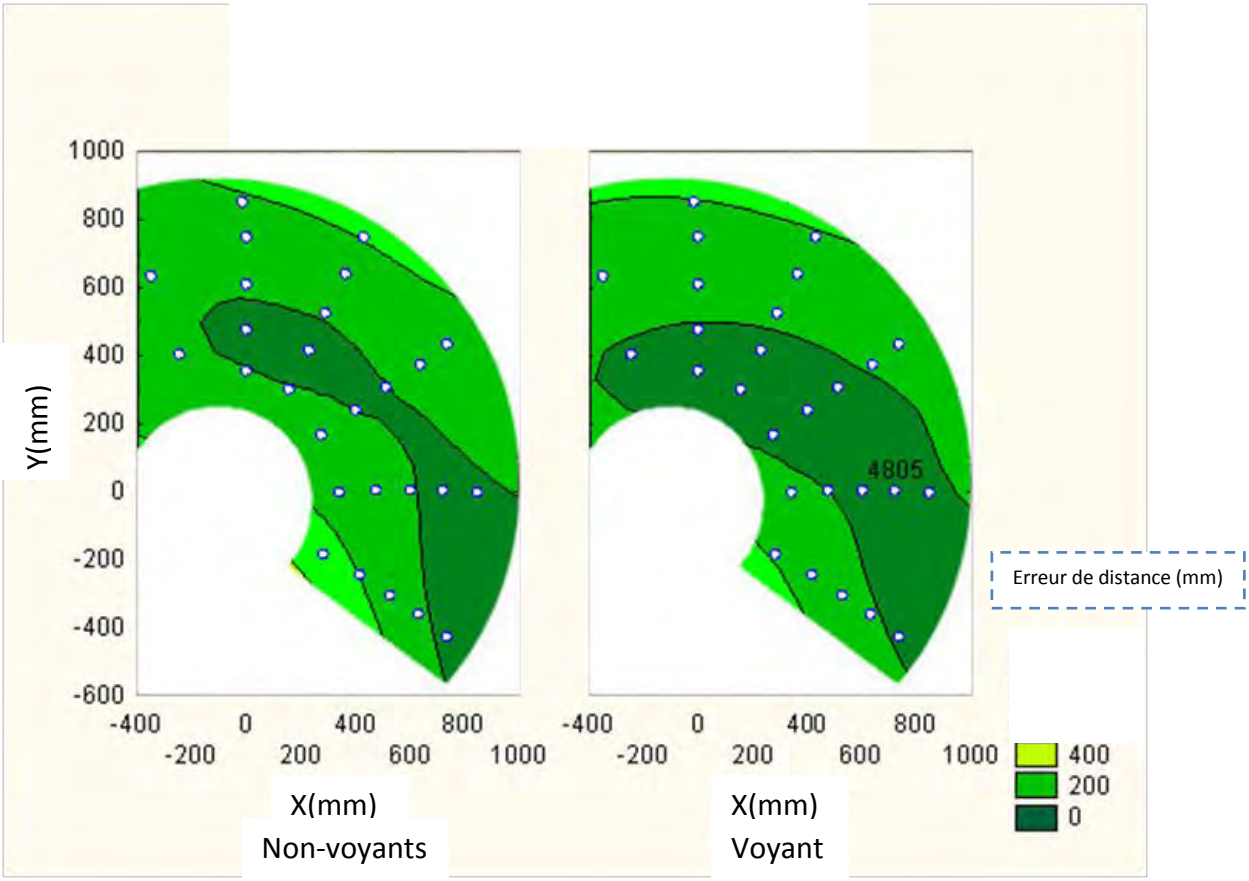


Figure 115: Moyenne des erreurs de distances pointées en fonction de la position de la source sonore sur le plateau pour les sujets non-voyants (à gauche) et voyants (à droite)

Discussion

La comparaison des performances de pointage chez les sujets voyants et non-voyants a permis de mettre en évidence des différences et des points communs pour les deux groupes de sujets. Pour les deux groupes, les deux conditions ayant engendré le plus faible pourcentage d'erreurs avant-arrière sont les conditions 5 (2*25 ms) et 6 (3*25 ms). Un résultat étonnant avait été développé dans la présentation des résultats moyennés chez tous les sujets : il semble y avoir une augmentation du nombre de ces erreurs au-delà de 3 sons de 25 ms. En effet, la condition 4*25ms fais significativement augmenter le pourcentage d'erreurs avant-arrière chez les deux groupes de sujets. La précision de

pointage angulaire chez les deux groupes de sujets évolue de la même manière quelle que soit l'azimut de la source : la précision diminue avec l'azimut de la source.

L'erreur de distance pointée fait apparaître des différences de performance de pointage entre les deux groupes de sujets. En effet, la précision du pointage dans la composante distance est très différente d'un groupe à l'autre en fonction de la position de la source. Les deux groupes ont les mêmes performances de pointage lorsque l'azimut de la source se situe proche de 0° mais l'erreur augmente ensuite avec l'azimut autant que la différence entre les deux groupes de sujets, comprise entre 16 mm de différence (azimut 60°) et 32 mm de différence (azimut120°). La différence de précision de pointage entre les deux groupes de sujets en fonction de la distance de la source est quasiment nulle pour des distances entre 345 mm et 595 mm. Au-delà, le groupe des sujets voyants obtient une précision moyenne plus élevée, avec une différence atteignant 18 mm à 120 mm de distance et 38 mm à 835 mm de distance. L'erreur de distance absolue mesure la distance entre la source et la position atteinte par le doigt du sujet. Elle reflète l'écart réel entre la source et là où le sujet la situe. La comparaison des sujets voyants et non-voyants pour cette mesure fait apparaître une zone devant le sujet où l'erreur de pointage est minimale (<120 mm en moyenne). Cette zone apparaît autour du point d'azimut 0° et de distance 470 mm pour les sujets voyants et pour le même azimut, autour du point à 340 mm de distance pour les sujets voyants. La zone de précision maximale s'avère être plus large chez les sujets voyants et évolue en moyenne comme l'erreur de distance présentée précédemment, l'erreur d'azimut étant presque négligeable dans le calcul de l'erreur absolue dans l'espace proche.

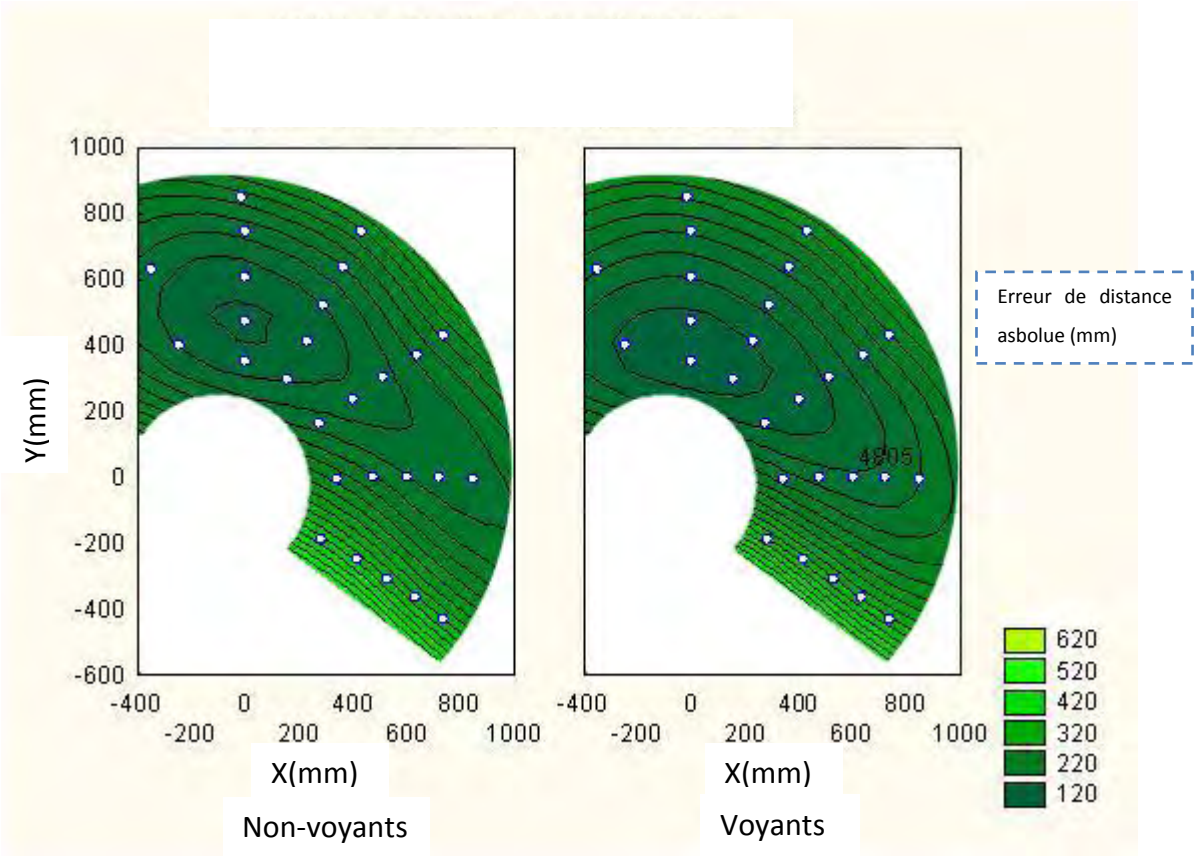


Figure 116: Moyenne de l'erreur en distance absolue en fonction de la position de la source sur le plan horizontal, pour les deux groupes de sujets : non-voyants (à gauche) et voyants (à droite)

La durée du son et le nombre de répétitions influent de la même façon pour les sujets voyants et non-voyants sur la précision dans les deux composantes de la position atteinte. La condition pour laquelle l'erreur est minimale dans les deux composantes est la condition 6 (3*25ms). C'est aussi la condition engendrant le moins d'erreur avant-arrière. Cette condition présente des erreurs de distance comprises entre 90+/-10 mm (azimut 60°) et 110+/-10 mm (azimut 120°) pour les voyants et entre 100+/-10 mm (azimut 60°) et 135+/-10 mm (azimut 120°) pour les non-voyants. L'erreur de distance est minimale pour les sujets voyants aux distances 345 mm, 470 mm, 595 mm avec une précision en distance de 109+/-10 mm. La précision pour cette condition des non-voyants est différente et atteint un minimum à 720 mm (102 mm d'erreur en moyenne). La précision atteint un maximum pour les sources éloignées (835 mm, 148+/-10 mm d'erreur) et les sources plus proches que 595+/-10 mm : l'erreur atteint 130+/-10 mm à 345 mm de distance. Finalement, la précision maximale (75 mm d'erreur de distance) est obtenue pour l'ensemble des sujets non-voyants à l'enceinte d'azimut 60° à 470 mm de distance. Pour les voyants, c'est l'enceinte située au

même azimut (60°) et à 595 mm de distance (60 mm d'erreur moyenne). L'erreur absolue dans cette condition reflète la distance absolue entre la position pointée et la source sonore. La précision dans cette mesure atteint un maximum avec 100+/-10 mm d'erreur moyenne chez les sujets voyants pour un azimut 60° et une distance de 720 mm. Cette zone de précision maximale est un peu plus proche pour les sujets non-voyants.

Discussion sur les résultats de l'expérience 2

Les résultats obtenus sur la localisation de sources sonores dans l'espace péripersonnel montrent qu'il est possible de localiser une source sonore dans l'espace proche et que de nombreux paramètres influents sur la précision du pointage. Les voyants et les non-voyants n'ont pas la même précision de localisation mais semblent réagir de la même façon aux différents paramètres. Nous avons dans les deux précédentes études établi qu'il était possible de donner une information spatiale pour aller atteindre une cible dans l'espace proche (<84cm de distance). Les non-voyants et les voyants ont dans l'ensemble été plus précis pour un stimulus composé de 3 sons de 25ms que dans les 6 autres conditions faisant varier le nombre et la durée des sons. Les sujets avaient la tête fixe quand le stimulus était joué et n'avaient pas le temps de bouger la tête pour se calibrer pendant le stimulus (durée totale du stimulus < 200ms). L'augmentation de la durée du son et le nombre de répétitions permettent d'augmenter la précision en azimut et en distance.

Dans le cadre de l'élaboration d'un outil de suppléance visuelle basé sur la localisation d'objets par la synthèse d'un son spatialisé, nous avons voulu établir quel était le stimulus le plus adéquat pour répondre à ce besoin. Pour ne pas surcharger l'utilisateur avec des sons qui seraient difficiles à localiser et qui auraient une durée trop longue pour ne pas interférer avec les sons environnants, notre objectif était de définir le son le plus efficace avec la durée la plus courte. Nous avons montré que 3 sons de 25ms entrecoupés de 30 ms de pause (135 ms au total) est le stimulus ayant permis la meilleure performance de localisation chez les sujets voyants et les sujets non-voyants. Avec ce stimulus, la précision maximale est atteinte pour les deux groupes de sujet à l'azimut 0° et à 470 mm de distance. La précision moyenne de pointage est alors de 100 mm pour les sujets voyants et 130 mm pour les sujets non-voyants. Les non-voyants ont, dans cette tâche, obtenu des résultats moins précis que les sujets voyants dans le pointage en distance.

L'objectif expérimental que nous poursuivons est de répliquer ces résultats obtenus avec des sons réels dans une interface de restitution utilisant la synthèse binaurale. Les résultats

obtenus serviront à la conception d'une interface de restitution pour un système de suppléance. L'équipe de Brian FG. Katz au Limsi à Orsay avec laquelle nous collaborons dans cette étude dispose d'une base de données de 160 HRTF enregistrées sur des sujets sains dans une salle anéchoique. Il a été montré qu'il est possible d'attribuer un filtre existant se rapprochant le plus possible de la réalité pour ensuite améliorer les capacités de localisation avec celui-ci par apprentissage (Blum et al., 2004). La tâche d'apprentissage consiste à déplacer un capteur autour de soi, un son virtuel étant alors produit comme s'il provenait du capteur lui-même. La mise en marche d'une boucle perception-action par l'utilisateur lui permet d'accroître ses capacités de localisation en s'habituant à ces paramètres. On ne sait en revanche pas s'il peut garder ces facultés apprises sans les stimuler régulièrement. Notre dispositif augmente l'information auditive réelle et n'enlève rien à la perception des bruits environnants. Une question persiste encore quant à la capacité à pouvoir discerner et s'adapter aux sons réels et ceux synthétisés en même temps, les facultés de localisation des sons synthétisés ayant été apprises. Il est donc possible de concevoir une interface de restitution par synthèse binaurale en temps réel pour restituer la position d'un objet dans l'espace.

Étude préliminaire : modélisation d'une neuroprothèse visuelle

Les neuroprothèses posent pour une large part les mêmes problématiques que les systèmes de suppléance non-invasifs. La résolution d'une image à restituer doit être de très haute résolution pour pouvoir être interprétable. Dans le cas d'une image de dimensions 320 pixels par 240 pixels, il faudrait pouvoir « allumer » environ 76000 points alors que les matrices de stimulation actuelles ne permettent l'évocation que d'une centaine de percepts (Figure 117). Comme pour les systèmes non-invasifs, il est nécessaire de filtrer l'information avant de la restituer. Pour cela, nous proposons d'utiliser le capteur de localisation d'objets et ainsi ne restituer qu'une seule information utile : la position de l'objet. L'objectif de cette étude est de modéliser le comportement d'une neuroprothèse destinée à la restauration de la position 3D des objets. Ce que nous voulons ainsi évaluer est le nombre minimum de phosphènes utiles sur la plus petite zone de cerveau implantée assurant cependant une restitution des informations de localisation rapide et précise.

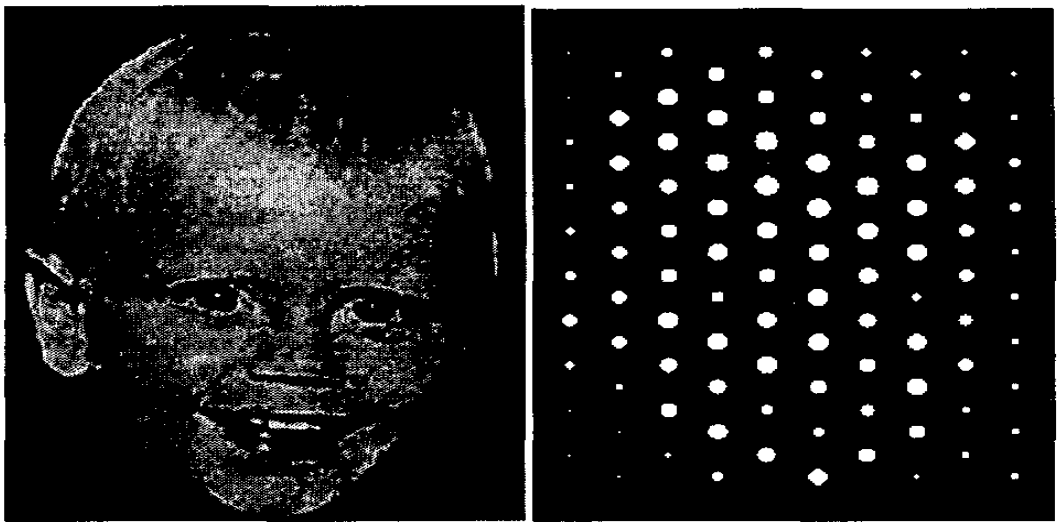


Figure 117: Image d'un petit garçon à gauche et son équivalent converti en une matrice de 100 phosphènes (correspondant à environ 100 électrodes).

Les recherches menées dans ce cadre sont relativement exploratoires. En effet, nous ne disposons que de très peu d'informations précises sur la position des phosphènes et sur leur aspect en fonction des paramètres de stimulation. De plus certaines études (Brindley, 1982;Brindley and Lewin, 1968;Dobelle et al., 1974;Dobelle, 2000b;Schmidt et al., 1996;Srivastava et al., 2007) montrent que la relation entre le nombre d'électrodes stimulées et le nombre de phosphènes évoqués n'est pas linéaire. Une électrode peut générer un ou plusieurs phosphènes ou bien encore aucun.

En choisissant la voie de la modélisation informatique, nous nous donnons la possibilité de simuler ce dispositif en manipulant librement tous les paramètres issus de ces observations. Grâce à cette approche, nous avons pu chercher à déterminer le nombre d'électrodes nécessaire et suffisant pour reconnaître et localiser un objet dans l'espace. Cette étude a été effectuée en collaboration avec Marc Macé Post-Doctorant de l'équipe dans laquelle je travaille ainsi que Yannick Adelise que j'ai co-encadré avec Marc Macé pendant son stage de M2 neurosciences.

1) Matériel et méthodes :

La stimulation du système visuel humain est étudiée dans trois principaux sites de stimulation : la rétine, le nerf optique et le cortex visuel (Figure 33). La littérature montre qu'une stimulation électrique du système visuel fait apparaître des percepts visuels (points lumineux appelés phosphènes. Nous avons mené une expérience de psychophysique pour évaluer la possibilité de saisir des objets sur la base de l'apparition de phosphènes restituant

la position d'un objet. Nous voulons ainsi donner une information de position 3D dans un casque de réalité virtuelle en simulant l'apparition de percepts au plus proche de ceux qui auraient pu être générés par une stimulation nerveuse du système visuel. Les percepts lumineux observés dans la littérature (Brindley, 1982;Schmidt et al., 1996) ont la particularité d'être invariants au mouvement des yeux. Afin de simuler l'apparition de ces percepts dans le casque de réalité virtuelle, nous utilisons un outil de suivi du regard (ASL 6000) permettant de suivre la position de l'œil et de mettre à jour en temps réel la position du percept en fonction de la position de l'œil. Durant cette expérimentation exploratoire, pour des raisons techniques, il était demandé aux sujets de fixer le regard droit devant eux et l'outil de suivi du regard n'était utilisé ici que pour contrôler la position des yeux.

La faible résolution de l'interface de restitution rend la problématique très proche de celle décrite précédemment pour une restitution auditive. Il est en effet essentiel de filtrer l'information pour ne restituer qu'une information pertinente. Dans cette étude, nous mesurons le temps d'atteinte d'une cible dans un monde virtuel. Nous avons choisi le temps comme unité de mesure afin d'évaluer de manière objective et précise l'évolution du comportement visuo-moteur du sujet dans la réalisation de cette tâche.

Le matériel utilisé pour notre expérience était composé des éléments suivants :

-Un casque de réalité virtuel Nvisor SX, de résolution 1280*1024.

-Un ordinateur Dell Latitude E5500.

Figure 118: 1- Écran de contrôle de la scène virtuelle pour l'expérimentateur. 2- Écran de contrôle du rendu du modèle pour l'expérimentateur. 3- Casque de réalité virtuelle porté par le sujet.

Dans l'expérience réalisée, le sujet pouvait naviguer dans l'espace virtuel à l'aide des 4 flèches de direction d'un clavier, ces touches permettaient d'avancer, de reculer et de tourner à gauche et à droite. Pour cette première étude, nous avons choisi d'évaluer trois cartographies de phosphènes constituées de 9, 100 et 400 phosphènes possibles. Ces phosphènes étaient répartis selon une grille déformée rendant compte de la variabilité sur leur position observée dans les études de cartographie des phosphènes (Schmidt et al., 1996). L'expansion corticale est modélisée par des percepts dont la taille augmente avec leur excentricité : les phosphènes dans la région fovéale seront les plus petits. Des tests préliminaires nous ont permis de sélectionner un nombre approprié de phosphènes affichés simultanément à l'écran. Ce nombre de phosphènes correspond théoriquement au nombre d'électrodes stimulées simultanément afin de générer plusieurs phosphènes. Nous avons choisi d'afficher pour chaque cartographie un ou cinq percepts (phosphènes).

De plus, nous avons ajouté à notre condition de référence (1 phosphène parmi 9 phosphènes possibles) deux facteurs supplémentaires pour indiquer la distance de l'objet sous la forme d'un clignotement du phosphène ou d'une variation de sa luminance. Finalement, la position de l'objet sera décrite suivant 2 paramètres : sa position angulaire (azimut et élévation) sera caractérisée par la position du phosphène et sa distance par la fréquence de clignotement ou sa luminance (Figure 119).

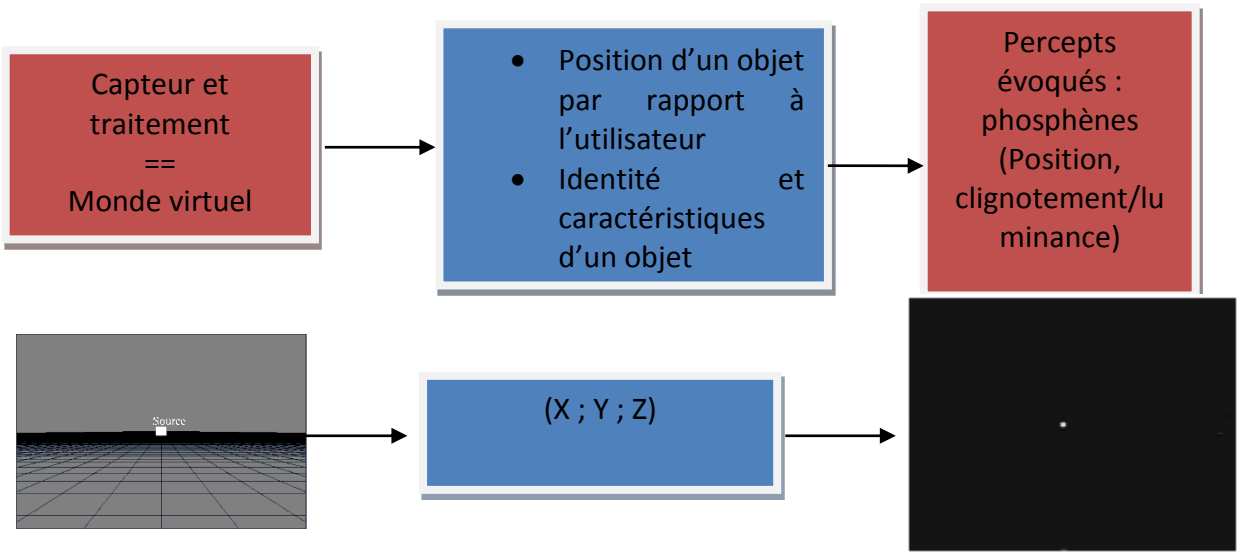


Figure 119: Informations disponibles à la restitution dans le dispositif de localisation de cibles visuelles. Seule l'information de localisation est transmise pour l'affichage.

Protocole expérimental

Nous avons voulu ici tester de manière isolée la restitution d'une information de localisation par l'évocation de phosphènes. Afin de n'évaluer que le mode de restitution, nous nous sommes placés dans une situation contrôlée simulant un capteur de restitution de la position des objets parfait : le déplacement dans un monde virtuel. Pour cela, l'utilisateur se déplaçait virtuellement dans un monde dans lequel il fallait saisir des objets virtuels qui n'étaient indiqués que par des phosphènes. Le champ virtuel de la caméra était de 60° et la caméra était située à 1,50 mètre de hauteur.

Un objet était placé aléatoirement à entre 12 mètres et 24 mètres de distance de l'utilisateur. De plus, l'objet pouvait être positionné à une hauteur aléatoire comprise entre 0 et 2,50 m. Il n'y avait aucun obstacle entre le sujet et l'objet. Lorsque qu'un nouvel objet virtuel était positionné, l'utilisateur devait naviguer avec les flèches directionnelles dans l'espace virtuel afin de l'atteindre (Figure 120).

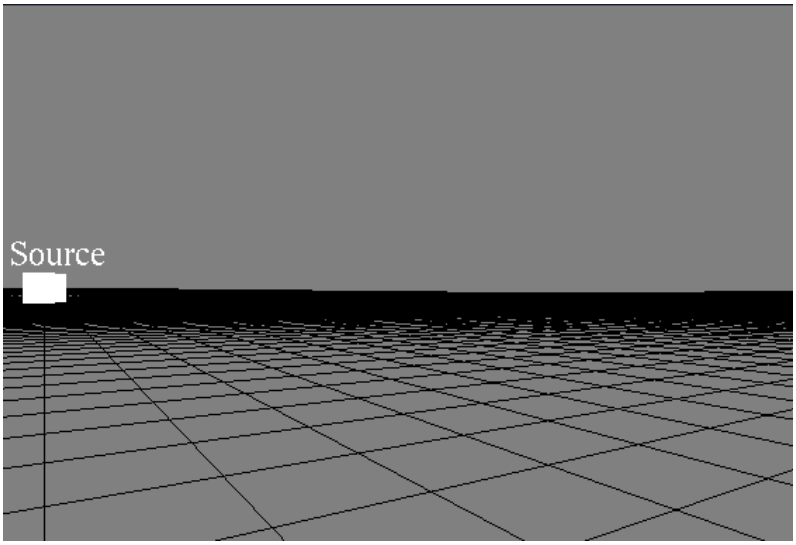


Figure 120: Monde virtuel simulant un utilisateur portant une caméra tridimensionnelle sur la tête, permettant de localiser des objets dans un champ visuel de 60°. L'utilisateur évoluait dans cet environnement avec les flèches directionnelles.

Les coordonnées en trois dimensions de l'objet (X, Y, Z) relativement à l'utilisateur sont envoyées au module de restitution pour la génération des phosphènes. Les sujets évoluaient dans l'environnement virtuel dans un casque de réalité virtuelle grand angle et ne voyaient que des phosphènes.

Paramètres de l'expérience

Le modèle est caractérisé par la fonction de transformation entre une stimulation électrique du système nerveux et le percept évoqué en réponse. Cette fonction est très complexe à établir et de nombreux défis sont encore à relever dans ce domaine pour connaître avec précision les paramètres de stimulation à utiliser pour évoquer un percept précis. La littérature en revanche propose de nombreuses règles basées sur l'observation du fonctionnement du système visuel humain et c'est ce que nous avons tenté de modéliser. N'ayant pas de cartographie précise à disposition, certains choix ont été faits dans cette expérience préliminaire. Nous avons émis l'hypothèse d'une bijection entre le nombre d'électrodes stimulées et le nombre de percepts évoqués même s'il a parfois été observé des comportements différents (Schmidt et al., 1996). Nous avons considéré que les phosphènes sont topiques, avec une précision toutefois limitée. Une électrode stimulée ne donne pas nécessairement un phosphène à l'emplacement attendu (Brindley and Lewin, 1968;Dagnelie et al., 2003;Dobelle, 2000b;Srivastava et al., 2007). Le modèle développé est entièrement paramétrable et repose sur l'étude de la littérature en microstimulation du système visuel humain. Afin de définir les paramètres d'affichage des phosphènes simulés,

autrement dit la taille, la forme, le clignotement, la variation de contraste, le champ visuel, nous avons développé une interface nous permettant d'interagir avec le modèle (Figure 121).

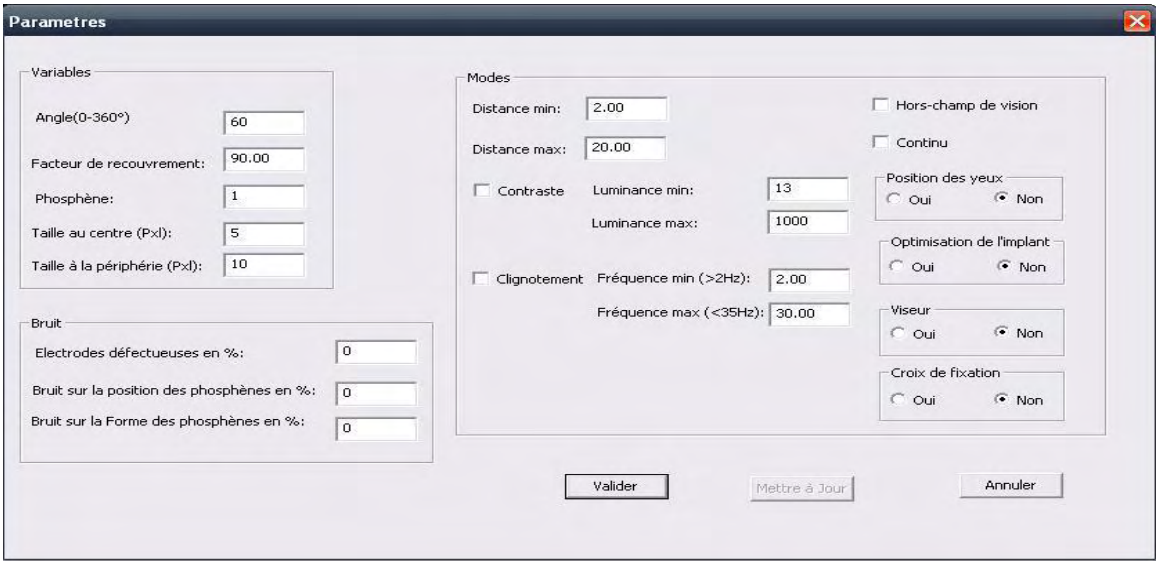


Figure 121: Interface de paramétrisation des phosphènes: il est possible de générer différentes configurations de phosphènes correspondant à différentes implantations d'interfaces cerveau-machine.

Paramètres du modèle

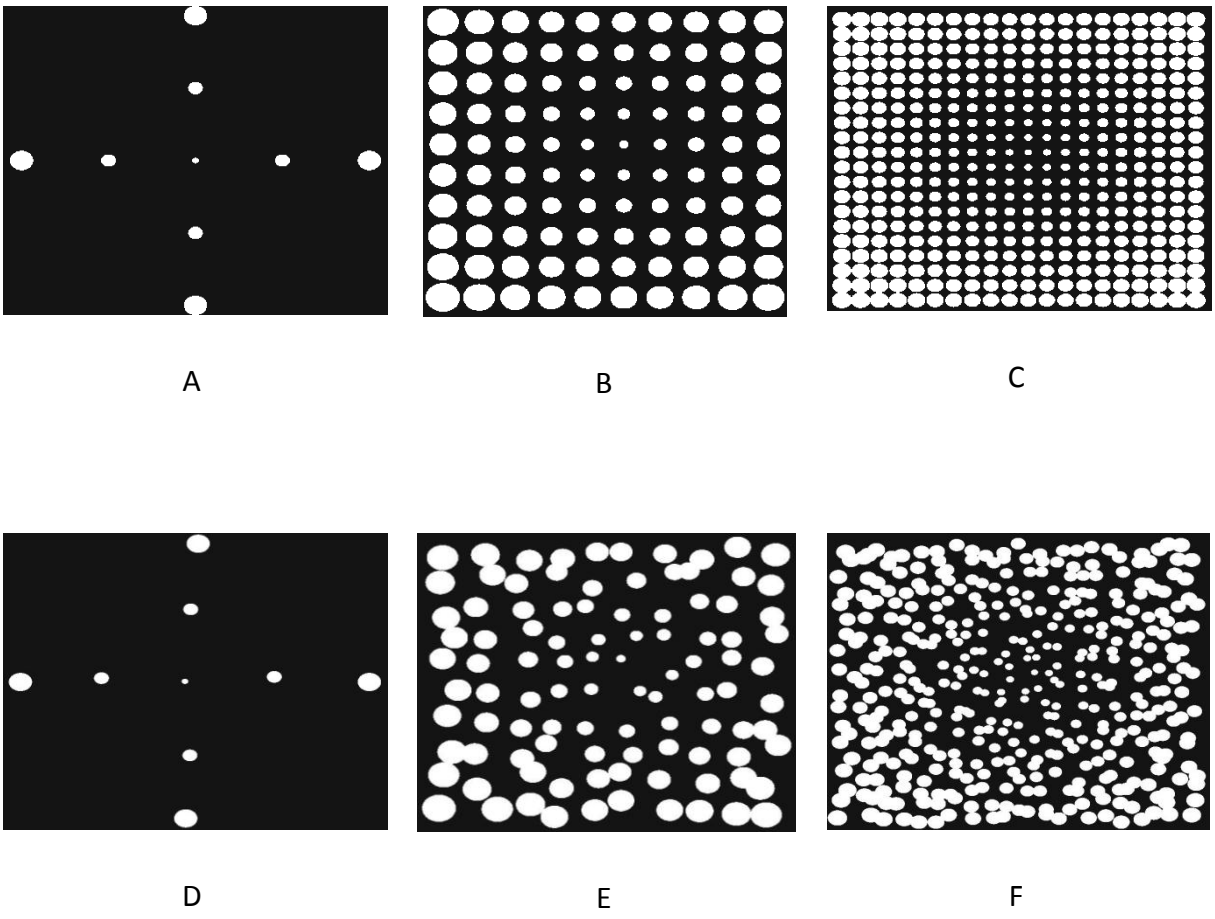
L'étude de la littérature nous a permis d'établir des règles et des restrictions pour l'affichage des phosphènes dans le casque de réalité virtuelle :

- L'angle de vue possible à restituer (il est inférieur ou égal à celui de la caméra).
- Nombre de phosphènes qu'il est possible d'afficher sur l'écran simultanément
- Nombre de phosphènes possible à afficher
- Taille des phosphènes dans la région fovéale et périphérique
- Pourcentage d'électrodes périphériques
- Bruit sur la position et la forme des phosphènes
- Distances minimale et maximale que l'on peut restituer sur la base de modulation de la luminance ou du clignotement, la valeur de la luminance et la fréquence de clignotement pour les distances minimales et maximales à restituer.

- Simulation ou non de la rétinotopie des percepts (un outil de suivi du regard ASL 6000 permet de connaître la position des yeux à une fréquence de 240 Hz).

Nous avons voulu dans cette expérience évaluer le temps pour atteindre une cible en fonction de trois paramètres : le nombre de phosphènes, leurs positions et le bruit dégradant la position des percepts. Pour cela, nous avons étudié trois cartographies de phosphènes avec 9, 100 et 400 phosphènes possibles avec ou sans bruit pour rendre compte de la variabilité des phosphènes observée dans la littérature (Figure 122).

1) La première cartographie était composée de neuf électrodes positionnées de façon à représenter les directions cardinales (A et D).
2) La seconde était composée de 100 électrodes placées de façon à former une matrice carrée d'électrodes (B et E).
3) La troisième était composée de 400 électrodes placées de façon à former une matrice carrée d'électrodes (C et F).



4) **Figure 122 : Cartographies des phosphènes utilisés dans le prototype de simulation de la neuroprothèse. Ces cartes sont constituées de 9, 100 et 400 phosphènes disposés régulièrement (A, B & C) ou de manière pseudo-aléatoire (D, E & F).**

Stimuli et variables étudiés

Le temps d'atteinte de la cible a été mesuré en faisant varier le nombre de phosphènes affichés simultanément : 1 ou 5 pour chaque cartographie. Les phosphènes pouvaient clignoter ou changer de luminance pour restituer la distance à l'objet. La première cartographie (9 positions de phosphènes possibles) a été testée dans quatre conditions différentes :

1. 1 Phosphène seul phosphène affiché sans notion de distance
2. 1 Phosphène seul phosphène affiché avec notion de distance (Contraste)
3. 1 Phosphène seul phosphène affiché avec notion de distance (Fréquence de clignotement)
4. 5 Phosphènes simultanément affiché sans notion de distance

Les deux cartographies suivantes (100 et 400 phosphènes) ont été testées dans deux conditions différentes chacune (1 ou 5 phosphènes affichées simultanément).

Le comportement de chaque sujet était étudié dans les 16 séries (deux fois les 8 conditions) de 25 essais chacune.

- Condition 1 : 9 positions de phosphènes possibles, 1 seul phosphène affiché (9-1 standard)
- Condition 2 : 9 positions de phosphènes possibles, 1 seul phosphène affiché avec variation de contraste en fonction de la distance (9-1 cont)
- Condition 3 : 9 positions de phosphènes possibles, 1 seul phosphène affiché avec variation de la fréquence de clignotement en fonction de la distance (9-1 clign)
- Condition 4 : 9 positions de phosphènes possibles, 5 phosphènes affichés simultanément (9-5 standard)
- Condition 5 : 100 positions de phosphènes possibles, 1 seul phosphène affiché (100-1 standard)
- Condition 6 : 100 positions de phosphènes possibles, 5 phosphènes affichés simultanément (100-5 standard)
- Condition 7 : 400 positions de phosphènes possibles, 1 seul phosphène affiché (400-1 standard)
- Condition 8 : 400 positions de phosphènes possibles, 5 phosphènes affichés simultanément (400-5 standard)

L'expérience a été menée auprès de 8 sujets, une phase d'apprentissage de 25 essais était réalisée avec le dispositif dans une condition avec 5 phosphènes (1 à chaque cardinalité et 1 au centre) et un seul phosphène affiché à un instant donné. Une fois cette phase d'apprentissage effectuée, chaque sujet démarrait la 1ère série d'évaluation jusqu'à la 16ème en s'arrêtant entre chaque série.

Huit sujets (2 femmes / 6 hommes), âgés de 25 à 45 ans ont participé à cette expérience. Les sujets avaient tous une vue normale ou corrigée à la normale et étaient habitués à utiliser un clavier ordinateur au quotidien.

2) Résultats

La Figure 123 affiche les temps d'atteinte de cibles en fonction du nombre de phosphènes (9, 100, 400) utilisés en n'affichant qu'un seul phosphène à la fois.

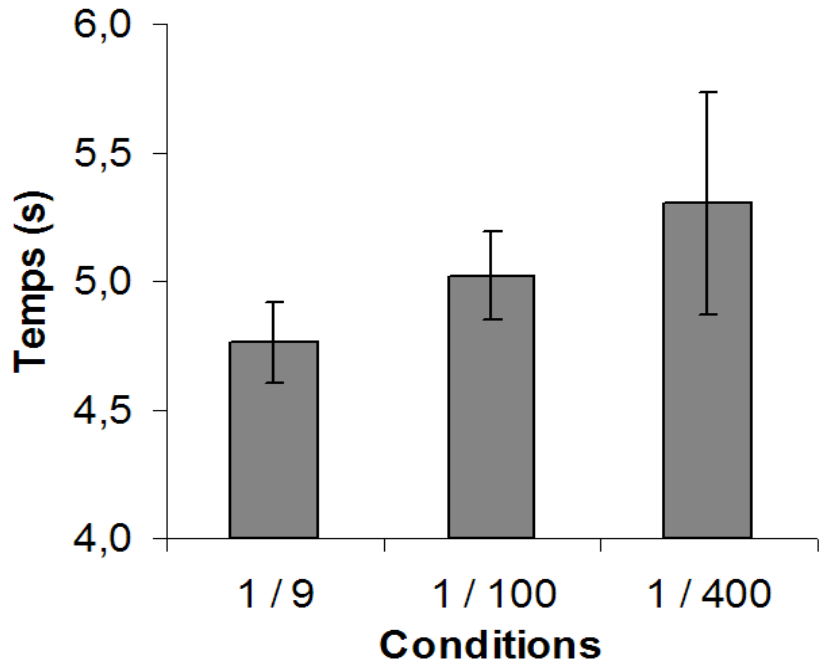


Figure 123: Temps moyen pour saisir un objet dans un environnement virtuel avec un phosphène allumé parmi 9, 100 ou 400 emplacements de phosphènes possibles.

Le nombre de phosphènes pouvant être affichés pour restituer une information de position d'un objet virtuel n'est pas meilleur avec un nombre élevé de percepts. En effet, le temps de saisie avec 9 phosphènes (4,75 s) n'est pas supérieur aux deux autres conditions (5 s et 5,75 s). Une tendance inverse semble même se dégager. Dans le cas de notre tâche exploratoire, la vitesse de déplacement jusqu'aux objets virtuels était fixe (les sujets pouvaient s'arrêter ou avancer), ce qui rend les temps d'atteinte très proches. Ce résultat nous conforte dans l'idée que seuls quelques phosphènes suffisent à restituer l'information de position. Comme le montre la Figure 124, la perception de distance n'améliore pas les temps d'atteinte de la cible virtuelle. En effet, les sujets savaient qu'ils pouvaient se déplacer à vitesse maximale jusqu'à la cible sans rencontrer d'obstacles. La condition clignotante en revanche ralentit les sujets.

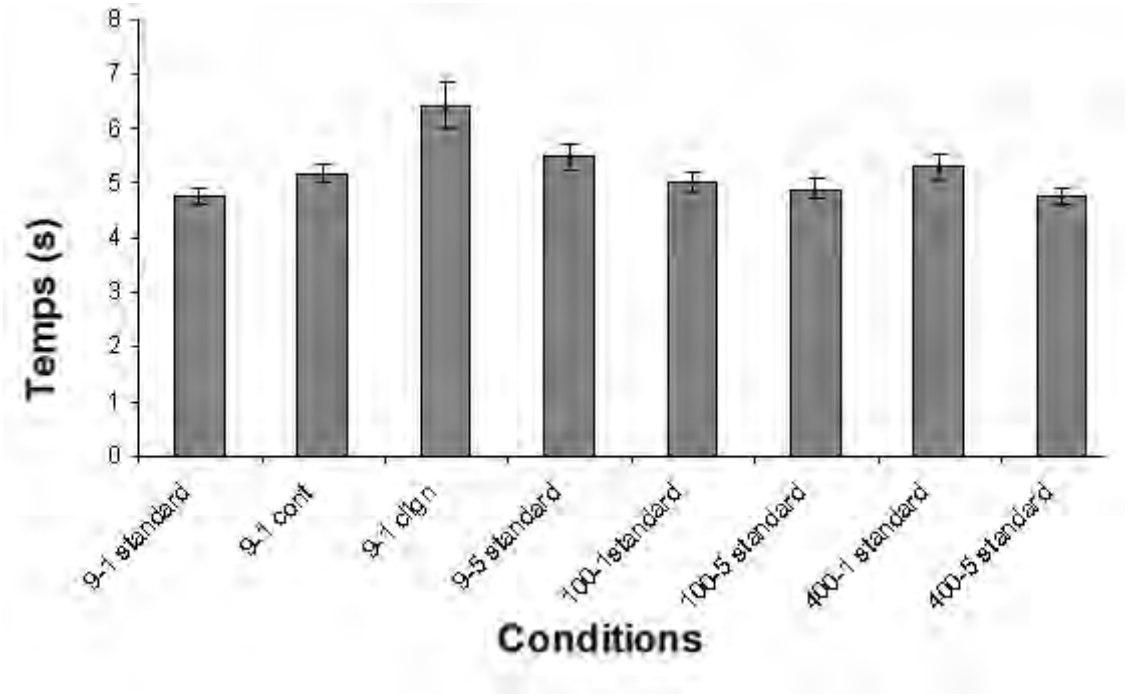


Figure 124 : Temps moyen (s) mis par l'ensemble des sujets pour localiser et attraper l'objet dans chaque condition (1-8)

Le temps moyen mis par les sujets dans la condition 9-1 cont (condition 2) a été de 5,16 s et de 6,41 s dans la condition 9-1 clign (condition 3).

Le balayage angulaire (mouvements de balayage de la tête) est plus que doublé dans la condition avec le clignotement car les utilisateurs avaient tendance à tourner rapidement pour chercher le percept qui pouvait facilement être manqué pendant la période où le percept disparaissait. Ce résultat explique pourquoi le temps moyen mis par les sujets dans la tâche avec clignotement est plus élevé que dans l'ensemble des autres tâches. La nécessité d'une perception de distance pourrait être efficace dans des tâches où l'utilisateur aurait des choix d'itinéraire à effectuer en fonction de l'objet à percevoir, avec des obstacles sur l'itinéraire. En effet, le fait que les sujets avaient à atteindre la cible sans se soucier de la distance entre eux et l'objet revenait pour eux à avancer jusqu'à l'atteindre. Ce protocole n'a donc pas permis de mettre en évidence la nécessité d'une perception de distances.

3) Discussion

Analyse des performances des sujets sans la variation de contraste et le clignotement

D'après les résultats présentés dans la Figure 124, nous constatons que les sujets ont obtenu des performances similaires dans les conditions 9-1 standard et 400-5 standard (différence non significative avec une ANOVA, $p<0,05$). De plus, les sujets ont été plus rapides dans la condition 9-1 standard (4,76 s) que dans les conditions 9-5 standard (5,48) et 400-1 standard (5,30) (ANOVA, $p < 0,05$). En effet, dans la condition 9-5 standard, les phosphènes affichés étaient trop nombreux par rapport à la cartographie. Il suffisait que le sujet tourne la tête dans l'environnement virtuel pour voir la majeure partie des phosphènes se déplacer vers la périphérie. Les résultats montrent donc qu'il n'est pas indispensable d'afficher plus d'un phosphène afin de réaliser cette tâche. On peut en déduire qu'il n'est pas nécessaire d'utiliser un grand nombre d'électrodes et d'évoquer plus d'un percept en simultané pour une tâche d'orientation et d'atteinte de cible.

Analyse des performances des sujets avec la variation de contraste et le clignotement

Contrairement à nos attentes, la variation du contraste et le clignotement n'ont pas permis aux sujets de réaliser la tâche plus rapidement. En effet, l'ensemble des sujets a réalisé un temps moyen de 5,16 s avec la variation de contraste et de 6,41 s avec le clignotement alors qu'avec 9 électrodes en standard le temps moyen est de 4,75 s. La tâche qui était d'atteindre une cible sans obstacles ne faisait pas intervenir la notion de distance puisqu'il suffisait de continuer à marcher jusqu'à atteindre la cible. Toute information supplémentaire inutile pour la tâche (comme l'information de distance par clignotement ou variation de contraste) a pu augmenter le temps d'atteinte de la cible. Le clignotement a été rapporté comme étant très gênant par les sujets. En effet, pour réaliser le clignotement nous avons dû prendre en compte la limite de perception de ce type d'effet visuel chez l'homme qui se situe aux alentours de 30 Hz. Les fréquences utilisées dans l'expérimentation étaient ainsi comprises entre 2 Hz et 30 Hz, la fréquence la plus faible (2 Hz) représentait une distance élevée de l'objet. Par conséquent, les sujets effectuaient parfois plusieurs tours sur eux-mêmes avant de percevoir l'objet, ce qui a pu entraîner un temps moyen élevé dans cette condition.

4) Conclusion

Ces résultats préliminaires sont encourageants puisqu'ils montrent que le nombre de positions possibles pour afficher des phosphènes n'est pas déterminant dans la facilité à localiser et se diriger vers une cible. Le résultat semble plutôt contre-intuitif, mais il reflète bien l'idée qu'une quantité d'information limitée mais pertinente peut s'avérer suffisante pour réaliser une tâche. Ces premières données collectées vont nous guider pour les futures étapes de conception et d'évaluation du modèle de neuroprothèse visuelle. Des expérimentations sont actuellement en cours en collaboration avec un Post-Doctorant de l'équipe Marc Macé et un stagiaire de M2 Recherche en IHM, Valérian Guivarch. L'objectif est de coupler le capteur de reconnaissance et de localisation d'objet développé dans ce manuscrit et le modèle de neuroprothèse. L'objectif est de confirmer les résultats de localisation déjà obtenus avec ce modèle mais cette fois ci en situation très contrainte. Les travaux en cours chercheront aussi à évaluer de nouveau la pertinence de restituer une information de distance dans un protocole où elle est nécessaire. Pour cela, des expériences de psychophysique de saisie d'objets comparables à celles effectuées avec l'interface sonore sont en cours pour évaluer la capacité de sujets voyants équipés du casque de réalité virtuelle à aller saisir des objets.

Figure 125: Un sujet voyant équipé du modèle de neuroprothèse va saisir une souris sur la table sur la base de phosphènes affichés dans le casque de réalité virtuelle. La position de l'objet reconnu est envoyée au modèle par l'outil de reconnaissance et de localisation d'objets développé dans ce manuscrit.

Les résultats et observations obtenus dans cette série expérimentale auront un impact important sur le développement de la neuroprothèse. En faisant varier différentes la fonctions de transformation de la position d'un objet et sa restitution au cours de ces tâches de navigation spatiale, nous pourrons déterminer les paramètres d'affichage les plus importants pour guider une personne dans une tâche de navigation mais aussi une tache de saisie d'objets dans l'espace péripersonnel. Nous aurons alors défini une interface optimale pour la restitution de la position d'un objet pour notre modèle de neuroprothèse.

Discussion générale

Les systèmes de suppléance pour les non-voyants sont aujourd'hui en très forte progression grâce à deux principaux facteurs : la volonté politique d'allouer les financements nécessaires mais aussi la démocratisation de calculateurs de plus en plus puissants pour le grand public. Cette prise de conscience assez récente de la condition de vie des personnes en situation de

handicap a permis de faire de grandes avancées pour l'accessibilité des infrastructures. Il y a deux manières de s'intéresser à l'autonomie des non-voyants : en équipant les objets et le mobilier urbain d'outils d'accessibilité (mobilier parlant, écritures braille, …) ou en créant des systèmes permettant de rendre accessible la navigation non-visuelle. Ces derniers permettent une utilisation dans n'importe quel environnement mais aussi une maintenance ciblée des appareils contrairement à l'équipement du mobilier urbain qui nécessite une attention omniprésente et coûteuse. Je me suis intéressé dans cette thèse aux systèmes de suppléance en tant qu'outil pour répondre à un besoin créé par une déficience et non comme outil permettant de restaurer la vision.

5) Substitution sensorielle ou augmentation sensorielle : restauration sensorielle ou restauration fonctionnelle ?

L'étude des systèmes de substitution sensorielle permet de mesurer les limites des modalités sensorielles à interpréter des signaux destinés normalement à une autre modalité –ici la vision-. L'étude de ces systèmes permet d'établir la sensibilité de chaque modalité sensorielle, sa résolution et les informations qu'il est possible d'y faire transiter. Ils sont utilisés aujourd'hui pour étudier un mode de restitution mais ne s'intéressent pas aux phases amont du traitement du signal. Finalement, le fait d'évaluer ces systèmes en reconnaissant des objets noirs sur des fonds blancs revient à ne pas utiliser de caméra mais simuler l'apparition virtuelle d'une forme sur une matrice électro-tactile et d'étudier la capacité des sujets à reconnaître cette forme, dans différentes orientations et à différentes échelles. La faible résolution en entrée de ces systèmes les rend inopérants pour reconnaître les formes dans des environnements complexes. Nous verrons dans la suite de la discussion combien les résultats fondamentaux issus de l'étude de ces systèmes sont importants pour la conception des systèmes de suppléance pour les non-voyants.

Les systèmes d'augmentation sensorielle pour la navigation et la localisation prennent comme point de départ le besoin utilisateur. La conception d'un tel système passe en premier lieu par l'étude de ce besoin. Une manière d'analyser ce besoin est d'étudier de quelle manière le système visuel d'une personne voyante y répond pour isoler les fonctions particulièrement utiles à ce besoin. Il n'est en aucun cas ici question de restaurer la vision. Le fait d'étudier cette modalité sensorielle pour répondre aux déficiences est une méthode de travail pour étudier son utilité mais la caméra vidéo ne joue pas le rôle d'un œil : c'est un

capteur qui, couplé à des algorithmes, permet d'extraire **des informations utiles** dans la scène visuelle.

6) Choix du capteur et des algorithmes de traitement du signal

Une fois le besoin utilisateur défini, tout capteur couplé à des algorithmes de traitement adaptés répondant au besoin utilisateur est acceptable. La réflexion doit se situer au niveau fonctionnel du capteur et pas à son analogie avec une quelconque modalité sensorielle manquante. L'étude des fonctions du système visuel humain permet en outre de comprendre comment un système sensoriel déficient engendre un handicap et comment créer des systèmes d'aide adaptés. Les recherches en neuroscience sur le système visuel humain font apparaître qu'il est capable de localiser et reconnaître des objets de manière immédiate. A partir de ce constat, nous avons étudié l'influence de cette fonction du système visuel dans la navigation pour les sujets voyants et nous sommes aperçus que cette fonction particulière serait l'une des plus utiles à restaurer pour les personnes non-voyantes. Il existe une multitude d'installations permettant de reconnaître les objets et de les localiser : en pré-équipant les objets d'identifiants sans fil pour les détecter (RFID, patterns visuels à coller sur les objets, …). Une autre approche plus générique mais aussi beaucoup plus complexe a été choisie dans ce manuscrit : l'utilisation d'un capteur de vision artificielle pour localiser et reconnaître les objets. Ce capteur a l'avantage de pouvoir être utilisé dans n'importe quel environnement, même si le mobilier n'est pas pré-équipé. La principale difficulté de l'utilisation d'un tel capteur est le traitement d'un signal bruité et très complexe. L'étude d'une approche pour reconnaître et localiser des objets dans une scène visuelle décrite dans ce manuscrit montre qu'il est maintenant possible, d'analyser en temps réel une scène visuelle complexe. Aucun des algorithmes de vision existant ne permet d'analyser le signal de manière fiable dans toutes les situations mais chacun apporte une manière différente de traiter les informations visuelles et est adapté à des situations précises. Les méthodes faiblement tolérantes à l'orientation et au changement d'échelle s'avèrent très rapides et permettent d'analyser exhaustivement les objets visuels présents dans la scène. Ces méthodes sont en revanche moins adaptées aux recherches d'objets dans toutes orientations et toutes échelles. Pour cela, les méthodes définissant un objet par les relations spatiales reliant ses points d'intérêt à ceux d'un modèle à reconnaître s'avèrent plus adaptées. De ce constat, nous avons conçus un système capable de localiser et reconnaître des objets en essayant de restreindre au maximum les changements d'orientation et d'échelle. Nous montrons qu'il est possible d'utiliser l'algorithme de vision

Spikenet, couplé à une caméra binoculaire portée sur la tête pour reconnaître des amers visuels et aider les personnes non-voyantes à naviguer et à reconnaître des objets. Dans une version future du dispositif, un couplage de ces différents algorithmes en fonction du type d'objet recherché et son environnement permettrait d'accroître les situations dans lesquelles notre système de reconnaissance d'objets fonctionne.

La reconnaissance d'objets est une application directe de l'utilisation d'un outil de reconnaissance de formes pour répondre à ce besoin spécifique des personnes non-voyantes. Les personnes non-voyantes utilisent des méthodes pour retrouver des objets mais ces objets peuvent parfois être déplacés par d'autres personnes. Il est alors impossible de les retrouver. La reconnaissance et la localisation d'objets perdus répond donc à un des besoins identifiés pour augmenter l'autonomie des non-voyants mais répond aussi partiellement un à réel besoin en navigation. Le projet Navig fusionne des données provenant d'un GPS avec les données de la vision artificielle. Un système d'information géolocalisé comportant l'ensemble du mobilier urbain a été conçu dans des parcours prédéfinis pour le projet. Une modélisation du mobilier urbain pour la vision permettrait de charger à la volée des modèles géoréférencés à reconnaître, la position de l'utilisateur permettant de restreindre le nombre d'objets à charger. Les coordonnées GPS peuvent aider le système de vision en restreignant le nombre d'objets cherchés mais la vision peut aussi améliorer la précision du GPS, souvent inadaptée à une navigation pour le piéton. Pour cela, une reconnaissance et une localisation d'un ou plusieurs objets géoréférencés permettrait, par triangulation, d'inférer sur la position réelle de l'utilisateur. Il faut pour cela créer un modèle de fusion multi-capteurs GPS – Vision dans lequel chaque donnée est pondérée par la confiance affectée aux différents capteurs à un instant donné. En collaboration avec Olivier Gutierrez, nous travaillons actuellement à l'implémentation d'un réseau bayésien pour attribuer en temps réel la confiance accordée à chaque capteur. L'étude du fonctionnement de l'algorithme de vision présenté dans ce manuscrit nous permet de mieux appréhender la façon de l'utiliser : nous allons privilégier la reconnaissance de modèles « facilement » reconnaissables pour la vision : des objets d'intérêt, des logos, des façades, des amers qui peuvent n'avoir aucun sens pour l'utilisateur mais qui peuvent servir de point d'ancrage géolocalisés pour le système. La vision tient donc une place centrale dans le projet Navig.

7) Restitution : choix de la modalité et de la méthode

Comme pour le choix du capteur, la manière de restituer l'information doit être pensée de manière fonctionnelle et non pas par analogie au système sensoriel déficient. Certaines fonctions de la vision ont été identifiées comme particulièrement utiles et ce sont celles-ci que nous allons restaurer. La localisation d'objets est cette faculté immédiate du système visuel humain que nous souhaitons restituer par le biais d'une autre modalité.

Les systèmes de suppléance fonctionnent en capturant une information, en la traitant et enfin en la restituant. Les différentes études sur la substitution sensorielle nous informent de la sensibilité du système sensoriel pour reconnaître des formes mais aussi que cette méthode ne permet pas de localiser instantanément un objet dans l'espace : les sujets doivent balayer l'espace avec la caméra et interpréter les signaux pour inférer la position probable de l'objet. Ce que nous proposons est d'utiliser la faculté du système auditif à localiser précisément et rapidement des sons pour fournir des informations de position sur les objets. L'étude que nous avons menée auprès de sujets voyants et non-voyants nous montre qu'il est possible de localiser des cibles auditives avec un stimulus sonore de seulement 120 ms. Cette méthode de restitution permet donc de restituer très rapidement, avec une précision d'environ 10 cm d'erreur dans l'espace proche (<1 m) la position d'un objet sonore réel. Il n'est en revanche pas possible d'établir avec précision la distance à une cible sonore lorsque celle-ci se situe dans l'espace extra-personnel (Blauert, 1997). La littérature dans ce cas de figure contient des informations montrant en revanche que la précision angulaire (azimut et élévation) est excellente dans ce cas. Il est donc possible dans l'espace proche d'indiquer une position tridimensionnelle et dans l'espace lointain une position angulaire sans en connaître la distance. Les recherches en acoustiques sur la synthèse binaurale montrent qu'il est possible de synthétiser un son virtuel en établissant des fonctions de transfert relatives à la tête des sujets (HRTF), personnelles et uniques. Pour le projet Navig a été conçu un module de synthèse sonore permettant de générer un son à une position virtuelle voulue. Une banque de 160 HRTF enregistrée auprès de sujets humains permet d'étudier différentes stratégies pour établir une fonction de transfert pour chaque individu pour une utilisation en navigation dans des environnements réels, grâce à des sons virtuels ou réels. Une première approche est de mesurer les fonctions de transfert propres à chaque individu. Cette méthode est longue à mettre en place et coûteuse et n'est pas adaptée pour un système destiné à être utilisé par beaucoup de sujets. Une deuxième manière d'opérer est d'établir parmi la bibliothèque de fonctions de transfert, celles qui se

rapprochent le plus de celles du sujet au moyen de filtres morphologiques et perceptifs. Cette méthode est très pratique mais les sujets pourraient – si la synthèse n'est pas parfaite – commettre des erreurs d'interprétation entre les sons réels et les sons virtuels. Une troisième méthode serait d'établir pour un sujet des fonctions de transfert très différentes de la réalité pour ne pas induire d'erreurs d'interprétation entre des sons réels ou virtuels. Il faudrait pour cela que les sujets soient capables d'apprendre deux codages différents de la position d'une source sonore et qu'ils soient capables de passer très rapidement de l'un à l'autre. Les données manquent sur la précision de localisation de cibles sonores synthétisés avec des HRTF qui ne nous appartiennent pas , mais il est en tout à fait possible qu'il soit préférable d'utiliser une restitution très différente du réel (Van Wanrooij and Van Opstal, 2005) pour ne pas induire de perturbation sur la perception naturelle des sons de l'environnement.

La synthèse binaurale dans le cas d'objets proches peut donc servir à restituer instantanément la position 3D d'un objet, mais elle peut aussi servir à indiquer la direction d'un objet ou un cap à suivre dans un espace plus lointain. C'est l'idée que nous poursuivons dans ce projet : un des besoins établis par les sujets non-voyants est la difficulté de maintenir un cap pour traverser une place, ou une route de manière perpendiculaire (passage piéton), en l'absence de repère que la canne blanche puisse détecter. Couplé à un système de reconnaissance d'objets, il est alors possible à tout moment de demander au système de garder un cap et de se baser sur les informations visuelles de la scène pour restituer le cap à suivre par un son spatialisé provenant de la direction à prendre.

La méthode de restitution sera évaluée en réel dans trois parcours entièrement contrôlés. Une fois la méthode validée, nous tenterons d'utiliser d'autres dimensions du son pour augmenter l'information restituée par celui-ci. Les dispositifs de substitution sensorielle nous apportent de nombreux résultats fondamentaux sur la précision des modalités tactiles ou auditives et montrent que des motifs simples peuvent être reconnaissables avec un peu d'entraînement. Pour la modalité tactile, une carte électro-tactile permet de restituer la forme mais aussi la position de l'objet dans l'image par stimulation de la langue. Nous avons présenté trois manières de restituer un motif simple par le canal auditif : la méthode 'The vOICe', 'The Vibe' et EAV. Cette dernière présente l'avantage de restituer la position en trois dimensions de l'objet en plus de sa forme. Si ces systèmes sont aujourd'hui inutilisables dans des environnements complexes, ils sont évalués dans des conditions très simples où ils

pourraient devenir intéressants s'ils sont couplés à des modules d'analyse de scènes visuelles.

Dans le cas d'une neuroprothèse visuelle, la restitution se fait par le biais de la stimulation électrique du système visuel humain au niveau principalement de trois sites : la rétine, le nerf optique et le cortex visuel. Cette stimulation électrique fait percevoir aux sujets stimulés des sensations visuelles appelées phosphènes. La plupart des études sur la conception de neuroprothèses visuelles montrent que la résolution visuelle possible à restaurer est très faible par rapport à la résolution d'une image dans laquelle les objets seraient visuellement identifiables. L'idée que nous poursuivons est que restaurer la vision par le biais de telles interfaces n'est aujourd'hui pas réalisable. Nous proposons plutôt d'étudier la possibilité de donner une information de position tridimensionnelle sur des objets précis par le biais de phosphènes évoqués par des stimulations électriques du système visuel. L'idée proposée est donc de coupler un système capable de reconnaître des objets dans une scène visuelle avec une interface de restitution invasive. Différents modes de restitution sont actuellement testés et ce manuscrit présente des résultats préliminaires permettant de montrer qu'avec très peu de percepts évoqués, il est possible de guider des utilisateurs jusqu'à une cible, dans un monde virtuel. Ces résultats préliminaires sont très encourageants et nécessitent d'être évalués plus en profondeur en développant un modèle de neuroprothèse toujours plus proche des percepts réellement évoqués et évalué dans des tâches de localisation d'objets et de navigation réelles.

8) Navig : un système d'aide à la navigation et à la localisation d 'objets pour les non-voyants

L'ensemble des résultats présentés dans cette thèse ont contribué à la naissance du projet Navig : Navigation Assistée par VIsion embarquée et GNSS. Le dispositif comporte un outil de géolocalisation, un système d'information géoréférencé adapté au piéton non-voyant et des modèles visuels d'objets géolocalisés adaptés aux spécifications du module de vision, issues des recommandations de ce manuscrit. Ces trois modules permettent d'analyser l'environnement proche (vision) mais aussi de situer l'utilisateur dans un repère spatial plus grand et prédire un parcours vers une destination cible. Son fonctionnement peut être divisé en deux modes : Recherche d'objet en champ proche et Navigation.

Deux besoins spécifiques des personnes non-voyantes ont été identifiés : le besoin de localiser des objets (toute entité visuelle discernable ou utile à un instant donné) et le besoin d'aide à l'orientation. C'est à ceux-ci que nous tentons de répondre avec le dispositif Navig.

Aide à la recherche d'objets

Le capteur de reconnaissance et de localisation d'objets décrit dans cette thèse permet de répondre à un besoin spécifique des personnes non-voyantes. Le système est doté d'une interface utilisateur pourvue en entrée d'une reconnaissance vocale permettant de connaître les requêtes utilisateur (Figure 126). L'utilisateur peut alors demander au système la recherche d'un objet précis.

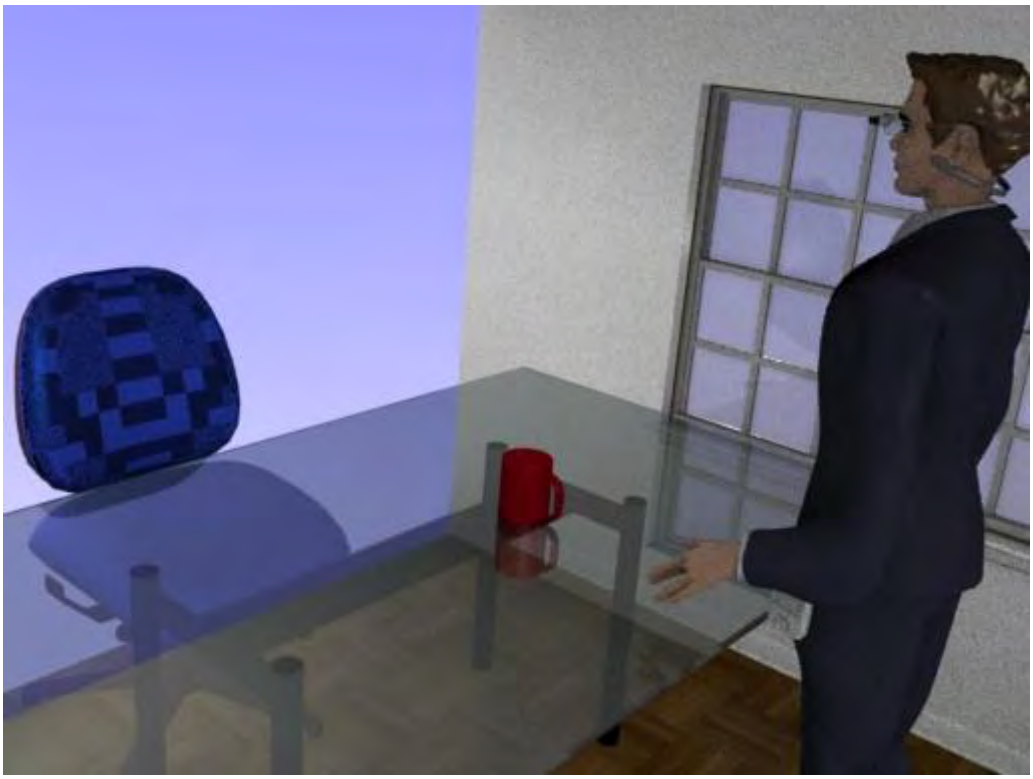


Figure 126: Illustration de l'utilisation du dispositif Navig en intérieur pour reconnaître et localiser un objet. L'utilisateur demande au système "où est ma tasse" et le système émet un son perçu comme s'il provenait de la tasse elle-même

Il balaye alors l'espace et un son est émis dans un casque audio comme si celui-ci provenait de l'objet recherché dès que celui-ci entre dans le champ de vision des caméras. Deux caméras sont aujourd'hui disposées sur un casque sur la tête des sujets. Le champ de vue des caméras utilisées est très important puisqu'il conditionne le temps de recherche de l'objet. En effet, des caméras grand-angle sont indispensables afin de limiter le travail de balayage de l'utilisateur. Les caméras aujourd'hui utilisées (100° d'angle de vue, résolution

de 640x480 maximum) sont adaptées à ce mode de fonctionnement mais il est évident qu'un capteur permettant la reconnaissance et la localisation à 360° serait très efficace pour limiter le temps de recherche par balayage. Le temps de traitement des informations visuelles est proportionnel à la taille des images à traiter et donc du champ de vue traité. Un capteur matériel de reconnaissance et de localisation d'objets avec un processeur dédié à chaque caméra est à l'étude pour répondre à ce besoin, permettant de largement réduire les temps de calcul du processeur central et d'augmenter le nombre de capteurs. Mais un large champ de vue couplé à une faible résolution (320x240 utilisé aujourd'hui) pose des problèmes pour la fiabilité de reconnaissance des petits objets par les algorithmes de vision, leur définition étant trop faible. Une solution matérielle permettrait donc de répondre de manière plus efficace à ce besoin de localisation d'objets.

Aide au déplacement

La localisation d'amers visuels permet de localiser des objets d'intérêts mais aussi de prévenir la rencontre d'obstacles. Les obstacles sont un problème complexe à traiter par analyse d'images puisqu'il est impossible de modéliser cette catégorie d'objets sans connaître l'identité de chacun d'eux. Certains amers visuels sont tout à fait reconnaissables (ex. boîte aux lettres) et ils constituent aussi des objets d'intérêt. En revanche, il est très difficile de détecter une barrière ou un trottoir par exemple. La détection de cette catégorie d'objets étant difficile, il est possible de donner des informations sur les volumes de l'espace environnant en détectant la distance à chaque surface dans l'image (par vision stéréoscopique) et prévenir qu'un objet volumineux se rapproche. Une autre manière de traiter ce problème serait par exemple d'utiliser des méthodes de détection d'appuyant sur l'analyse du flux optique au cours du déplacement. Cette fonction de détection d'obstacles est particulièrement sensible à destination de personnes aveugles et doit donc être très fiable. L'utilisation du capteur de vision pour traiter ce besoin en détection d'obstacles est possible mais ne sera pas fiable à 100% du fait des artéfacts visuels, des changements de luminosité etc. L'état de l'art sur les dispositifs de suppléance montre qu'il existe des dispositifs fiables pour cette tâche, s'appuyant sur des télémètres lasers ou des ultrasons. Le but de Navig n'est donc pas de répondre partiellement à ce problème alors que des solutions fiables existent. Il est en revanche possible de restituer la position tridimensionnelle des amers visuels présents dans l'environnement proche des personnes non-voyantes. Ces informations peuvent permettre aux personnes non-voyantes de se créer une représentation spatiale de leur environnement proche. Il est ainsi possible, en se trouvant

sur une place, que l'utilisateur demande au système de le guider jusqu'à un distributeur de billets. Pour cela, un son est émis à l'utilisateur comme si celui-ci provenait de la cible visuelle localisée en 3D. Il est alors possible de suivre et garder le cap jusqu'au distributeur de billets. Ce mode de fonctionnement n'est possible que dans l'espace proche de l'utilisateur : à portée des amers visuels reconnaissables et localisables par le système de vision artificielle. Au-delà de ce périmètre, un outil de géolocalisation est indispensable pour aider à l'orientation des personnes non-voyantes pour effectuer un trajet entre deux points éloignés.

Aide à l'orientation

L'aide à l'orientation consiste à permettre aux utilisateurs de s'orienter dans un lieu et vers une destination. La vision artificielle embarquée ne peut pas répondre seule à ce problème d'orientation. Les systèmes GPS ne sont aujourd'hui pas adaptés à l'orientation du piéton non-voyant du fait de leur précision insuffisante, particulièrement en ville et du manque de systèmes d'informations géolocalisés précis et adapté. Ces systèmes souffrent d'un autre problème majeur dans la manière avec laquelle l'information est restituée. Ils permettent en effet d'atteindre une cible avec un guidage pas à pas mais aucun système ne permet de se représenter l'espace de manière allocentrée. En partenariat avec le Grand Toulouse, nous établissons dans des trajets prédéfinis et contraints dans Toulouse un système d'information comportant l'ensemble du mobilier urbain nécessaire au piéton non-voyant. L'idée est de pouvoir prévenir les obstacles mais aussi guider l'utilisateur dans cet environnement. L'idée poursuivie est que la localisation d'amers visuels géolocalisés dans un système d'information peut améliorer la précision de la géolocalisation. La plupart des instruments de mesure d'orientation ne sont pas précis dans des milieux magnétiquement bruités et connaître avec précision la position et l'orientation du sujet est donc très imprécis en milieu urbain. Reconnaître et localiser des objets permettraient par triangulation de reconstruire ou améliorer la précision de l'évaluation de la position de l'utilisateur dans son environnement. Ce fonctionnement, complètement invisible pour l'utilisateur, permettrait d'accroître la fiabilité et la précision du système. La géolocalisation par fusion multi-capteurs à laquelle je collabore avec un ingénieur de l'équipe Olivier Gutierrez (vision, magnétomètre, GPS) nécessite des algorithmes complexes qui ne sont pas abordés dans cette thèse mais qui prennent une part importante dans le développement de l'outil de suppléance. Différents modes de guidage pour les non-voyants à partir du système de reconnaissance et localisation d'objets sont en cours d'évaluation et de nombreux travaux avec les utilisateurs

finaux sont encore à effectuer pour arriver à un guidage efficace et discret compatible avec une utilisation quotidienne.

Conclusion

Parmi les outils de suppléance développés pour les non-voyants, seule une minorité est aujourd'hui utilisée au quotidien. La canne blanche améliorée par augmentation tactile ou auditive en est un des rares exemples. L'étude de l'état de l'art montre que la plupart des outils de suppléance ne sont pas utilisés car ils ne répondent pas à des besoins identifiés des utilisateurs non-voyants. A partir d'une étude non-exhaustive du besoin utilisateur et de l'état de l'art sur les systèmes de suppléance, nous avons étudié dans ce travail de thèse la faisabilité d'un système de suppléance pour l'aide à la reconnaissance et la localisation d'objets d'une part et à la navigation d'autre part. L'évaluation du système dans son ensemble dans des tâches de navigation nous permettra d'améliorer l'utilisabilité du système jusqu'à l'obtention d'un système fonctionnel et utilisable par les utilisateurs finaux. La première version du prototype Navig est pourvue de capteurs inertiels et magnétiques, d'un GPS et d'un système de vision artificielle pour la localisation d'objets. Elle est opérationnelle et permet dans des trajets simples et contraints d'être guidé de manière fiable. Ces recherches montrent qu'il est possible aujourd'hui de répondre à une partie du besoin des non-voyants par la localisation d'objets avec une interface de restitution sonore nécessitant un faible apprentissage et qui s'avère peu envahissante. Un portage sur du matériel dédié accroîtrait les possibilités du système de vision artificiel en permettant d'analyser les images à des résolutions plus élevées ou avec un grand champ de vue plus large. Le principal frein au développement de systèmes de suppléances est aujourd'hui la relative faiblesse des algorithmes d'analyse de la scène visuelle. Le capteur développé permettant la reconnaissance et la localisation de cibles visuelles est un point de départ à de nombreux travaux qui ont déjà débuté à l'IRIT. Une des applications à plus long terme de son utilisation est le développement d'une neuroprothèse visuelle sur la base de ce capteur.

Bibliographies

Arno, P., C. Capelle, M. C. Wanet-Defalque, M. Catalan-Ahumada and C. Veraart, Auditory coding of visual patterns for the blind, *Perception, 28*(8), 1013-1029, 1999.

Auvray, M., Immersion et perception spatiale. L'exemple des dispositifs de substitution sensorielle, 2004.

Auvray, M., S. Hanneton and J. K. Regan, Learning to perceive with a visuo - auditory substitution system: Localisation and object recognition with 'The vOICe', *Perception, 36*, 416-430, 2007.

Bach-y-Rita, P., Tactile vision substitution: past and future, *Int. J. Neurosci., 19*(1-4), 29-36, 1983.

Bach-y-Rita, P., C. C. Collins, F. A. Saunders, B. White and L. Scadden, Vision substitution by tactile image projection, *Trans Pac. Coast. Otoophthalmol. Soc. Annu. Meet., 50*, 83-91, 1969.

Bach-y-Rita, P., K. A. Kaczmarek, M. E. Tyler and J. Garcia-Lara, Form perception with a 49-point electrotactile stimulus array on the tongue: a technical note, *J. Rehabil. Res. Dev., 35*(4), 427-430, 1998.

Bay, H., A. Ess, T. Tuytelaars and L. V. Gool, SURF: Speeded Up Robust Features, 2008.

Blauert, J., Binaural localization, *Scand. Audiol. Suppl, 15*, 7-26, 1982.

Blauert, J., Spatial Hearing, The psychophysics of Human Sound Localization, 1997.

Blum, A., B. F. G. Katz and O. Warusfel, Eliciting adaptation to non-individual HRTF spectral cues with multi-modal training, 2004.

Brindley, G. S., Effects of electrical stimulation of the visual cortex, *Hum. Neurobiol., 1*(4), 281-283, 1982.

Brindley, G. S. and W. S. Lewin, The sensations produced by electrical stimulation of the visual cortex, *J. Physiol, 196*(2), 479-493, 1968.

Brungart, D. S., Auditory localization of nearby sources. III. Stimulus effects, *J. Acoust. Soc. Am., 106*(6), 3589-3602, 1999.

Brungart, D. S., N. I. Durlach and W. M. Rabinowitz, Auditory localization of nearby sources. II. Localization of a broadband source, *J. Acoust. Soc. Am., 106*(4 Pt 1), 1956-1968, 1999.

Brungart, D. S., A. J. Kordik, B. D. Simpson and R. L. McKinley, Auditory localization in the horizontal plane with single and double hearing protection, *Aviat. Space Environ. Med., 74*(9), 937-946, 2003.

Brungart, D. S. and W. M. Rabinowitz, Auditory localization of nearby sources. Head-related transfer functions, *J. Acoust. Soc. Am., 106*(3 Pt 1), 1465-1479, 1999.

Cheng, C., B. O'Leary, L. Stearns, S. Caperna, J. Cho, V. Fan, A. Luthra, A. Sun, R. Tessler, P. Wong, J. Yeh, B. Bobo, R. Chellappa and C. M. Tang, Developing a Real-Time Identify-and-Locate System for the Blind (on line), 2008.

Chow, A. Y., V. Y. Chow, K. H. Packo, J. S. Pollack, G. A. Peyman and R. Schuchard, The artificial silicon retina microchip for the treatment of vision loss from retinitis pigmentosa, *Arch. Ophthalmol., 122*(4), 460-469, 2004.

Dagnelie, Yin, Hess and Yang, Phosphene mapping strategies for cortical visual prosthesis recipients, *J. Vis., 3*(9), 222, 2003.

Deprés, O., Mécanismes de localisation spatiale chez l'homme : interaction entre le système visuel et le système auditif, 2006.

Dobelle, W. H., Artificial vision for the blind by connecting a television camera to the visual cortex, *ASAIO J., 46*(1), 3-9, 2000a.

Dobelle, W. H., Artificial vision for the blind by connecting a television camera to the visual cortex, *ASAIO J., 46*(1), 3-9, 2000b.

Dobelle, W. H., M. G. Mladejovsky and J. P. Girvin, Artificial vision for the blind: electrical stimulation of visual cortex offers hope for a functional prosthesis, *Science, 183*(123), 440-444, 1974.

Doucet, M. E., J. P. Guillemot, M. Lassonde, J. P. Gagne, C. Leclerc and F. Lepore, Blind subjects process auditory spectral cues more efficiently than sighted individuals, *Exp. Brain Res., 160*(2), 194-202, 2005.

Dramas, F., B. F. Katz and C. Jouffrais, Auditory-guided reaching movements in the peripersonal frontal space, *The Journal of the Acoustical Society of America, 123*, 3723, 2008.

Duret, F., M. E. Brelen, V. Lambert, B. Gerard, J. Delbeke and C. Veraart, Object localization, discrimination, and grasping with the optic nerve visual prosthesis, *Restor. Neurol. Neurosci., 24*(1), 31-40, 2006.

Durette, B., N. Louveton, D. Alleysson and J. Hérault, Visuo-auditory sensory substitution for mobility assistance: testing TheVIBE, 2008.

Eckhorn, R., M. Wilms, T. Schanze, M. Eger, L. Hesse, U. T. Eysel, Z. F. Kisvarday, E. Zrenner, F. Gekeler, H. Schwahn, K. Shinoda, H. Sachs and P. Walter, Visual resolution with retinal implants estimated from recordings in cat visual cortex, *Vision Res., 46*(17), 2675-2690, 2006.

Farcy, R. and R. Damaschini, Guidance – Assist systems for the blind, paper presented at EBIOS 2000, Amsterdam, 2000.

Farcy, R., R. LEROUX, R. Damaschini, R. LEGRAS, Y. Bellik, C. Jacquet, J. Greene and P. Pardo, Laser telemetry to improve the mobility of blind people: Report of the 6 month training course, 2003.

Farne, A. and E. Ladavas, Auditory peripersonal space in humans, *J. Cogn Neurosci., 14*(7), 1030-1043, 2002.

Fukuda, T., T. Horiuchi, H. Hokari and S. Shimada, Relative distance perception by manipulating the ILD of HRTFs, 2003.

Gekeler, F., P. Szurman, S. Grisanti, U. Weiler, R. Claus, T. O. Greiner, M. Volker, K. Kohler, E. Zrenner and K. U. Bartz-Schmidt, Compound subretinal prostheses with extra-ocular parts designed for human trials: successful long-term implantation in pigs, *Graefes Arch. Clin. Exp. Ophthalmol.,* 2006.

Gold, D. and H. Simson, Identifying the needs of people in Canada who are blind or visually impaired: Preliminary results of a nation-wide study, *International Congress Series, 1282*, 139-142, 2005.

Gonzalez-Mora, J. L., A. Rodriguez-Hernandez, E. Burunat, F. Martin and M. A. Castellano, Seeing the world by hearing: Virtual Acoustic Space (VAS) a new space perception system for blind people, 2006.

Hausfeld, S., R. P. Power, A. Gorta and P. Harris, Echo perception of shape and texture by sighted subjects, *Percept. Mot. Skills, 55*(2), 623-632, 1982.

Helal, A., S. E. Moore and B. Ramachandran, Drishti: An Integrated Navigation System for Visually Impaired and Disabled, 2001.

Hub, A., T. Hartter and T. Ertl, Interactive Tracking of Movable objects for the Blind on the Basis of Environment Models and Perception-Oriented Object Recognition Methods, 2007.

Hui, T. and D. J. Beebe, Design and microfabrication of a flexible oral electrotactile display, *Microelectromechanical Systems, Journal of, 12*(1), 29-36, 2003.

Hui, T. and D. J. Beebe, An oral tactile interface for blind navigation, *Neural Systems and Rehabilitation Engineering, IEEE Transactions on [see also IEEE Trans. on Rehabilitation Engineering], 14*(1), 116-123, 2006.

Humayun, M. S., J. de, Jr., J. D. Weiland, G. Dagnelie, S. Katona, R. Greenberg and S. Suzuki, Pattern electrical stimulation of the human retina, *Vision Research, 39*(15), 2569-2576, 1999.

Kaczmarek, K. A., Sensory Augmentation and Substitution, in The Biomedical Engineering Handbook, 2000.

Kaczmarek, K. A., M. E. Tyler and P. Rita, Pattern identification on a fingertip-scanned electrotactile display, 1997.

Kammer, T., K. Puls, M. Erb and W. Grodd, Transcranial magnetic stimulation in the visual system. II. Characterization of induced phosphenes and scotomas, *Experimental Brain Research, 160*(1), 129-140, 2005.

Kulkarni, A. and H. S. Colburn, Role of spectral detail in sound-source localization, *Nature, 396*(6713), 747-749, 1998a.

Kupers, R. and M. Ptito, "Seeing" through the tongue: cross-modal plasticity in the congenitally blind, *International Congress Series, 1270*, 79-84, 2004.

Lenay, C., O. Gapenne, S. Hanneton, Marque Catherine and Genouëlle Christelle, Sensory Substitution: Limits and Perspectives, in Touching for knowing, John Benjamins Publishing Company, 2003.

Lessard, N., M. Pare, F. Lepore and M. Lassonde, Early-blind human subjects localize sound sources better than sighted subjects, *Nature, 395*(6699), 278-280, 1998.

Liu Xu, A camera phone based currency reader for the visually impaired, 2008.

Loomis, J. M., R. G. Golledge and R. Klatzky, Navigation System for the Blind: Auditory Display Modes and Guidance, *Presence,* 1998.

Loomis, J. M., R. G. Golledge, R. L. Klatzky, J. M. Speigle and J. Tietz, Personal guidance system for the visually impaired, 1994.

Lowe, D. G., Object recognition from local scale-invariant features, 1999.

Maeda, Y., E. Tano, H. Makino, T. Konishi and i. I. Ishi, Evaluation of a GPS-based guidance system for visually impaired pedestrians, 2002.

Makino, H., F. Morishita, Y. Abe, S. Yamamiya, M. Hasegawa, I. Ishii and M. Nakashizuka, 3-D object recognition and description: A method for the visually impaired using an invisible bar code, *Systems and Computers in Japan, 29*(8), 1-8, 1998.

Meijer, P. B., An experimental system for auditory image representations, *IEEE Trans. Biomed. Eng, 39*(2), 112-121, 1992.

Middlebrooks, J. C. and D. M. Green, Sound localization by human listeners, *Annu. Rev. Psychol., 42*, 135-159, 1991.

Normann, R. A., E. M. Maynard, P. J. Rousche and D. J. Warren, A neural interface for a cortical vision prosthesis, *Vision Res., 39*(15), 2577-2587, 1999.

Pissaloux, E. E., R. Velazquez and F. Maingreaud, On 3D world perception: towards a definition of a cognitive map based electronic travel aid, *Conf. Proc. IEEE Eng Med. Biol. Soc., 1*, 107-109, 2004.

Ran, Helal, Moore and Drishti, An Integrated Indoor/Outdoor Blind Navigation System and Service., 2004.

Rice, C. E. and S. H. Feinstein, Sonar system of the blind: Size discrimination, *Science, 148*, 1107-1108, 1965.

Rizzo, J. F., J. Wyatt, J. Loewenstein, S. Kelly and D. Shire, Perceptual efficacy of electrical stimulation of human retina with a microelectrode array during short-term surgical trials, *Invest Ophthalmol. Vis. Sci., 44*(12), 5362-5369, 2003.

Roder, B., F. Rosler and E. Hennighausen, Different cortical activation patterns in blind and sighted humans during encoding and transformation of haptic images, *Psychophysiology, 34*(3), 292-307, 1997.

Roder, B., W. Teder-Salejarvi, A. Sterr, F. Rosler, S. A. Hillyard and H. J. Neville, Improved auditory spatial tuning in blind humans, *Nature, 400*(6740), 162-166, 1999.

Sadato, N., A. Pascual-Leone, J. Grafman, V. Ibanez, M. P. Deiber, G. Dold and M. Hallett, Activation of the primary visual cortex by Braille reading in blind subjects, *Nature, 380*(6574), 526-528, 1996.

Sampaio, E., S. Maris and P. Rita, Brain plasticity: 'visual' acuity of blind persons via the tongue, *Brain Research, 908*(2), 204-207, 2001.

Schmidt, E. M., M. J. Bak, F. T. Hambrecht, C. V. Kufta, D. K. O'Rourke and P. Vallabhanath, Feasibility of a visual prosthesis for the blind based on intracortical microstimulation of the visual cortex, *Brain, 119 (Pt 2)*, 507-522, 1996.

Simpson, W. and L. Stanton, Head movement does not facilitate perception of the distance of a source of sound., *American Journal of Psychology, 86*, 151-159, 1973.

Srivastava, N. R., P. R. Troyk, V. L. Towle, D. Curry, E. Schmidt, C. Kufta and G. Dagnelie, Estimating Phosphene Maps for Psychophysical Experiments used in Testing a Cortical Visual Prosthesis Device, 2007.

Thorpe, S., Ultra-rapid scene categorization with a wave of spikes, *Biologically Motivated Computer Vision, Proceedings, 2525*, 1-15, 2002.

Thorpe, S., A. Delorme and Van Rullen R., Spike-based strategies for rapid processing, *Neural Netw., 14*(6-7), 715-725, 2001a.

Thorpe, S., D. Fize and C. Marlot, Speed of processing in the human visual system, *Nature, 381*(6582), 520-522, 1996.

Thorpe, S. J. and M. Fabre-Thorpe, Neuroscience. Seeking categories in the brain, *Science, 291*(5502), 260-263, 2001.

Thorpe, S. J., K. R. Gegenfurtner, M. Fabre-Thorpe and H. H. Bulthoff, Detection of animals in natural images using far peripheral vision, *Eur. J Neurosci, 14*(5), 869-876, 2001b.

Thorpe, S. J., R. Guyonneau, N. Guilbaud, J. M. Allegraud and R. VanRullen, SpikeNet: real-time visual processing with one spike per neuron, *Neurocomputing, 58-60*, 857-864, 2004.

Uhl, F., T. Kretschmer, G. Lindinger, G. Goldenberg, W. Lang, W. Oder and L. Deecke, Tactile mental imagery in sighted persons and in patients suffering from peripheral blindness early in life, *Electroencephalogr. Clin. Neurophysiol., 91*(4), 249-255, 1994.

Van Wanrooij, M. M. and A. J. Van Opstal, Relearning Sound Localization with a New Ear, *J. Neurosci., 25*(22), 5413-5424, 2005.

Viola, P. and M. Jones, Rapid object detection using a boosted cascade of simple features, 2001.

Voss, P., M. Lassonde, F. Gougoux, M. Fortin, J. P. Guillemot and F. Lepore, Early- and late-onset blind individuals show supra-normal auditory abilities in far-space, *Curr. Biol., 14*(19), 1734-1738, 2004.

Wanet-Defalque, M. C., C. Veraart, V. A. De, R. Metz, C. Michel, G. Dooms and A. Goffinet, High metabolic activity in the visual cortex of early blind human subjects, *Brain Res., 446*(2), 369-373, 1988.

Willis, S. and S. Helal, RFID Information Grid for Blind Navigation and Wayfinding, 2005.

Zabihaylo, C., Y. Jalbert and M. Gauthier, Evaluation de la satisfaction et de l'utilisation d'une aide à l'orientation, le trekker, par des utilisateurs ayant une déficience visuelle, 2006.

ANNEXE

Questionnaire pour l'étude du besoin utilisateur en navigation et localisation de cibles

Le but de ce questionnaire est de connaître vos habitudes de vie, vos besoins dans la vie de tous les jours, dans un milieu connu ou inconnu afin de mieux les évaluer et y répondre.

Il vous faudra environ 5 à 10 minutes pour le remplir.

Une fois vos réponses validées, elles seront envoyées automatiquement par mail à l'adresse dramas@irit.fr.

---

1) I. Questions générales

1. Quel est votre âge ?

- ○ Moins de 20 ans
- ○ 20 - 40 ans
- ○ 40 - 60 ans
- ○ plus de 60 ans

2. Sortez-vous à l'extérieur seul...

- ○ Plus d'une fois par jour ?
- ○ une fois par jour ?
- ○ Deux fois par semaine ?
- ○ une fois par semaine ?
- ○ deux fois par mois ?
- ○ Une fois par mois ?
- ○ Jamais

3. Quel type d'aide utilisez vous ?

- ☐ Canne blanche
- ☐ Chien
- ☐ Outil electronique tel une canne laser; Teletact ...
- ☐ Aucune car je n'en ai pas l'utilité
- ☐ Aucune mais j'envisage d'en acquérir une

☐ Autre (précisez :)

4. Quels sont les problèmes que vous rencontrez avec ces aides et quelles en sont les limites ?

5. prenez vous les transports en commun ...

- ○ Plus d'une fois par jour ?
- ○ une fois par jour ?
- ○ Deux fois par semaine ?
- ○ une fois par semaine
- ○ deux fois par mois ?
- ○ Une fois par mois ?
- ○ Jamais ?

6. En etes vous content, pourquoi ?

---

2) II. Navigation en intérieur

7. Y a-t-il des pièces de votre domicile où localiser ce que vous voulez vous pose problème ?

- ☐ Salle de Bain
- ☐ Cuisine
- ☐ Salon
- ☐ Chambre
- ☐ Salle à manger

- [ ] Garage
- [ ] Toutes les pièces de mon domicile

8. Quels sont les principaux obstacles auquels vous vous confrontez dans un espace clos ?

- [ ] Chaises
- [ ] Portes
- [ ] Jouets d'enfant
- [ ] Autres

9. Vous arrive-t-il de chercher ...

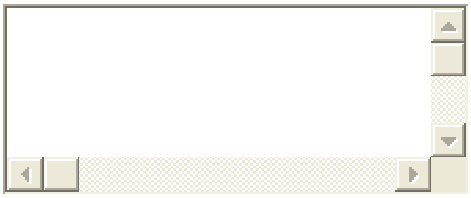
- [ ] Des ustensiles de toilette ?
- [ ] de l'alimentaire ?
- [ ] votre radio ?
- [ ] le Téléphone ?
- [ ] vos Clés?
- [ ] les ustensiles de cuisine ?
- [ ] Autres

10. Les principaux obstacles sont ils...

- [ ] A hauteur du visage (extincteur par exemple)
- [ ] Longs et verticaux (porte ouverte par exemple)
- [ ] Longs et horizontaux (table par exemple)
- [ ] Par terre (Jouets d'enfants par exemple)
- [ ] Autres

11. Quelle activité vous pose le plus de problèmes ?

- ☐ Détecter les obstacles
- ☐ retrouver des objets
- ☐ Autres

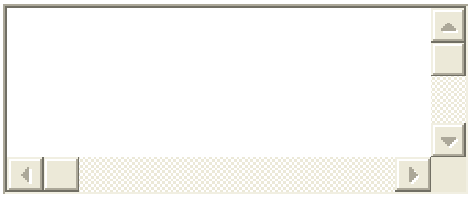


---

3) III. Navigation en Extérieur

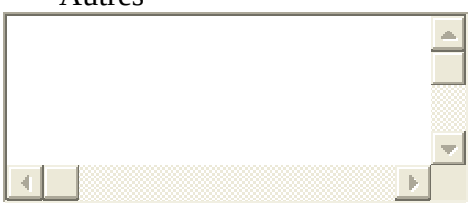
12. Quels sont les principaux obstacles en exterieur ?

- ☐ Voiture
- ☐ Cabine téléphonique
- ☐ Dénivellations brutales (trottoirs ....)
- ☐ Personnes
- ☐ Autres



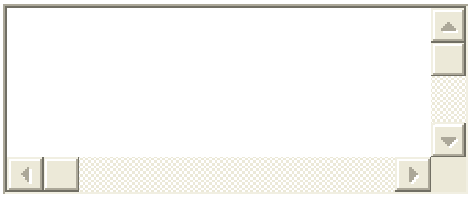
13. Que recherchez vous particulièrement quand vous êtes en extérieur ?

- ☐ Cabine téléphonique
- ☐ Passage piéton
- ☐ Rue
- ☐ Personnes
- ☐ Arret de bus
- ☐ Arbres
- ☐ Poubelles
- ☐ bancs
- ☐ Autres



14. Quelle activité vous pose le plus de problèmes en extérieur ?

- [ ] Atteindre des objets
- [ ] detecter des obstacles
- [ ] trouver une direction
- [ ] Garder un Cap
- [ ] Savoir où l'on se situe
- [ ] Autres



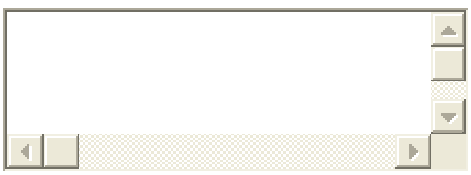
Motivation de ce questionnaire :

Ce questionnaire s'inscrit dans un projet à l'IRIT (Institut de Recherche en Informatique de Toulouse), en collaboration avec le Cerco (Cerveau et Cognition). Son objectif est de proposer des solutions au problème de la navigation et de la saisie d'objets dans des lieux connus ou inconnus en s'approchant au maximum aux besoins des non-voyants, en tirant des leçons des échecs des précédents systèmes. Le principe est simple et peut s'expliquer par un scénario de fonctionnement. Deux caméras miniatures sont disposées sur une monture de lunettes, "voient" et "comprennent" la scène. La personne demande "où sont mes clés ? ". Le système localise alors les clés et émet un son comme si celui-ci provenait des clés. Il est alors trivial et naturel de les retrouver. Le principe peut être appliqué de même à la détection d'obstacles.

Nous autorisez-vous à vous recontacter ultérieurement afin d'affiner notre évaluation du besoin après avoir analysé les réponses de ce premier questionnaire ?

- ( ) Oui
- ( ) Non

Si oui, Veuillez nous laisser votre adresse mail afin que nous puissions vous recontacter



Le questionnaire est terminé.

Merci d'avoir pris le temps de répondre.

Vous pouvez maintenant valider vos réponses. Un mail va être automatiquement envoyé à dramas@irit.fr. Ce mail contient uniquement vos réponses. Hormis votre adresse mail, aucune information personnelle ne sera envoyée.