Assistance multimodale à l'exploration de visualisations 2D interactives

Suzanne Kieffer

Département de formation doctorale en informatique
UFR STMIA

École doctorale IAEM Lorraine

Assistance multimodale à l'exploration de visualisations 2D interactives

THÈSE

présentée et soutenue publiquement le 6 Juillet 2005

pour l'obtention du

Doctorat de l'université Henri Poincaré – Nancy 1

(spécialité informatique)

par

Suzanne Kieffer

Composition du jury

Président :	M. Frédéric Alexandre, Professeur, Université Henri Poincaré Nancy 1
Rapporteurs :	M. Alistair Sutcliffe, Professeur, Université de Manchester Mme Catherine Pélachaud, Professeur, Université de Paris 8
Examinateurs :	M. Éric Lecolinet, Maître de Conférence, Telecom Paris M. Jean Vanderdonckt, Professeur, Université de Louvain-la-Neuve Mme Noëlle Carbonell, Professeur, Université Henri Poincaré Nancy 1

Laboratoire Lorrain de Recherche en Informatique et ses Applications — UMR 7503

Mis en page avec la classe thloria.

À ma mère

Remerciements

Je tiens à exprimer tous mes remerciements à Noëlle Carbonell pour m'avoir accueillie dans l'équipe MErLIn au LORIA. Grâce à elle, j'ai pu travailler dans un environnement scientifique rigoureux et orienter librement mes recherches dans des directions qu'elle a toujours su m'aider à exploiter. J'ai pu apprécier, pendant ces quatre ans, ses grandes qualité d'écoute et le précieux soutien qu'elle m'a apporté dans des moments un peu pénibles de ma vie.

Je remercie vivement Catherine Pélachaud et Alistair Sutcliffe qui ont accepté d'être les rapporteurs de ce travail et dont j'ai apprécié les nombreuses appréciations et suggestions, ainsi que l'intérêt qu'ils ont porté à ma thèse. Je remercie également Frédéric Alexandre qui m'a fait l'honneur de présider le jury, Jean Vanderdonckt et Éric Lecolinet pour leurs remarques et leurs questions enrichissantes lors de la soutenance.

Je remercie tous mes collègues de l'équipe MErLIn pour les moments d'échange durant ma thèse, et particulièrement Toni pour ses conseils avisés à la fois professionnels et personnels, Jérôme pour son implication dans la phase finale de mon travail, et plus récemment Olivier et Marius pour leur bonne humeur. Un grand merci à Feirouz, membre du bureau B216, et à Djamé, occupant invité du dit bureau, pour les conversations souvent explosives qui ont animé nos soirées au laboratoire pendant plus de trois ans.

Je tiens à remercier vivement tous les participants aux trois expérimentations que j'ai menées dans le cadre de mes recherches.

Merci à Lucy, Serges, Sergio et Ricardo du groupe d'anglais qui m'ont beaucoup apporté tant sur un aspect humain que sur l'aspect "oral présentation". Merci à Céline, Tamara, Betty, Iadine, pour les soirées où le service est toujours "de grande qualité", Mohammed et Suzanne, Cédric, Émily et Greg, les péons hors norme, et John qui m'ont tous permis de m'évader un peu. Un grand merci aux membres du club de bowling de Vandœuvre et, particulièrement à mes entraîneurs et amis Christine et Dominique.

Merci à Fix qui m'a fait profiter de ses connaissances en statistiques, m'a accueillie moult fois à Paris et avec qui même des débats sur Brad Pitt peuvent remplir une soirée entière.

Je tiens à remercier chaleureusement Alain Zimmerlé, qui m'a donné le goût des mathématiques et des études en général. Un merci tout particulier à Madeleine et Jean, ma deuxième famille, à Francette et Livia, à Nicolas et Kathia, et à mon père.

Enfin, je tiens à remercier mes proches qui m'ont encouragée, poussée, réconfortée et qui ont vécu au plus près de moi les moments forts de cette thèse. Merci donc à mes amis et confidents, Stéphanie, Jean-Marc et Daniel.

Résumé

Ce travail porte sur la conception d'une nouvelle forme d'interaction Homme-Machine : la multimodalité parole+présentation visuelle en sortie du système. Plus précisément, l'étude porte sur l'évaluation des apports potentiels de la parole, en tant que mode d'expression complémentaire du graphique, lors du repérage visuel de cibles au sein d'affichages 2D interactifs.

Nous avons adopté une approche expérimentale pour déterminer l'influence d'indications spatiales orales sur la rapidité et la précision du repérage de cibles, et évaluer la satisfaction subjective d'utilisateurs potentiels dans cette forme d'assistance à l'activité d'exploration visuelle.

Les différentes études réalisées ont montré d'une part que les présentations multimodales facilitent et améliorent les performances des utilisateurs pour le repérage visuel, en termes de temps et de précision de sélection des cibles. Elles ont montré d'autre part que les stratégies d'exploration visuelle des affichages, en l'absence de messages sonores, dépendent de l'organisation spatiale des informations au sein de l'affichage graphique.

Abstract

This work is about the design of a new form of Human-Computer Interaction: the speech+visual presentation combination as an output multimodality. More precisely, it deals with the valuation of the potential benefits from the speech, as a graphical expression mode, during visual target detection tasks in 2D interactive layouts.

We used an experimental approach to show the influence of oral messages, including the spatial localization of the target in the layout, on users speed and accuracy. We also valued the subjective satisfaction of the users about this assistance to visual exploration.

Three experimental studies showed that multimodal presentations of the target facilitate and improve users performances, considering both speed and accuracy. They also showed that visual exploration strategies, without any oral message, depend on the spatial organization of the informations displayed.

Table des matières

Chapitre 1

Présentation générale

Les techniques interactives de visualisation d'informations ont vu le jour et ont été commercialisées au début des années 90. Elles visent principalement à faciliter aux utilisateurs l'accès à l'information graphique ou multimédia. Elles ont connu un essor considérable en raison, principalement, du volume important des travaux de recherche publiés en psychologie et en neurosciences sur la perception visuelle ou le système visuel humain. Parmi les très nombreuses publications mensuelles sur la perception visuelle, beaucoup trouvent une application dans l'affichage efficace d'informations. À noter qu'il existe d'autres approches de la visualisation comme l'approche artistique, l'approche graphique [Bertin, 1983], ou encore, l'algorithmique géométrique. Mais la seule approche susceptible de fournir des règles de conception efficaces de visualisations interactives est celle basée sur la perception visuelle humaine, et plus particulièrement, les principes ascendants (*bottum-up*) de la vision [Ware, 2004].

Dès lors, de nombreuses techniques de visualisation ont été proposées, comme par exemple, les vues zoomables et les interfaces multiéchelles [Furnas et Bederson, 1995], ou encore les vues hiérarchiques [Pirolli et Card, 1995]. Il existe de nombreuses publications portant sur la conception et l'implémentation de grands ensembles d'informations. On peut citer, à titre d'exemple, l'interface "intelligente" GeoSpace, dont l'affichage s'adapte progressivement en fonction de requêtes formulées par l'utilisateur. Son implémentation repose sur la combinaison de techniques visuelles de conception (cf. la typographie, l'utilisation de couleurs et de la transparence) et d'un mécanisme simple d'apprentissage qui permet au système de supporter des activités comme l'exploration de visualisations complexes d'informations [Lokuge et Ishizaki, 1995]. On peut citer également DEVise, un système qui permet à l'utilisateur de développer, hiérarchiser, et partager des présentations visuelles de vastes ensembles de données [Livny *et al.*, 1997]. En outre, de nombreux travaux portant sur la recherche, ou attention visuelle, au sein de représentations complexes d'informations ont été publiés en psychologie ou en neurosciences, dans le cadre d'une approche perceptive [Underwood, 1998; Doll et Home, 2001; Léger *et al.*, 2003]. En revanche, il n'existe que très peu d'études ergonomiques publiées sur l'efficacité de la recherche visuelle dans de telles organisations spatiales.

Par ailleurs, avec l'avènement du multimédia, un nombre croissant d'applications n'est plus purement graphique, mais tend à présenter simultanément information visuelle et information

verbale, écrite ou orale. La plupart des visualisations d'informations combinent désormais image et langage parlé. On parle, dans ce cas, d'interaction Homme-Machine multimodale parole + présentation visuelle.

Mais quel peut être l'apport spécifique des messages verbaux à l'interaction Homme-Machine au sein d'organisations spatiales complexes, sachant que les visualisations d'informations véhiculent davantage d'informations pour les utilisateurs que les autres médias combinés ? En effet, la vision est le sens qui fournit aux utilisateurs plus d'informations que les autres sens réunis [Ware, 2004]. Plus particulièrement, comment combiner information visuelle et information verbale au sein d'une même intervention du système, afin de faciliter l'interaction au sein de visualisations d'informations dans un contexte d'interaction Homme-Machine multimodale ?

La parole est considérée comme le mode le plus naturel, le plus élaboré et le plus complet de communication humaine, en particulier pour ce qui est des échanges d'informations. En outre, les messages verbaux doivent être préférés aux présentations visuelles pour véhiculer des concepts abstraits, ou des concepts logiques, ou encore des conditions d'utilisation. En revanche, les images doivent être préférées aux messages verbaux pour véhiculer les détails d'une scène ainsi que ses propriétés graphiques.

L'objectif de notre recherche est de concevoir de nouvelles modalités et formes d'interaction Homme-Machine multimodale qui exploitent au mieux l'enrichissement de l'interaction que permet l'intégration de la parole aux modalités de sortie actuelles, à savoir les affichages graphiques. Nous avons donc choisi de centrer notre recherche sur l'évaluation de l'apport spécifique des messages verbaux, en tant que mode d'expression complémentaire des présentations visuelles, dans un contexte d'interaction Homme-Machine multimodale. En particulier, nous avons choisi d'étudier l'association parole + présentation visuelle au sein d'une même intervention du système, en tant qu'assistance multimodale à l'exploration de visualisations 2D interactives.

Cette étude est motivée, d'une part, par la volonté d'étudier la contribution de messages oraux à l'efficacité de la recherche d'informations dans des affichages denses, et de définir des recommandations ergonomiques qui permettent la conception de messages oraux susceptibles d'améliorer l'efficacité et de faciliter la recherche d'informations dans des scènes visuelles complexes. Elle est motivée, d'autre part, par la volonté de comparer l'efficacité de différentes structures 2D simples pour la recherche d'informations visuelles, et de définir des recommandations ergonomiques qui permettent de favoriser la conception de visualisations 2D interactives efficaces.

Nous avons adopté une approche expérimentale. Le programme expérimental élaboré comporte trois études. Dans chaque étude, la tâche expérimentale proposée aux sujets était de sélectionner, à la souris et le plus rapidement possible, des cibles visuelles dans des affichages denses et complexes. Nous avons choisi de centrer l'étude sur le repérage de cibles :

- soit connues visuellement *a priori* ;
- soit définies verbalement de façon non ambiguë ;

au sein de scènes visuelles :

- abstraites *versus* réalistes ;
- structurées *versus* non structurées.

Le travail se présente comme suit. Le chapitre 2 décrit les motivations de l'étude sur la multimodalité parole + présentation visuelle en tant que mode d'expression du système. Nous

décrivons le contexte au sein duquel l'interaction Homme-Machine multimodale a vu le jour, les utilisations classiques de la parole, en entrée du système notamment, ainsi que les nouvelles formes de multimodalité qui émergent actuellement. Après avoir décrit l'intérêt potentiel de la recherche pour l'interaction avec les visualisations de grands ensembles d'information, nous présentons les apports de la parole aux affichages, et au cas particulier des visualisations 2D interactives.

Le chapitre 3 décrit la démarche adoptée ainsi que les choix méthodologiques sur lesquels se fonde la conception du programme expérimental. Celui-ci vise à déterminer l'influence de la parole, sous forme d'indications orales à caractère spatial, sur l'efficacité de la recherche d'informations, les performances, et la satisfaction d'utilisateurs potentiels dans des tâches de repérage visuel de cibles au sein d'affichages complexes. Des réflexions relatives aux affichages 2D statiques, et l'évaluation de la contribution de la parole en tant qu'assistance multimodale à la détection de cibles, concluent ce chapitre.

Le chapitre 4 décrit la première étude que nous avons menée et qui porte sur les apports de la multimodalité parole + présentation visuelle de la cible à son repérage au sein d'affichages abstraits ou réalistes. Dans cette étude préliminaire, trois types de présentation de la cible ont été expérimentés : la présentation visuelle, la présentation orale et la présentation multimodale, i.e., présentation visuelle et présentation orale simultanées. Deux types d'affichages ont été testés :

- les scènes abstraites, constituées de formes arbitraires comme les formes géométriques, ou de structures graphiques symboliques, comme les cartes ;
- les scènes réalistes constituées de photographies, classées en fonction de leur thème comme suit : objets complexes, paysages et groupes de personnages.

Lors de l'expérimentation, 18 sujets devaient réaliser des tâches de repérage sur 36 scènes en tout. L'analyse a porté sur les temps de réponse des sujets, leurs erreurs, et les informations fournies par les questionnaires et les entretiens post-expérience auxquels ils ont participé. Des analyses comparatives quantitatives ont été menées sur les performances des sujets en fonction des modes de présentation de la cible, et des types de scènes qui leur étaient présentés.

Compte tenu des résultats de l'analyse qualitative des données recueillies au cours de la première étude expérimentale, nous avons choisi de centrer les travaux de la deuxième étude sur l'analyse de l'influence de la structure visuelle des scènes sur l'efficacité de la recherche d'informations, en particulier celle du repérage de cibles avec ou sans l'assistance d'indications orales à caractère spatial. Cette étude fait l'objet du chapitre 5 et a également été conçue dans le cadre d'une approche expérimentale. Le protocole expérimental est, dans l'ensemble, semblable à celui adopté pour l'étude préliminaire. Les scènes présentées aux sujets étaient constituées d'un nombre constant d'éléments (30) formant une collection. Quatre types d'organisation spatiale des affichages ont été testés :

- les affichages non structurés ;
- les affichages structurés en ellipse ;
- les affichages structurés en matrice ;
- les affichages à structure radiale.

Pour pouvoir effectuer des traitements statistiques pertinents sur les données quantitatives recueillies (temps de réponse et erreurs), le nombre de sujets a été augmenté (24) et le nombre

des conditions a été réduit à deux : présentation visuelle ou multimodale de la cible. En outre, le nombre de tâches de repérage a été augmenté (240 images en tout) et les sujets ont traité les mêmes images dans chacune des deux conditions, soit 120 par condition et 30 par structure. Cette étude expérimentale s'inscrit dans le projet pluridisciplinaire Micromegas [Micromégas, 2003] qui porte sur le conception et l'évaluation d'approches multiéchelles pour la navigation dans les masses de données familières[1].

La mise en œuvre du protocole expérimental adopté imposait, pour être réalisable facilement, l'informatisation de la création des scènes visuelles présentées aux sujets lors de l'expérimentation, de même que l'informatisation de la création des différentes séquences d'images présentées aux sujets. Nous avons donc constitué manuellement une base de données de photographies[2] et automatisé la construction des images, en particulier la disposition des photographies dans les différentes structures, ainsi que l'ordre d'affichage des images pour les différents sujets. Ces développements importants sont également présentés dans le chapitre 5.

Le chapitre 6 décrit la dernière étude qui est centrée sur les stratégies d'exploration visuelle adoptées par les utilisateurs pour les tâches de repérage de cibles. Le protocole est, dans l'ensemble, semblable à celui adopté pour la deuxième étude. Il n'y a plus qu'un seul mode de présentation des cibles : la présentation visuelle. Le nombre de sujets a été réduit (10), de même que le nombre de scènes (120). Les structures des affichages sont les mêmes que lors de la deuxième étude. Comme dans les deux études précédentes, les performances des sujets sont évaluées en termes de temps et de précision de la sélection des cibles. Un eye-tracker permet d'enregistrer les fixations oculaires lors de chaque détection de cible. Ainsi, nous dressons des parcours oculaires types par sujet et en fonction de l'organisation spatiale des items contenus dans les scènes visuelles.

Le dernier chapitre présente, d'une part, les conclusions relatives à l'évaluation de la multimodalité parole + présentation visuelle en tant qu'assistance au repérage visuel de cibles et, d'autre part, l'influence de la structure spatiale des organisations 2D interactives testées sur l'efficacité de l'interaction en termes de rapidité et de précision des participants aux expérimentations. Nous discutons les résultats obtenus, notamment ceux relatifs à la troisième étude expérimentale qui porte, principalement, sur les stratégies d'exploration visuelle adoptées par les participants.

[1] Les partenaires sont, outre l'INRIA-Lorraine (projet commun MErLIn), l'INRIA-Futurs (projet commun In-Situ), le PLM (Laboratoire Mouvement et Perception, UMR CNRS, Marseille).

[2] Actuellement, 6000 images.

Chapitre 2

Motivations de l'étude

Ce chapitre présente le cadre dans lequel s'inscrit notre recherche. Il s'agit de l'étude de la multimodalité parole + graphique en sortie du système dans un contexte d'interaction Homme-Machine multimodale. Plus précisément, il s'agit d'étudier l'assistance orale à la navigation dans des ensembles d'informations visuelles. Après avoir défini ce qu'est la multimodalité en sortie du système et décrit le contexte scientifique dans lequel se situe notre étude, nous présentons les objectifs ainsi que le sujet de recherche. Enfin, nous montrons quels sont les intérêts de cette étude, avant de conclure par un récapitulatif des notions et réflexions abordées dans ce chapitre.

2.1 Multimédia et multimodalité : définitions

2.1.1 Média *versus* modalité

D'après [Coutaz et Caelen, 1991], [Maybury, 1993; Maybury, 2001] et [Bernsen, 1993; Bernsen, 1994], entre autres, il convient d'attribuer aux termes *média* et *modalité* deux définitions distinctes. Le mot *média* fait référence aux canaux matériels ou logiciels qui véhiculent l'information. En revanche, le mot *modalité* désigne l'association entre un média et les processus d'interprétation nécessaires à la transformation des représentations physiques de l'information en des messages ou symboles porteurs de sens. Il convient, en outre, de distinguer les médias et modalités en entrée, i.e., qui véhiculent l'information de l'utilisateur vers le système, des médias et modalités en sortie du système, i.e., qui véhiculent l'information du système vers l'utilisateur. Dans le cadre par exemple de la saisie au clavier, le clavier est le média en entrée, et la modalité est le texte, porteur d'informations. De la même façon, dans le cadre de la consultation de courrier électronique, le média en sortie est l'écran alphanumérique, tandis que la modalité est la vision et l'interprétation des caractères affichés.

En d'autres termes, du point de vue des médias et modalités en sortie du système uniquement, nous considérons avec Maybury, que le terme *média* désigne les supports qui permettent de transmettre des informations sous forme textuelle, graphique, audio ou vidéo. La notion de support de l'information comprend tous les dispositifs physiques de présentation de l'information nécessaires à l'interaction, comme haut-parleur et écran. Par *mode* ou *modalité*, nous faisons

référence aux capacités perceptives de l'utilisateur mises en jeu lors de l'interaction pour traiter l'information véhiculée par le système, en sortie, à savoir la vision, l'ouïe et le toucher. [Maybury, 2001]

Dans la définition plus générale de Vernier et Nigay, un *média* désigne le support de l'information échangée plutôt que son contenu. On distingue les médias en entrée, comme le clavier, la souris, les systèmes de commande orale et gestuelle, des médias en sortie comme les écrans, les manettes à retour d'effort et les haut-parleurs. Une *modalité*, en entrée comme en sortie, est le moyen de communication qui met en œuvre un média et un langage d'interaction. [Vernier et Nigay, 2000]

De ces nombreuses définitions des termes média et modalité, découlent les notions d'interaction multimodale et de présentation multimédia. Ce sont ces notions que nous allons décrire ci-dessous.

2.1.2 Système multimédia *versus* système multimodal

D'après Vernier et Nigay, l'utilisation simultanée de plusieurs médias dans un contexte d'interaction Homme-Machine est appelée *interaction multimédia* [Vernier et Nigay, 2000]. Dans un tel contexte, le multimédia en sortie désigne simplement la présentation simultanée d'informations sur plusieurs supports qui mettent en jeu plusieurs capacités perceptives de l'utilisateur. De plus, d'après Coutaz et Caelen dans [Coutaz et Caelen, 1991], une application multimédia permet d'acquérir, de sauvegarder et de restituer des informations exprimées dans plusieurs médias. Par extension, on parle de système multimédia lorsque celui-ci est capable d'acquérir, restituer, mémoriser et organiser des informations présentées sur des supports différents. D'après Coutaz et Caelen dans [Coutaz et Caelen, 1991], un système, pour être multimodal, doit pouvoir accepter en entrée les modalités humaines telles que gestes, écriture et langage naturel. Il doit être en outre capable de "comprendre" le contenu des messages en provenance de l'utilisateur et de l'application.

Au total, un système multimédia doit permettre d'acquérir, restituer, mémoriser et organiser les informations véhiculées par plusieurs médias ou supports de l'information. En revanche, la notion de système multimodal sous-entend que le système est capable de "comprendre" le contenu des informations exprimées par l'utilisateur à l'aide de plusieurs médias.

On peut étendre cette définition de la multimodalité centrée sur les entrées du système pour qu'elle inclue également l'expression multimodale du système. On parle alors de multimodalité en sortie. Il s'agit de la capacité, pour le système, de générer à partir d'un ensemble d'informations des messages qui expriment ces informations à l'aide de plusieurs médias et modalités. On peut citer à titre d'exemple, la multimodalité parole + graphique, ou encore la multimodalité texte + graphique, etc.

Dans toute la suite, nous parlerons de *multimodalité parole + graphique en sortie* pour désigner les messages émanant du système qui associent simultanément :

- une présentation visuelle classique (i.e., la modalité visuelle) via un écran d'ordinateur (i.e., le média associé) ;

et :

- un message oral (i.e., la modalité sonore) via des haut-parleurs (i.e., le média associé).

2.1.3 Hypertexte, hypermédia et multimédia

On peut s'interroger sur les relations entre les concepts exprimés par hypermédia et multimédia qui ne diffèrent, dans la forme, que par la présence, devant le terme média, du préfixe "hyper-" ou "multi-".

D'après [Münz, 2001], l'*hypertexte* est du texte, à ceci près que la notion d'hypertexte ne sous-entend pas la contrainte de séquentialité présente implicitement dans le mot texte, i.e., lire du début à la fin, page après page. Le préfixe hyper- signifie qu'il existe une manière d'organiser le texte en unités d'information distinctes de façon intelligente et de permettre à l'utilisateur de choisir un ordre de parcours adapté à ses besoins et intérêts.

Le prolongement logique de l'essor de l'hypertexte en informatique grâce à la diversification des médias en sortie (images, vidéos, sons, etc.) est l'*hypermédia.* L'hypermédia sous-entend que les informations véhiculées par les médias sont organisées, structurées visuellement de façon intelligente.

La différence entre multimédia et hypermédia tient aux préfixes hyper- et multi-. Le préfixe multi- signifie plusieurs médias en même temps, sans intégrer la notion de structuration des unités d'informations. En ce sens, l'hypermédia est du multimédia structuré.

2.1.4 Classifications des modalités

L'émergence de nouvelles applications graphiques interactives, mais surtout de nouvelles techniques d'interaction, entraîne de nouvelles modalités, donc par suite, de nouvelles formes de multimodalité, ou combinaisons de modalités. Pour caractériser les combinaisons possibles entre modalités, de nombreuses classifications - on parle aussi de taxonomies - ont été proposées.

Dans [Coutaz et Caelen, 1991], la multimodalité est caractérisée en fonction des stratégies d'interaction offertes, c'est-à-dire en fonction de l'utilisation des différentes modalités d'expression offertes (en entrée comme en sortie). D'un point de vue sémantique, les modalités peuvent être combinées (synergie) ou non (utilisation concurrente). D'un point de vue syntaxique (i.e., temporel), leur usage peut être simultané ou alterné. De plus, [Coutaz *et al.*, 1995] définit quatre propriétés, les propriétés CARE, pour caractériser l'interaction multimodale : le type d'information exprimé par chaque modalité (i.e., équivalence *versus* assignation) et la contribution sémantique de chaque modalité (i.e., redondance *versus* complémentarité).

Enfin, la taxonomie des modalités de Bernsen présente un intérêt particulier, en raison de son exhaustivité [Bernsen, 1994]. Chaque modalité en sortie y est caractérisée en fonction des propriétés, ou traits binaires, suivants :

- linguistique (e.g., hiéroglyphes, discours) *versus* non-linguistique (e.g., photographies, signaux sonores) ;
- analogique (e.g., langage parlé, cartes géographiques) *versus* non-analogique (e.g., mots, langages de programmation) ;

- arbitraire (e.g., icones informatiques) *versus* non-arbitraire (e.g., barres de défilement, fenêtres) ;
- statique (e.g., dessins, diagrammes, fenêtres) *versus* dynamique (e.g., films, compteurs Geiger, Braille).

De plus, cette taxonomie distingue trois médias d'expression pour le système ; ce sont le graphique, le son et l'haptique qui sollicitent différentes capacités sensorielles de l'utilisateur (la vue, l'ouïe et le toucher, respectivement). Elle comprend quatorze modalités visuelles, sept modalités sonores et sept modalités tactiles. C'est sur la base de cette taxonomie, en raison de son exhaustivité, que nous avons débuté notre recherche sur la multimodalité parole + présentation visuelle en sortie du système[3].

2.2 Contexte scientifique

Il y a 10 ou 15 ans, les P.C. ne fournissaient, comme moyens d'interaction en entrée, que la manipulation directe d'objets stéréotypés comme les icones, les barres de défilement, etc. On parlait d'interfaces graphiques utilisateur, ou en anglais de GUIs pour Graphical User Interfaces. Le terme GUI[4] désigne les affichages comportant essentiellement du texte et des fenêtres (Windows).

Désormais, il est possible d'en faire bien plus. En entrée, il est possible pour l'utilisateur d'interagir avec le système :

- par la parole, grâce aux systèmes de reconnaissance vocale qui sont présents même dans les téléphones mobiles ;
- ou par geste de la main, par exemple, sur un écran tactile avec un doigt ou un stylo numérique [Guyomard *et al.*, 1995; Siroux *et al.*, 1997] ou, dans l'espace, avec un gant numérique [Ehrenmann *et al.*, 2001] ;
- ou encore, par le regard, notamment dans le domaine médical [Blois *et al.*, 1999].

En sortie, les réactions de la machine aux commandes ou actions des utilisateurs sont un peu plus limitées en raison principalement du manque de dispositifs de sortie [Vernier et Nigay, 2000]. Elles peuvent être de nature :

- tactile : le média associé est, par exemple, une manette à retour d'effort, dispositif utilisé surtout dans les jeux, les visites virtuelles d'environnements 3D, ou les simulateurs ;
- sonore (messages ou bips sonores) : le média associé est un haut-parleur, les haut-parleurs constituant la seule alternative aux affichages écran classiques ;
- visuelle : le média associé est l'écran ; on parle de présentation visuelle ou de présentation graphique.

On peut citer à titre d'exemple de présentation graphique, les organisations visuelles en 3D, comme les "Data Mountains" [Robertson *et al.*, 1998] pour visualiser des données, les logiciels de C.A.O., les jeux, etc. Il est également possible d'interagir avec des environnements logiciels de

[3] cf. infra chapitre 4 paragraphe 4.1.2 page 28.

[4] À noter que, contrairement à leur nom, les GUIs ne contiennent que peu de "graphique" au sens premier du terme. En réalité, le terme "graphique" signifie que les interfaces utilisateur requièrent des écrans offrant des possibilités graphiques (cf. les travaux de Krause dans [Krause, 1997]).

réalité virtuelle tels que les "Reality Centers" ou encore les "Workbenches". De façon pratique, il n'existe plus de limite à la taille des affichages, du moins d'un point de vue technologique.

C'est dans ce contexte que l'interaction Homme-Machine multimodale a vu le jour, parallèlement à l'émergence de nouvelles applications qui offrent à l'utilisateur la possibilité d'échanger des informations avec le système :

- selon diverses modalités d'interaction, comme la parole ou le geste ;
- via divers médias, comme haut-parleurs, ou manettes à retour d'effort.

En effet, l'intégration de plusieurs médias et modalités au sein d'une interface graphique utilisateur fait de l'interaction Homme-Machine multimodale - en particulier, celle qui associe la parole, soit au geste en entrée, soit à l'image en sortie - un thème de recherche en plein essor. Il convient de noter que la plupart des recherches actuelles portent sur le multimédia (cf. les travaux de Mabury dans [Maybury, 1993]) et non sur la multimodalité en sortie. Enrichir les moyens d'expression multimodale du système a suscité peu d'intérêt de la part des chercheurs, en dépit de la distinction entre les définitions de multimédia et de multimodalité[5].

Nous avons choisi de faire porter nos recherches sur l'association parole + graphique en tant que mode d'expression du système dans un contexte d'interaction Homme-Machine multimodale, c'est-à-dire dans un contexte où le système dispose d'une représentation des informations qu'il doit transmettre à l'utilisateur. Après avoir décrit les formes possibles d'utilisation de la parole en IHM, nous montrerons en quoi la multimodalité parole + graphique en sortie est un thème de recherche qui présente beaucoup d'intérêt pour enrichir et faciliter les échanges entre l'utilisateur et la machine. Enfin, nous présenterons les objectifs de cette recherche.

2.2.1 Utilisations classiques de la parole en IHM

Initialement, une modalité alternative

Les premières interfaces utilisateur intégrant la parole en sortie du système furent commercialisées au cours des années 80. Elles mettaient en œuvre une interaction exclusivement orale entre l'homme et la machine. En effet, le recours à la parole y était motivé par la volonté d'offrir aux utilisateurs une modalité d'interaction de substitution à l'utilisation des médias et modalités classiques, les présentations visuelles. On peut citer, à titre d'exemple, les services téléphoniques.

Désormais, la parole en sortie du système est le plus souvent utilisée pour pallier les insuffisances des canaux d'échange dans des contextes d'utilisation où ne sont disponibles que des affichages de taille réduite. On peut citer, par exemple, l'informatique mobile, et plus particulièrement les P.D.A.[6] et les wearable computers [Baber, 2001], où la modalité sonore est utilisée pour développer l'information affichée (cf. la redondance au sens des propriétés CARE [Coutaz *et al.*, 1995]). Mais on utilise également la parole dans des contextes où l'utilisation des présentations visuelles classiques est impossible en raison :

- soit de la mobilisation des capacités motrices de l'utilisateur par d'autres activités simultanées multiples, par exemple dans le cadre du pilotage d'avion ;

[5] cf. supra paragraphe 2.1 à la page 5.
[6] Personal Digital Assistant.

- soit de la mobilisation des capacités sensorielles de l'utilisateur, par exemple lors de la conduite de véhicule [De Vries et Johnson, 1997] ;
- soit d'un handicap moteur et/ou perceptif de l'utilisateur, par exemple, pour permettre aux déficients visuels d'accéder à l'outil informatique [Yu et Brewster, 2003].

À l'heure actuelle, on tente d'enrichir l'interaction Homme-Machine en associant la parole à d'autres modalités d'interaction disponibles, comme par exemple, le geste de désignation en entrée, ou le graphique en sortie.

En entrée du système

Les formes de multimodalité en entrée associant la parole au geste de désignation ont fait l'objet de nombreuses études, que ce soit d'un point de vue logiciel, ou d'un point de vue ergonomique. Les principaux travaux sur les aspects logiciels, issus pour un grand nombre d'entre eux de la communauté scientifique francophone, comprennent notamment : [Bourget, 1992; Baudel et Braffort, 1993; Nigay et Coutaz, 1993; Guyomard *et al.*, 1995; Siroux *et al.*, 1997]. Les principaux travaux sur les aspects ergonomiques, qui ont suscité l'intérêt de la communauté scientifique internationale, comprennent notamment : [Hauptmann et McAvinney, 1993; Catinis et Caelen, 1995; Oviatt *et al.*, 1997] et au LORIA [Mignot et Carbonell, 1996; Robbe *et al.*, 2000].

La plupart de ces études associe la parole à des gestes 2D de désignation, réalisés avec le doigt ou un stylo à la surface d'un écran tactile. C'est le cas des travaux de Guyomard *et al.* et Siroux *et al.* [Guyomard *et al.*, 1995; Siroux *et al.*, 1997] qui portent sur l'observation d'utilisateurs d'une base de données géographiques et touristiques qui supporte des requêtes, à la fois, tactiles et linguistiques[7]. Cette forme de multimodalité a également fait l'objet des travaux suivants : [Hauptmann et McAvinney, 1993; Catinis et Caelen, 1995; Mignot et Carbonell, 1996; Oviatt *et al.*, 1997; Robbe *et al.*, 2000]. Les études publiées associent plus rarement la parole à des gestes de désignation 3D. C'est le cas des travaux de Baudel et Braffort [Baudel et Braffort, 1993] qui proposent une application de navigation dans un système hypertexte au moyen de gestes 3D de la main, réalisés avec un gant numérique.

Enfin, les travaux de Bourguet [Bourget, 1992] substituent la parole, en langage naturel ou quasi naturel, associée à des gestes de désignation réalisés avec la souris, à la manipulation directe. À noter également, les travaux de Nigay et Coutaz [Nigay et Coutaz, 1993] portant sur la fusion des modalités au sein des systèmes multimodaux VOICEPAINT et NOTEBOOK.

2.2.2 Vers de nouvelles formes de multimodalité

Si la multimodalité qui associe, en entrée, la parole à d'autres modalités a suscité de nombreux travaux, en revanche, l'association de la parole au graphique ou au texte, en sortie, n'a motivé que peu d'études. Il n'existe que peu de travaux scientifiques publiés sur les possibilités offertes par la parole en tant que modalité d'expression du système, complémentaire du graphique ou du

[7]Les dispositifs physiques et logiciels associés comprennent, respectivement, un écran tactile et un système de reconnaissance vocale.

texte. En outre, les rares travaux publiés en informatique sur le rôle de la parole associée au graphique, animé ou non, analysent sa contribution dans des contextes de présentation multimédia d'informations, et non dans une situation d'interaction Homme-Machine multimodale[8].

On peut citer, à titre d'exemple, les recherches ergonomiques qui portent sur le rôle de la parole dans les présentations multimédias, ou encore les travaux publiés qui évaluent l'apport de la parole à la génération automatique de telles présentations [André et Rist, 1993; Pan et McKeown, 1996; Faraday et Sutcliffe, 1997; Maybury, 2001].

Les autres travaux de recherche publiés sur l'association de la parole au graphique sont issus de la psychologie cognitive, comme par exemple, les travaux de Kalyuga *et al.* qui portent sur la charge cognitive supplémentaire introduite par des messages sonores, les messages sonores étant redondants à un texte affiché à l'écran. Dans [Kalyuga *et al.*, 1999], deux expérimentations sont présentées : l'une montrant que les messages sonores interfèrent avec l'apprentissage, l'autre montrant que les couleurs permettent de réduire la charge cognitive dans un contexte de recherche dans un texte. On peut citer également les travaux publiés au cours des années 70, sur la représentation linguistique ou la mémoire verbale [Paivio, 1977]. Quoi qu'il en soit, il n'existe à notre connaissance que peu de travaux publiés, même en Neurosciences, sur les interférences provoquées par l'association de la parole à des affichages graphiques, ou même sur les apports potentiels pour l'interaction de cette forme de multimodalité en sortie.

Ce manque d'intérêt des chercheurs en ergonomie de l'interaction Homme-Machine ou en génie des interfaces utilisateurs, illustré précédemment, peut éventuellement s'expliquer par la confusion fréquente entre présentation multimédia et réaction multimodale aux actions et commandes de l'utilisateur [Maybury, 1993]. Alors même que les problèmes d'ordre logiciel sont résolus depuis longtemps, l'étude de la contribution de la parole à l'efficacité des interventions du système (messages d'erreurs, comptes rendus d'exécution, aide en ligne) en est encore à ses débuts. Or, on ne peut raisonnablement proposer à l'utilisateur une interface avec laquelle il pourrait interagir oralement mais qui resterait muette. C'est donc, dans ce contexte, que nous avons choisi de nous intéresser, d'abord et en priorité, à la multimodalité parole + graphique en sortie du système, et d'étudier les apports éventuels de la parole en tant que mode d'expression complémentaire du graphique dans un contexte d'interaction entre la machine et l'utilisateur.

2.3 Objectifs et sujet de recherche

Comme nous l'avons vu dans la section précédente, l'interaction Homme-Machine multimodale est un thème de recherche en plein essor, suite à l'avènement de nouvelles interfaces utilisateur qui offrent désormais, aux professionnels comme au grand public, la possibilité d'interagir avec la machine selon diverses modalités, par le biais de divers médias. Étudier l'association de la parole au graphique dans une même intervention du système, et ce, dans un contexte d'interaction Homme-Machine multimodale, est un sujet de recherche qui présente un intérêt particulier actuellement, notamment si l'on veut :

- enrichir les échanges entre le système et l'utilisateur ;

[8] cf. supra paragraphe 2.1 à la page 5.

- accroître la facilité d'utilisation des interfaces utilisateur classiques, à savoir l'interaction avec les applications graphiques.

D'une part, l'association de la parole aux modalités de sortie actuelles, essentiellement l'affichage graphique, devrait enrichir l'interaction en raison des modes de perception spécifiques qu'offrent les messages sonores, par rapport aux affichages. En outre, la parole est considérée, à juste titre semble-t'il, comme le mode le plus naturel de communication humaine, en particulier pour ce qui est des échanges d'informations ; l'intégration de cette modalité aux interfaces des applications, qu'elles soient destinées au grand public ou aux professionnels, doit donc permettre d'accroître sensiblement la diffusion de l'informatique dans la société, et promouvoir une société de l'information accessible à tous.

D'autre part, si la manipulation directe suffit pour interagir avec les applications classiques qui, en sortie, combinent exclusivement graphique et texte, la diversification des contextes d'utilisation font de la parole une modalité d'interaction utile, voire indispensable. En effet, on peut citer à titre d'exemple l'émergence de l'informatique embarquée ou portable qui entraîne de nouvelles formes d'utilisation de l'outil informatique où l'interaction ne peut se passer de la parole en sortie du système. Considérons les systèmes de positionnement et de navigation par satellite[9] qui permettent un positionnement absolu ou relatif en tout point du globe. Ils sont plus connus sous le nom de G.P.S.[10]. Désormais, ces dispositifs sont accessibles au grand public sous forme portable ou embarquée. C'est le cas dans les véhicules automobiles où les indications d'itinéraire sont orales, puisque la vision est déjà sollicitée par la conduite automobile.

Enfin, il est intéressant de constater que les concepteurs d'applications surchargent de plus en plus les affichages, en augmentant la partie visuelle des interfaces, sans pour autant se passer des composantes sonores (paroles, musique, bips sonores) ; et ce, en dépit du manque de travaux publiés sur l'augmentation éventuelle de la charge cognitive induite par l'intégration de la modalité sonore à ces interfaces. En effet, associer parole et graphique dans une même intervention du système soulève des problèmes d'ordre ergonomique qui n'ont pas encore été abordés, mis à part peut-être pour des catégories spécifiques d'utilisateurs ou pour des contextes spécifiques d'utilisation[11] [Wang *et al.*, 2000]. L'ajout d'autres médias ou modalités, tels que l'animation, crée des difficultés de conception supplémentaires.

Il est donc utile de fournir aux concepteurs des guides et des recommandations efficaces pour la conception d'interfaces utilisateur qui offrent des médias supplémentaires appropriés en sortie, comme par exemple, la génération automatique de signaux auditifs. Les nouvelles modalités faisant intervenir ces médias pourront constituer une alternative à la manipulation directe (cf. les travaux de [Shneiderman, 1983]). En effet, la manipulation directe, ne permet des réactions du système qu'exclusivement visuelles. En ce sens, elle constitue un facteur de limitation pour l'interaction.

Ce sont les raisons pour lesquelles nous avons choisi d'étudier et de clarifier les apports potentiels de la parole aux interfaces utilisateur classiques, en tant que mode d'expression complémentaire du graphique et du texte. Nous avons choisi comme application la recherche d'informations

[9] Ensemble de satellites artificiels particuliers, dont les plus connues sont les éphémérides.
[10] Global Positioning System.
[11] cf. supra paragraphe 2.2.1 page 9.

au sein de visualisations de grands ensembles d'informations, ou sur le Web ou encore dans les banques d'images. D'une part, il s'agit d'applications en plein essor touchant, à la fois, le grand public comme les spécialistes. D'autre part, étant donné la complexité de la tâche, la charge cognitive induite par la multimodalité peut être considérée comme négligeable.

L'objectif peut se résumer comme suit. Nous souhaitons concevoir de nouvelles modalités et formes d'interaction Homme-Machine multimodales qui exploitent au mieux l'enrichissement de l'interaction que permet l'intégration de la parole aux modalités de sortie actuelles. Plus particulièrement, nous avons retenu pour objectif d'étudier l'assistance orale à la navigation dans les ensembles d'informations visuelles, en vue de contribuer à enrichir les travaux de recherche publiés sur la mise en œuvre de la multimodalité parole + présentation visuelle.

2.4 Intérêt potentiel de la recherche

On assiste aujourd'hui à une surcharge croissante des affichages en raison, d'une part, de l'augmentation du volume des informations échangées au cours de l'interaction, et d'autre part, de la multiplication des fonctionnalités logicielles offertes aux utilisateurs. Au fil des versions, les logiciels grand public et professionnels doivent supporter toujours plus de fonctionnalités dans le but de satisfaire la diversité grandissante des utilisateurs et donc, de leurs besoins et des tâches qu'ils souhaitent réaliser. Ce phénomène conduit à une multiplication des fenêtres et des barres d'outils affichées simultanément et des icones au sein de celles-ci. Par ailleurs, l'évolution des dispositifs d'affichage, les progrès des techniques d'affichage graphique ont suscité le développement de nouvelles fonctions d'interaction. Ces nouvelles fonctionnalités - vues d'ensemble, zoom, filtrage, ou encore présentation des relations entre les objets graphiques - ont fait l'objet d'une classification décrite dans [Shneiderman, 1996]. Désormais, le flot d'informations visuelles[12] qu'il est nécessaire de transmettre à l'utilisateur est en constante augmentation [Krause, 1997].

2.4.1 Limites des techniques de visualisation interactive d'informations

Pour permettre à l'utilisateur d'accéder rapidement à une vaste quantité de données visuelles[13] tout en lui fournissant davantage d'indices sur les relations entre les objets en sollicitant moins sa mémoire et en diminuant les charges cognitives, des techniques de visualisation d'informations ont été élaborées [Ware, 2004]. Elles reposent sur des travaux déjà anciens, comme la cartographie ou au $20^{ème}$ siècle, la sémiologie graphique, ou la "Graphique" [Bertin, 1981; Bertin, 1983]. Les apports de l'informatique ont permis d'améliorer ces techniques et d'aboutir à de nouvelles représentations graphiques de données, en vue de "rendre compréhensibles les phénomènes n'ayant pas de représentation standard". [Micromégas, 2003]

[12] Affichages en 2 ou 3 dimensions, animations, etc.

[13] cf. la société de l'imagerie.

2.4.2 Interaction avec les visualisations de grands ensembles d'informations : l'existant

La fin des années 90 a vu émerger un grand nombre de types de visualisation de grands ensembles d'informations, comme les visualisations scientifiques (graphes, diagrammes, graphes multidimensionnels, etc.), les diagrammes de noeuds et de liens, les arbres hyperboliques, les Tree-Maps [Card et Mackinlay, 1997]. Plusieurs taxonomies ont été proposées pour caractériser ces nombreuses visualisations. On peut citer, entre autres, la taxonomie des tâches de visualisation de Chuah et Roth [Chuah et Roth, 1996] qui distingue, au sein d'une classification hiérarchique, trois types d'opérations dans les visualisations interactives (B.V.I. pour Basic Visualization Interaction) : les opérations graphiques qui modifient l'apparence des visualisations, les opérations sur les données qui manipulent les données encodées au sein des visualisations, et enfin les opérations sur les ensembles qui créent et manipulent des ensembles de données. À noter que les données sont exprimées grâce à des objets graphiques. Par exemple, la manipulation d'objets graphiques n'entraîne pas leur modification en termes des données qu'ils représentent.

Parmi les opérations graphiques, on peut citer l'encodage de données qui fait référence aux opérations qui transforment la cartographie existant entre les données et leur représentation graphique. Parmi les opérations sur les données, on peut citer la suppression d'une donnée, qui supprime également sa représentation graphique au sein de la visualisation. Parmi les opérations sur les ensembles, on peut citer la création d'ensembles au sein de la représentation, qui entraîne une nouvelle classification de l'information. On peut citer également la matrice des tâches de Shneiderman basée sur le type des données contenues dans les visualisations [Shneiderman, 1996]. Les tâches, d'un haut niveau d'abstraction contiennent, entre autres, les vues d'ensemble, le zoom, les filtres. Les types de données comprennent les données linéaires (e.g., le texte), les données planes (e.g., les cartes), les données 3D (e.g. les objets du monde réel comme les molécules), les données temporelles, multidimensionnelles, les arbres et les réseaux.

Lorsque l'utilisateur recherche visuellement une information dans un grand ensemble, ou banque, d'informations, il fait appel au modèle mental qu'il s'est construit de son organisation ou extrait directement l'information d'une représentation de l'organisation de la banque fournie par le concepteur. Ainsi, il préfère souvent "naviguer" dans la banque d'informations visuelles, plutôt qu'effectuer une recherche par des outils d'indexation. À noter que nous avons pris soin, dans nos travaux de recherche, de distinguer la recherche visuelle d'un item, de la recherche par outils d'indexation, comme la recherche par mots-clés. En outre, il convient de distinguer également (cf. la section "Navigation dans les données familières" dans [Micromégas, 2003]) :

- la recherche visuelle d'un élément : l'élément recherché est caractérisé par des propriétés visuelles (couleur, forme, taille, etc.) ou par des critères que l'on peut exprimer verbalement et qui peuvent être fournis, dans le cas du Web, à un moteur de recherche ;
- la navigation vers un élément : l'élément recherché est caractérisé par un chemin d'accès. Il est "connu" visuellement ou non.

2.4.3 Recherche visuelle d'une information spécifique

Malgré des techniques de visualisation telles que les arbres hyperboliques [Lamping *et al.*, 1995] ou la superposition de vues transparentes [Harisson et Vicente, 1996], la recherche dans les grands ensembles d'informations exige un effort cognitif important de la part des utilisateurs. De plus, l'augmentation exponentielle du nombre d'objets graphiques présents dans les affichages, aussi bien que la mise en relief, ou "pop out", d'informations[14], ou encore la densité d'informations à l'écran, sont autant de facteurs qui ralentissent et rendent fastidieuse la recherche visuelle [Pirolli *et al.*, 2000]. En effet, il est prouvé que la densité d'informations affichées affecte l'attention visuelle, en réduisant la taille du champ visuel utile, ou U.F.O.V. pour Useful Field Of View. On pense même que la recherche visuelle dans l'affichage entier est moins efficace s'il comprend des zones plus denses que d'autres [Drury et Clement, 1978].

2.4.4 Exploration et navigation vers un élément

De nouvelles fonctions d'interaction (zoom, filtrage, vues focalisées) ont été proposées pour explorer les grands ensembles d'informations [Shneiderman, 1996]. Cependant, l'évaluation ergonomique de ces techniques de visualisation interactive reste à faire. On ne dispose actuellement que d'études ponctuelles sur certaines d'entre elles, arbres hyperboliques [Pirolli *et al.*, 2000], superposition de vues transparentes [Harisson et Vicente, 1996] ou vues hiérarchiques multiples [Mukherjea *et al.*, 1995].

D'autre part, la navigation dans des documents hypermédias[15] s'avère souvent longue et fastidieuse, même si on tente de faciliter la recherche d'informations par la structuration et la cartographie des sites [Shipman *et al.*, 1995]. En effet, la navigation a fait la preuve de son inefficacité lorsque l'utilisateur recherche une information qui se situe dans un grand ensemble d'informations [Pirolli *et al.*, 2000]. Plus le nombre de données croît, plus les chemins de navigation deviennent longs. Les tâches de navigation dans les grands ensembles d'information se révèlent donc de plus en plus longues et fastidieuses, et se soldent souvent par un échec.

2.4.5 Apports potentiels de la parole aux affichages

En vue de faciliter l'exploration de visualisations denses et plus généralement la recherche d'informations dans des affichages complexes, nous avons choisi d'étudier et d'évaluer les possibilités d'assistance à cette activité offertes par des messages oraux appropriés.

En outre, l'évaluation ergonomique de l'apport d'informations spatiales d'assistance à l'exploration d'affichages complexes, exprimées oralement, est un sujet de recherche qui n'a pas encore été abordé dans un contexte d'interaction Homme-Machine, bien que l'exploration des scènes compte, avec la lecture, parmi les activités visuelles qui ont suscité le plus de travaux de recherche en psychologie, mais uniquement dans le contexte de tâches expérimentales en laboratoire ; voir par exemple [Findlay et Gilchrist, 1998] et [Henderson et Hollingworth, 1998].

[14] Les informations qui "sautent aux yeux" de l'utilisateur.

[15] cf. supra paragraphe 2.1 page 5.

Les applications potentielles d'une telle étude ne se réduisent pas au domaine de la visualisation interactive de grands ensembles d'informations. L'informatique mobile - les "wearable computers", par exemple - ou embarquée a entraîné également une augmentation du flot d'informations visuelles qu'il est nécessaire de transmettre à l'utilisateur ; il s'agit, entre autres, d'informations concernant son environnement et son itinéraire (cf. la lecture de cartes ou la recherche d'itinéraire). Il en est de même pour les services automatisés comme les pages Web multimédias, dont l'exploration est rendue de plus en plus difficile en raison de l'augmentation des icones, dessins, images, photographies, etc., qu'elles contiennent. La réalité virtuelle [Plesniak et Ravikanth, 1998] ou augmentée [Mackay *et al.*, 1998], ainsi que les environnements 3D [Robertson *et al.*, 1998], sont également concernés.

2.5 Résumé

L'avènement de nouvelles applications, comme la recherche d'informations, la navigation sur le Web, la recherche d'itinéraires, les réalités virtuelles, augmente considérablement la quantité de données graphiques affichées à l'écran. Ce constat du volume et de la densité croissants des informations affichées simultanément explique pourquoi nous avons choisi d'explorer l'interaction Homme-Machine multimodale qui associe la parole au graphique ou à l'image en sortie. L'étude de la contribution de la parole à l'efficacité des interventions du système (messages d'erreurs, compte-rendu d'exécution, aide en ligne) est encore à ses débuts, alors même que les problèmes d'ordre logiciel sont résolus.

En outre, les rares travaux publiés sur l'association de la parole au graphique analysent la contribution potentielle de cette nouvelle modalité dans des contextes de présentation multimédia d'informations, et non dans une situation d'interaction Homme-Machine multimodale.

La parole est considérée comme le mode le plus naturel de communication humaine, en particulier pour ce qui est des échanges d'informations ; l'intégration de cette modalité aux interfaces des applications grand public doit donc permettre d'accroître sensiblement la diffusion de l'informatique dans la société, et promouvoir une société de l'information accessible à tous.

L'objectif est de concevoir de nouvelles modalités et formes d'interaction Homme-Machine multimodale qui exploitent au mieux l'enrichissement de l'interaction que permet l'intégration de la parole aux modalités de sortie actuelles.

Cette étude est motivée par la volonté d'apporter une contribution aux travaux de recherche publiés sur la mise en œuvre de la multimodalité parole + présentation visuelle. Les résultats d'une telle étude en effet, sont susceptibles d'améliorer sensiblement la facilité d'utilisation d'une vaste classe d'applications : les interfaces utilisateur graphiques, actuellement surchargées par l'augmentation croissante de la quantité d'informations affichées à l'écran, notamment les visualisations interactives de grands ensembles d'informations.

Chapitre 3

Démarche et méthode

Nous avons adopté une démarche expérimentale afin de déterminer l'influence d'indications spatiales formulées oralement sur l'efficacité de la recherche d'informations et la satisfaction d'utilisateurs potentiels dans des activités d'exploration visuelle d'affichages complexes. Ce chapitre présente la démarche et la méthode que nous avons adoptées pour atteindre l'objectif visé : définir et évaluer un nouvelle forme d'interaction intégrant la parole aux modalités de sortie actuelles.

Après avoir décrit précisément la forme de multimodalité retenue, nous présentons l'approche expérimentale que nous avons adoptée. Puis, nous décrivons la tâche expérimentale ainsi que la méthode employée, notamment certains choix concernant le type d'affichage que nous avons testé. Nous terminons ce chapitre par la présentation du programme expérimental comportant trois volets.

3.1 Forme de multimodalité retenue

Associer parole et graphique dans une même intervention du système soulève des problèmes d'ordre ergonomique qui n'ont pas encore été abordés. Pour caractériser la forme de multimodalité en sortie que nous allons étudier, nous nous basons sur la taxonomie de Coutaz *et al.* dans [Coutaz *et al.*, 1995].

Une stratégie proposée pour associer la parole au graphique en sortie, consiste à assigner chaque modalité à un type d'information particulier. Par exemple, la parole est assignée à l'expression d'informations d'aide, et le graphique est assigné aux réactions de l'application aux commandes ou manipulations licites de l'utilisateur. Si cette forme de multimodalité présente l'avantage d'être simple à mettre en œuvre sur les plans logiciel et ergonomique, elle reste cependant d'un intérêt limité. Considérons une application de type tableur, où les informations d'aide ne sont disponibles qu'oralement. Alors, il est clair que, pour l'utilisateur "débutant", la terminologie spécifique aux tableurs constitue un facteur de limitation au bon déroulement de l'interaction. En effet, dans ce cas, l'expression d'aide ne peut se passer de l'affichage graphique pour être compréhensible.

C'est pour cette raison que nous écartons de nos recherches la parole en tant que modalité **assignée** à l'expression d'un type d'information. Dans cette étude, nous nous plaçons délibérément dans un contexte d'interaction Homme-Machine multimodale. Notre objectif n'est pas d'étudier la parole en tant que modalité **équivalente** ou **redondante** d'une autre, et donc capable de la remplacer, mais bien en tant que modalité **complémentaire** d'autres modalités. Dans notre étude, la parole sera complémentaire de la modalité graphique qui sollicite la vision. Complémentarité, assignation, redondance et équivalence font référence ici aux propriétés CARE [Coutaz *et al.*, 1995].

Nous souhaitons évaluer les apports éventuels de la parole à l'interaction Homme-Machine multimodale en tant que modalité d'expression complémentaire du graphique en sortie du système. Plus précisément, la forme de multimodalité, parole + graphique, que nous allons étudier comprend :

- des affichages graphiques qui sollicitent la vision ;
- des messages sonores complémentaires du graphique qui sollicitent l'audition.

3.2 Démarche globale

3.2.1 Choix de l'activité

Nous avons vu au chapitre précédent, paragraphe 2.4.5 page 15, que les apports potentiels de la parole à l'interaction Homme-Machine multimodale pouvaient être mis en évidence dans le domaine de l'exploration de grands ensembles d'informations et la recherche interactive d'informations dans des affichages complexes. Nous avons choisi de centrer notre étude sur la seconde activité.

Plus précisément, notre objectif est de déterminer l'influence d'indications orales à caractère spatial, sur l'efficacité de la recherche d'informations, les performances et la satisfaction d'utilisateurs potentiels dans des activités de recherche visuelle d'informations dans des affichages complexes.

La situation d'interaction multimodale retenue est le repérage visuel d'objets graphiques, désigné dans toute la suite, comme repérage visuel ou détection de cibles. Il s'agit donc d'étudier expérimentalement la contribution potentielle de messages oraux à caractère spatial à l'efficacité de la recherche d'informations dans des affichages complexes, et de définir des recommandations ergonomiques qui permettent la conception de messages oraux susceptibles de faciliter et d'améliorer l'efficacité de la détection de cibles dans des affichages visuels complexes.

Nous avons retenu la détection de cibles comme situation d'interaction multimodale car elle intervient dans de nombreuses activités interactives, comme :

- la navigation dans les grands ensembles d'informations[16] ;
- la navigation dans les réalités virtuelles [Jacob, 1993; Tanriverdi et Jacob, 2000] ;
- la navigation sur Internet [Byrne *et al.*, 1999] ;

[16] cf. supra chapitre précédent paragraphe 2.4.3 page 15.

- l'inspection visuelle, humaine ou assistée par ordinateur, au sein des processus de fabrication de produits industriels comme la maintenance ou la sécurité [Drury, 1992].

On peut citer également la manipulation directe d'objets graphiques sur un écran puisque cette activité lie la détection de cibles à l'action (e.g., clics souris, actions de cliquer-glisser, etc.) [Rasmussen, 1986]. On parle alors de "boucle sensori motrice" pour désigner l'ensemble des phénomènes perceptifs, cognitifs et moteurs intervenant entre le moment où l'utilisateur détecte l'objet graphique sur lequel porte l'action envisagée et celui où il agit sur lui. Dans le cas de la détection de cibles, il s'agit des phénomènes qui interviennent entre le moment où le sujet détecte visuellement la cible et celui où il la sélectionne à la souris.

En outre, bien qu'il existe une quantité significative de travaux de recherche en psychologie publiés sur la détection de cibles, la plupart d'entre eux portent sur des tâches artificielles de laboratoire. On peut citer à titre d'exemple les travaux de Kramer *et al.* sur l'influence de l'apparition de distracteurs[17] lors de la détection visuelle de cibles [Kramer *et al.*, 2001]. On peut citer également ceux de Diederich *et al.* portant sur les effets de l'interaction visuelle et tactile sur les temps de réaction des sujets [Diederich *et al.*, 2003].

D'autres disciplines comme les neurosciences ou les sciences de la vision se sont également intéressées à l'activité de recherche visuelle, mais pour explorer les mécanismes neuronaux sous-jacents à cette activité chez les humains et les singes dans le but de comprendre les principes de la perception visuelle et les facteurs susceptibles d'influencer l'attention visuelle sélective. Dans [Chelazzi, 1999], Chelazzi propose une revue critique des récentes contributions scientifiques issues des neurosciences dans ce domaine.

En revanche, rares sont les études ergonomiques centrées sur l'activité choisie en situation d'interaction Homme-Machine. À noter les travaux de Doll qui discutent et comparent des modèles pour la recherche visuelle et la détection de cibles, et soulignent les effets des propriétés visuelles de la cible, de l'attention visuelle, de l'apprentissage et de la charge cognitive sur les performances humaines dans de ce type d'activités [Doll et Home, 2001].

3.2.2 Méthodologie

En l'absence de résultats scientifiques suffisants sur la recherche visuelle en situation d'interaction Homme-Machine, nous avons adopté une démarche expérimentale en deux étapes, débouchant sur un programme de recherche comportant initialement deux études :

- la première, exploratoire, fondée sur des hypothèses générales posées *a priori* et de bon sens, en raison de l'absence de résultats publiés sur le thème étudié ;
- la deuxième, fondée sur des hypothèses précises issues des résultats de la première étude, qu'elle tente de valider.

Une troisième étude expérimentale s'est avérée nécessaire pour comprendre et interpréter les résultats de la seconde étude, obtenus lors de l'analyse quantitative des performances des sujets, en termes de la rapidité et de la précision de la sélection des cibles. En particulier, cette troisième étude visait à mettre en évidence leurs stratégies de mouvement oculaire et les facteurs qui les influencent. Nous avons utilisé un dispositif de suivi du regard pour mener cette troisième étude.

[17] Objets graphiques qui ne sont pas des cibles visuelles.

3.3 Choix de la tâche expérimentale

Nous avons choisi d'étudier la contribution potentielle d'indications orales à caractère spatial au repérage visuel d'objets, ou d'éléments, graphiques dans des affichages complexes sur la base de l'hypothèse suivante :

> *Des indications orales sur la localisation spatiale de la cible dans la scène sont susceptibles de faciliter son repérage visuel. En effet, en réduisant la zone de recherche de la cible, ces indications orales, à caractère spatial, devraient permettre une exploration visuelle plus efficace de la scène, et donc devraient entraîner la réduction des temps de recherche visuelle de la cible.*

Les applications potentielles du repérage visuel de cibles sont nombreuses. On peut citer à titre d'exemple, le repérage d'icones (cf. les logiciels courants), le repérage d'éléments d'environnements 2D ou 3D (dessins, images, photographies, etc.), l'exploration de pages Web, la recherche géographique sur un plan ou une carte, la recherche d'itinéraires, etc. La tâche expérimentale retenue est la sélection à la souris de cibles visuelles dans des affichages graphiques, que nous appellerons dans la suite scènes visuelles.

3.3.1 Mise en relief visuelle des cibles

Pour améliorer et faciliter les tâches visuelles, comme le repérage ou l'exploration, différentes techniques de mise en relief visuelle ont été élaborées. La mise en relief visuelle d'un objet consiste à lui donner des propriétés visuelles qui le distinguent des autres éléments de la scène et lui confèrent une saillance visuelle. Différentes techniques existent : couleur, encadrement, clignotement, mouvement (dont le zoom), etc. On peut classer les différentes techniques de mise en relief visuelle de la cible dans la scène affichée, en fonction des critères utilisés par Bernsen dans [Bernsen, 1994] pour caractériser les différentes modalités visuelles, à savoir "statique" et "dynamique". Nous avons enrichi cette classification pour tenir compte de la spécificité de la tâche de repérage visuel. Nous avons ajouté deux critères supplémentaires, à savoir "local" et "global". La signification des quatre critères retenus ressort des deux exemples que nous donnons pour illustrer les deux principales classes de mise en relief visuelle :

- la mise en relief visuelle statique et locale consiste à rendre la cible visuellement saillante dans la scène par la couleur, l'encadrement, le contraste, entre autres ;
- la mise en relief visuelle dynamique et globale consiste à guider le regard vers la cible, par exemple, en visualisant (e.g. par une flèche tracée dynamiquement) le trajet que doit suivre le regard à partir d'un point de référence (statique) à l'écran pour atteindre la cible.

Nous n'avons retenu ni la mise en relief statique et locale, ni la mise en relief dynamique et globale, en raison de la surcharge croissante des affichages graphiques. En effet, comme nous l'avons montré au chapitre précédent, paragraphe 2.4 page 13, le flot d'informations visuelles transmises à l'utilisateur est en constante augmentation. Ceci se traduit par la multiplication des objets graphiques affichés simultanément à l'écran qui entraîne fatigue et baisse des performances des utilisateurs dans leurs tâches visuelles. Pour aller plus loin encore, il est prouvé

que le clignotement, tout comme certains motifs rayés[18] peuvent entraîner un stress visuel qui se traduit par des crises d'épilepsie [Ware, 2004]. C'est pour ces raisons que nous n'avons pas retenu pour nos expérimentations la mise en saillance visuelle.

L'assistance orale au repérage de la cible apparaît donc comme une alternative possible à la mise en relief visuelle de la cible.

3.3.2 Assistance orale ou multimodale ?

L'usage de la langue naturelle orale offre deux possibilités d'assistance au repérage de la cible en facilitant sa localisation.

Certaines expressions déictiques[19], telles "ici" ou "là" permettent au locuteur de faire référence à l'environnement réel dans lequel il se trouve au moment où il parle. Cependant, ces références au contexte d'énonciation sont souvent ambiguës même si l'on prend en compte le contexte pragmatique de l'énoncé (i.e., les échanges verbaux qui le précèdent). Dans ces cas, le locuteur peut avoir recours à une désignation gestuelle (e.g., mouvement du regard, geste de la tête, de la main) pour lever l'imprécision ou l'ambiguïté. Cette forme d'utilisation multimodale des déictiques (énoncé + geste manuel de pointage) a été proposée dès la fin des années 80 ; voir la commande "Put that there" du prototype décrit dans [Bolt, 1980].

Des tentatives, plus récentes ont été réalisées pour doter le système de moyens d'expression similaires, mais avec un succès mitigé. En effet, cette forme de multimodalité impose une "incarnation" du système sous la forme d'un agent conversationnel animé, ou ACA[20], si l'on veut reproduire avec réalisme les gestes manuels de désignation qui accompagnent les déictiques. Voir, par exemple, la P.P.P.[21], persona avec bâton de pointage dans [André, 1997]. Or, la contribution de ces représentations anthropomorphiques du système à l'utilisabilité et à l'efficacité de l'interaction Homme-Machine semble discutable [Mulken *et al.*, 1999]. C'est pourquoi nous n'avons pas retenu cette technique.

En effet, d'autres études sont nécessaires pour déterminer l'utilité de telles techniques dans le contexte de l'interaction Homme-Machine. En outre, la présence d'un ACA en surchargeant l'affichage est un facteur de fatigue visuelle supplémentaire. Enfin, le dispositif dynamique de pointage (flèche ou bâton) présente, outre les inconvénients des mises en saillance visuelles statiques, celui de ralentir l'interaction en raison même de son caractère dynamique. Le guidage du regard qu'offre cette technique et son caractère "naturel" sont des avantages que nous avons estimé insuffisants pour compenser les inconvénients.

La seconde possibilité est le recours à des expressions linguistiques spatiales autosuffisantes que l'on peut classer en deux catégories selon le repère utilisé. Les indications spatiales absolues expriment la position d'un élément de la scène par rapport à l'écran ou à l'image dans son en-

[18] Le texte sur papier quadrillé est un exemple de motif stressant visuellement car des rayures horizontales sont le support du texte. De plus, certaines polices de caractères semblent être plus mauvaises que d'autres (cf. [Ware, 2004] pages 62 et 63, fig. 2.29).

[19] Les expressions déictiques sont des expressions linguistiques qui font référence au contexte de l'énonciation. Par exemple, "je" désigne le locuteur, "tu" le destinataire de l'énoncé, etc.

[20] e.g., *embodied system agent* ou *persona*.

[21] P.P.P. signifie "Personalized Plan based Presenter".

semble ; ainsi, "en haut" signifie "en haut de l'écran/de l'image". Les indications spatiales relatives en revanche spécifient la position d'un élément par rapport à un autre élément de la scène, par exemple, "à gauche du pont". Il convient de noter que cette dernière expression fait référence implicitement au point de vue de l'utilisateur : gauche et droite sont définies par rapport à lui.

Nous avons retenu cette forme d'assistance à la localisation des cibles, car elle ne présente aucun des inconvénients des mises en saillance visuelles ou des aides à la localisation multimodales. Reste à déterminer si elle améliore effectivement de façon significative la précision et la rapidité du repérage de cibles.

3.4 Choix méthodologiques

3.4.1 Situations expérimentales étudiées

Pour évaluer les apports potentiels d'indications orales, à caractère spatial, à la localisation de cibles visuelles dans des affichages denses, nous avons choisi de comparer les performances d'utilisateurs potentiels dans des situations d'interaction Homme-Machine représentatives des deux principales activités de recherche visuelle, à savoir :

- la recherche d'un objet/élément graphique connu visuellement ;

et :

- la recherche d'un objet/élément graphique non familier visuellement, mais caractérisé par un ensemble de propriétés que l'on peut décrire verbalement.

Pour être en mesure de comparer les performances des utilisateurs, la description verbale doit être suffisamment précise pour caractériser la cible de manière unique. Les messages oraux doivent donc contenir, outre des indications spatiales sur la position de la cible dans la scène, une description de ses propriétés caractéristiques. Mais cette contrainte ne diminue en rien le réalisme des activités étudiées.

Par conséquent, compte tenu de notre objectif, trois situations sont à considérer :

- celle où la cible est familière visuellement ;
- celle où elle est décrite et localisée dans la scène ;
- et celle où elle est familière visuellement et où l'utilisateur connaît ses caractéristiques et sa position dans la scène.

La réalisation de la première situation impose de présenter à l'utilisateur la cible isolée avant la scène dans laquelle il doit la chercher ; et ce, pour qu'il puisse se familiariser visuellement avec elle. Pour rendre possibles des comparaisons entre les trois situations, nous avons choisi de présenter la cible avant la scène. Les trois conditions expérimentales considérées comprennent deux étapes, une présentation de la cible suivie d'un affichage de la scène où elle doit être recherchée. Ces trois conditions se distinguent uniquement par le mode de présentation de la cible :

- visuel : il s'agit de l'affichage de la cible isolée ;
- oral : il s'agit de l'énoncé verbal de la caractérisation de la cible et de l'indication de sa position dans la scène ;

- multimodal : il s'agit de la présentation simultanée de la cible dans les deux modes précédents.

Pour un couple cible + scène donné, la tâche consiste donc à sélectionner, à la souris, une cible dans une scène, après que la cible ait été présentée, et ce, soit visuellement, soit oralement, soit visuellement et oralement (i.e., présentation multimodale de la cible). Dans le cadre de notre étude expérimentale, la présentation visuelle des cibles correspond à la situation de référence, i.e., celle où la présentation de la cible ne comporte pas de message oral.

3.4.2 Caractéristiques de affichages

Nous avons choisi de restreindre l'étude à la détection de cibles dans des affichages classiques, c'est-à-dire statiques et en deux dimensions.

Statique *versus* dynamique

La détection de cibles dans des présentations animées est une activité visuelle différente et bien plus complexe que dans les présentations statiques. En conséquence, nous nous sommes bornés à étudier les apports de la parole pour la recherche visuelle dans des affichages statiques exclusivement.

2-dimensions *versus* 3-dimensions

Les techniques d'affichage en trois dimensions ont connu un développement spectaculaire. Il est désormais possible d'interagir avec le système au sein d'espaces virtuels 3D. Il convient de noter cependant que les interfaces graphiques utilisateur 3D, en l'état actuel, engendrent pour l'utilisateur de nouvelles difficultés. Par exemple, la perception de l'espace n'est guère facilitée par la profusion des couleurs. De même, des améliorations sont encore nécessaires pour l'exécution de tâches comme l'évaluation des positions ou des mouvements relatifs et la planification de chemins d'accès à l'information (cf. [Ware, 2004]).

Par ailleurs, bien que les utilisateurs aient une préférence pour certains affichages 3D, leurs performances en termes de rapidité et de précision sont réduites par rapport aux affichages 2D. Par exemple, les sujets de l'étude comparative expérimentale présentée dans [Levy *et al.*, 1996] préfèrent les organisations en 3 dimensions dans les situations qui font appel à leur mémoire. Cependant, l'ajout de la profondeur altère leur précision dans les contextes d'évaluation des distances ou de perception dans l'espace. À noter également l'étude comparative 2D *versus* 3D dans [Sutcliffe et Patel, 1996] qui met en évidence que l'organisation 3D d'une hiérarchie de documents ne présente pas d'avantage sur son organisation 2D, en terme des temps de recherche de pages contenues dans les documents.

C'est pour cela que nous avons choisi de limiter notre recherche aux représentations statiques dont la structure est à deux dimensions.

3.4.3 Les apports perceptifs de l'assistance orale à la détection de cibles

Globalement, notre recherche a pour objectif de déterminer l'apport éventuel de la parole à l'interaction Homme-Machine pour des applications graphiques mettant en œuvre la perception visuelle. Nous avons donc choisi d'explorer, d'abord et en priorité, les apports perceptifs[22] de la parole en tant que modalité d'expression complémentaire du système. Ils seront évalués de la façon suivante :

- par la mesure des temps de sélection des cibles ;
- par la mesure de la précision de sélection des cibles ; i.e., dans la cible ou en dehors.

Les apports cognitifs, par exemple la compréhension des messages oraux leur mémorisation et celle des cibles ne seront pas abordés dans ce travail.

3.5 Description du programme expérimental

Le programme expérimental réalisé comporte trois études. La première étude porte sur le rôle de la parole sous forme d'indications spatiales dans le repérage visuel, en particulier dans le cas où l'affichage est dense et complexe [23].

Compte tenu des résultats de l'analyse qualitative des données recueillies lors de cette étude préliminaire, nous avons choisi de centrer la suite de nos travaux sur l'analyse de l'influence de la structure spatiale des affichages sur l'efficacité de la recherche d'informations visuelles, en particulier celle de leur repérage avec ou sans l'assistance d'indications orales, à caractère spatial. La deuxième étude de notre programme expérimental porte donc sur l'influence des structures spatiales d'affichages complexes sur les performances des sujets pour la même tâche de repérage que la première, avec ou sans l'assistance d'indications orales, à caractère spatial. Les mesures sont limitées à la précision et aux temps de réponse des sujets (cf. infra 3.5.2, page 25).

La troisième porte sur l'influence des structures spatiales d'affichage, étudiés dans le cadre de la seconde, sur les stratégies d'exploration visuelle mises en place par les sujets pour la tâche de détection de cibles. Dans cette étude, en plus de la précision et des temps de réponse des sujets, nous avons recueilli leurs fixations oculaires afin d'analyser leurs stratégies d'exploration visuelle (cf. infra 3.5.3, page 26).

3.5.1 Assistance orale à la détection de cibles : une étude préliminaire

La première étude dans le cadre de cette recherche sur la multimodalité parole + graphique est une étude à caractère préliminaire et exploratoire qui vise à déterminer l'apport éventuel de la parole à l'interaction homme machine. Plus particulièrement, il s'agit d'évaluer l'apport d'indications orales à caractère spatial dans un contexte d'interaction Homme-Machine multimodale. Cette étude préliminaire est destinée à fournir des éléments de réponse à la question suivante : des messages sonores contenant des indications spatiales sont-ils susceptibles d'améliorer l'efficacité de l'interaction entre le système et l'utilisateur en facilitant la tâche de repérage visuel ?

[22] Au sens de perception visuelle.

[23] cf. infra 3.5.1, page 24.

En effet, la désignation verbale de la cible visuelle, associée à une indication spatiale de sa position dans la scène, devrait en faciliter le repérage en termes des temps et de la précision de sélection des cibles.

3.5.2 Assistance orale à la détection de cibles au sein de structures visuelles symétriques

La deuxième étude est centrée sur l'étude de l'influence de la structure spatiale des scènes graphiques sur l'efficacité de leur exploration, avec ou sans l'assistance d'indications orales à caractère spatial. Cette étude est destinée à fournir des éléments de réponse aux questions suivantes :

- en l'absence et en présence d'informations de localisation des cibles, quelle est l'influence de l'organisation spatiale des scènes affichées sur l'efficacité de l'interaction dans ces deux situations ?
- existe-t'il une structure visuelle pour laquelle l'apport des messages sonores est plus important ?

Ce problème présente un double intérêt :

- Sur le plan scientifique d'une part, dans la mesure où il n'a pas encore été étudié de façon systématique, surtout dans des situations d'interaction Homme-Machine réalistes ;
- Sur le plan des applications potentielles d'autre part, en raison du développement des techniques de visualisation de grands ensembles d'informations qui mettent en œuvre des organisations spatiales variées, radiales (par exemple les arbres hyperboliques [Lamping *et al.*, 1995]) matricielles ou arborescentes (comme les cartes arborescentes [Plaisant *et al.*, 2002]).

Pour déterminer quelle est l'influence des structures visuelles sur les performances et la satisfaction des utilisateurs, nous avons élaboré une expérimentation où quatre structures visuelles sont testées :

- l'absence de structure, où les éléments constituants une scène sont disposés aléatoirement ; il s'agit de l'organisation spatiale de référence ;
- la structure matricielle, où les éléments sont présentés sous la forme d'un tableau à deux dimensions ;
- la structure radiale, où les éléments sont répartis suivant huit rayons partant du centre de l'écran et se dirigeant respectivement, en diagonale vers les quatre coins et, horizontalement et verticalement, vers les milieux des quatre côtés de l'écran ;
- la structure elliptique, où les éléments sont présentés sous la forme de deux ellipses concentriques.

À noter que les structures matricielle, radiale et elliptique sont toutes trois symétriques.

Comme dans l'expérience préliminaire, les performances des sujets sont évaluées en termes des temps et de la précision de sélection des cibles.

3.5.3 L'organisation des affichages : une forme de guidage visuel ?

La troisième étude est centrée sur l'étude des stratégies d'exploration visuelle adoptées par les utilisateurs au sein des mêmes organisations spatiales que pour la deuxième étude, en l'absence cette fois-ci, d'indications orales à caractère spatial. Comme dans les expérimentations précédentes, les performances des sujets sont évaluées en termes des temps et de la précision de sélection des cibles. De plus, un eye-tracker permet d'enregistrer les fixations oculaires lors de la recherche des cibles.

Cette étude est destinée à fournir des éléments de réponse aux questions suivantes. Quelle est l'influence de l'organisation spatiale des scènes graphiques sur les stratégies de parcours visuel de la scène mises en place par les utilisateurs pour détecter une cible dans une image complexe ? Autrement dit, les utilisateurs adaptent-ils leur stratégie d'exploration en fonction de la structure visuelle de la scène ? Sinon, leur stratégie de parcours visuel est-elle stable et indépendante de l'organisation spatiale de la scène ? Si c'est le cas, et si cette stratégie ne varie pas d'un utilisateur à l'autre, quelle est la structure de l'affichage la plus compatible avec cette stratégie ?

Chapitre 4

Étude préliminaire

Cette étude expérimentale à caractère exploratoire a pour objectif de comparer l'efficacité, en termes de précision et de rapidité, de trois modes de présentation de la cible : visuel, oral et multimodal. Les cibles à sélectionner sont des éléments appartenant à des scènes 2D réalistes ou abstraites en couleur. Après la présentation visuelle, orale ou multimodale de la cible, les sujets (18) ont pour consigne de "trouver" la cible dans la scène qui s'affiche à l'écran et de la sélectionner le plus vite possible à la souris.

La principale conclusion émergeant de cette étude est que les indications verbales de localisation spatiale de la cible associées à la présentation visuelle de celle-ci dans la scène - donc au sein des présentations multimodales - facilitent la recherche visuelle : elles améliorent, à la fois, temps et précision des sélections de la cible par rapport aux deux autres formes de présentations testées, i.e. visuelle et orale.

Après avoir décrit la méthodologie utilisée dans la conception du plan expérimental, nous décrivons le protocole expérimental ainsi que l'élaboration du matériel visuel et sonore utilisé au cours de l'expérimentation. Ensuite, nous présentons la méthodologie d'analyse des données recueillies, selon deux axes : analyse quantitative et analyse qualitative. Enfin, nous commentons les résultats.

4.1 Méthodologie

Afin d'estimer la contribution potentielle d'informations spatiales exprimées oralement pour faciliter le repérage visuel, nous avons conçu une étude expérimentale portant sur la sélection à la souris d'un objet graphique. Les sujets devaient localiser puis sélectionner une cible, le plus rapidement possible, à la souris, dans k scènes visuelles[24]

Chaque cible leur avait été préalablement présentée pendant n millisecondes selon les trois modes de présentation suivants[25] :

[24] L'entier k représente la taille minimale de l'ensemble des images, et doit être choisi de façon à permettre l'exploitation des données recueillies lors de l'expérience.

[25] L'entier n représente la durée en millisecondes de présentation des cibles, et doit être choisi en fonction des contraintes imposées par le protocole.

- présentation visuelle de la cible isolée au centre de l'écran (PV) ;
- désignation orale de la cible accompagnée d'indications spatiales (messages oraux d'une durée aussi proche que possible de n millisecondes) (PO) ;
- présentation multimodale de la cible (PM) : il s'agit de la présentation simultanée orale et visuelle de la cible[26].

Chacune des trois présentations PV, PO et PM était suivie de l'affichage de la scène, ou image, dans laquelle se situait la cible à repérer.

La tâche de repérage était effectuée pour chacun de ces trois types de présentation, définissant ainsi trois situations expérimentales distinctes, à savoir les situations PV, PO et PM. En associant présentation de la cible et affichage de la scène visuelle correspondante, on définit ainsi un couple de stimulus (présentation de la cible + affichage de la scène associée). Nous disposons donc de trois ensembles PV, PO et PM de k couples de stimuli (soit un couple par image).

Afin d'évaluer l'apport de la parole pour le repérage visuel, nous avons mené une étude comparative sur les performances des sujets dans les trois situations, en considérant la situation PV comme situation de référence. Le protocole expérimental[27] a été conçu en fonction d'hypothèses plausibles sur la contribution potentielle d'indications orales pour le repérage de cibles visuelles. Après la présentation de ces hypothèses, ce paragraphe décrit la caractérisation des matériels visuels et sonores présentés aux sujets.

4.1.1 Hypothèses à tester

L'expérience réalisée est destinée à tester la validité des hypothèses de bon sens suivantes, fondées sur la quantité d'informations fournies par les différentes présentations :

- hypothèse A : La présentation multimodale devrait réduire les temps de sélection des cibles et améliorer la précision des sélections, par rapport à la présentation visuelle de la cible isolée ;
- hypothèse B : La présentation orale devrait réduire les temps de sélection des cibles et améliorer la précision des sélections, par rapport à la présentation visuelle de la cible isolée ;
- hypothèse C : Le type d'information spatiale contenue dans les présentations orales des cibles devrait influencer la précision et les temps de sélection des sujets. En particulier, les informations spatiales absolues ou relatives devraient s'avérer plus efficaces que les informations faisant référence aux connaissances *a priori* des sujets.

4.1.2 Caractérisation des images présentées aux sujets

Dans tout ce paragraphe, nous ferons référence aux travaux de Bernsen sur la caractérisation des modalités en sortie présentée dans [Bernsen, 1994]. Cette taxonomie est intéressante en raison

[26] Les présentations visuelles et orales des cibles étaient les mêmes dans la situation PM que dans les situations PV et PO respectivement.

[27] cf. infra paragraphe 4.2 page 33.

de son exhaustivité[28] et comprend, entre autres, 14 modalités graphiques. Le tableau 4.1 page 29 donne, pour chacune des 14 modalités graphiques de Bernsen, sa description ainsi que son numéro de modalité dans [Bernsen, 1994], puis un ou plusieurs exemples l'illustrant.

Description des modalités graphiques de Bernsen [Bernsen, 1994]	
Modalités	Exemples
Langage sémiotique statique (1)	Hiéroglyphes.
Langage sémiotique dynamique (2)	Langage gestuel, hiéroglyphes dynamiques.
Langage non-sémiotique statique (5)	Lettres, mots, langages de programmation.
Langage non-sémiotique dynamique (6)	Texte défilant, sous-titres.
Schémas ou graphiques statiques (9 et 11)	Diagrammes, cartes, formes géométriques.
Images réalistes statiques (10)	Photographies, dessins.
Schémas ou graphiques animés (12 et 14)	Diagrammes animés, animations.
Images réalistes animées (13)	Films, vidéos, animations réalistes.
Graphiques arbitraires statiques (21)	Formes arbitraires, diagrammes géométriques.
Graphiques arbitraires animés (22)	Diagrammes géométriques animés.
Structures graphiques statiques (25)	Grilles, tables, arbres, fenêtres.
Structures graphiques animées (26)	Grilles, tables, arbres et fenêtres animés.

TAB. 4.1 – Description et exemples des modalités graphiques de Bernsen
À noter que pour chaque modalité, on présente sa description et son numéro dans le tableau 3 page 355 dans [Bernsen, 1994] (1ère colonne), puis des exemples l'illustrant (2ème colonne).

Nous avons analysé puis adapté cette taxonomie des modalités exclusivement graphiques de Bernsen pour tenir compte des médias et des contextes d'interaction Homme-Machine actuels (application, environnement d'utilisation, etc.). En effet, diverses considérations nous ont amenés à réduire, par fusion et suppression, le nombre de ces modalités. La classification résultante ne porte que sur les images présentées aux sujets lors de l'expérience (cf. infra tableau 4.2 page 30).

Vers une classification simplifiée des modalités graphiques de Bernsen

Comme nous l'avons vu précédemment[29], nous nous bornons à étudier les apports de la parole pour les tâches visuelles dans des affichages statiques exclusivement.

Nous supprimons également toutes les modalités linguistiques ou modalités associées par Bernsen au langage écrit. En effet, le repérage visuel de caractères, de mots ou de phrases met en jeu des processus cognitifs de nature et de complexité différentes de ceux qui interviennent dans le repérage d'objets graphiques ; c'est du moins la position que nous adoptons en l'absence de données expérimentales spécifiques sur ce point.

Nous fusionnons les images schématiques avec les graphiques que nous regroupons en une seule modalité "schémas ou graphiques". Nous justifions cette fusion de la manière suivante. Les

[28] cf. supra paragraphe 2.1.4 page 7.
[29] cf. chapitre 3 paragraphe 3.4.2 page 23.

images schématiques et les graphiques ont les mêmes propriétés binaires considérées par Bernsen pour classer les modalités, à savoir : non-linguistique, analogique, non-arbitraire. De plus, la distinction entre ces modalités ne nous semble pas de nature à influer sur la sélection des cibles, c'est-à-dire sur les stratégies et performances des sujets. Elle isole simplement les graphes (classe 11) des autres graphiques et schémas (classe 9).

Nous avons procédé à d'autres adaptations afin de ne conserver pour notre expérimentation que les types d'affichage utilisés le plus souvent, sur Internet par exemple.

Notre classification simplifiée contient :

- les schémas ou graphiques statiques dont les autres traits binaires sont non-linguistique, analogique, non-arbitraire. Ils correspondent à la fusion des modalités 9 et 11 de la taxonomie des modalités de Bernsen ;
- les images réalistes statiques dont les autres traits binaires sont non-linguistique, analogique, non-arbitraire. Elles correspondent à la modalité 10 de la taxonomie des modalités de Bernsen ;
- les graphiques arbitraires statiques dont les traits binaires sont non-linguistique, non-analogique et arbitraire. Ils correspondent à la modalité 21 de la taxonomie des modalités de Bernsen ;
- les structures graphiques statiques dont les traits binaires sont non-linguistique, non-analogique et non-arbitraire. Elles correspondent à la modalité 25 de la taxonomie des modalités de Bernsen.

Taxonomie du matériel visuel de l'expérimentation	
Modalités	Exemples
Schémas ou graphiques statiques (9 et 11)	Diagrammes, cartes, formes géométriques.
Images réalistes statiques (10)	Photographies, dessins.
Graphiques arbitraires statiques (21)	Formes arbitraires, diagrammes géométriques.
Structures graphiques statiques (25)	Grilles, tables, arbres, fenêtres.

TAB. 4.2 – Classification simplifiée des modalités graphiques de Bernsen utilisées dans l'étude. Chaque modalité est suivie du nombre correspondant à sa classe dans le tableau 3 page 355 de [Bernsen, 1994].

Classification du matériel graphique expérimental

Nous avons réduit encore la classification simplifiée des modalités du tableau 4.2 en éliminant les structures graphiques statiques. Nous les avons supprimées car elles peuvent inclure du texte. En effet, les mots, les chiffres sont susceptibles de modifier les stratégies de repérage visuel d'objets graphiques simples. En revanche, les structures visuelles feront l'objet du chapitre 5. Des études, comme celles de Cribbin et Chen [Cribbin et Chen, 2001a; Cribbin et Chen, 2001b], mettent en évidence des différences de traitement visuel entre les images structurées et non structurées.

En outre, nous fusionnons les schémas et les graphiques statiques avec les graphiques arbitraires statiques, distinguant ainsi les scènes abstraites (diagrammes, cartes, collections de formes géométriques) des scènes réalistes (photographies, dessins).

Au terme de notre analyse, nous obtenons donc deux classes de scènes graphiques statiques :
- la classe 1 d'images abstraites, qui contient les collections, structurées ou non, d'objets symboliques ou arbitraires, comme les cartes (schémas), les panneaux de circulation ou encore les formes géométriques (objets arbitraires) ;
- la classe 2 d'images réalistes, qui contient les représentations de la réalité, sous forme de photographies ou de dessins ; elle distingue les objets complexes isolés, des paysages ou scènes d'intérieur et des groupes de personnages.

C'est cette classification que nous avons utilisée.

4.1.3 Caractérisation des messages sonores

Les messages sonores peuvent être caractérisés d'après les traits binaires utilisés par Bernsen pour caractériser l'une des modalités auditives. Il s'agit du langage parlé décrit comme linguistique, non analogique, non arbitraire et dynamique. Cependant, si la taxonomie de Bernsen est intéressante pour décrire le matériel visuel, sa précision est insuffisante dans le contexte de notre expérience, pour décrire le matériel sonore. En effet, elle ignore la structure linguistique des énoncés et le type d'information apportée par le message sonore.

Les messages sonores utilisés pour notre expérience doivent contenir la désignation de la cible visuelle et des indications spatiales destinées à en faciliter la localisation dans la scène. Par analogie avec les légendes de journaux et, comme le montre [Burhans *et al.*, 1995], les désignations verbales doivent énoncer des caractéristiques discriminant la cible des autres éléments non cibles de la scène. Par exemple, la désignation verbale d'une cible peut contenir l'énoncé de ses propriétés graphiques (comme sa couleur ou sa forme), ou son nom.

Contenu des messages sonores

D'une part, une scène est composée de collections, structurées ou non, d'objets graphiques. Chaque objet peut donc être désigné oralement par son nom dans le message sonore.

D'autre part, les travaux de [Burhans *et al.*, 1995] caractérisent les informations spatiales en distinguant :
- les prépositions spatiales ou topologiques (sous, sur, dans, etc.) ;
- les prépositions projectives (gauche, bas, centre, etc.) ;
- les relations spatiales implicites (de gauche à droite, de bas en haut) utilisées dans des messages du type "Dans la dernière ligne, ..." qui désigne la ligne en bas d'une scène ;
- les verbes ou adjectifs spatiaux (tenir un objet, être posé sur une table) ;
- ou encore l'absence d'information spatiale, si l'objet à localiser est saillant ou si on ne spécifie aucun objet précis dans la scène.

En outre, d'après [Frank, 1998], le langage parlé n'offre que quelques formes d'indications spatiales, que l'on peut caractériser comme suit. L'indication de la position de la cible, déterminée à partir d'une simple observation de la scène, est donnée :

- soit sous forme d'indication spatiale absolue (ISA), comme par exemple, "à gauche/droite", "en haut/bas" de l'écran ;
- soit sous forme d'indication spatiale relative (ISR), c'est-à-dire en donnant la position de la cible par rapport à un autre élément de la scène, comme par exemple, "à gauche/droite de...", "au-dessus de...".

Nous avons approfondi l'analyse de ces formes d'indications verbales, en ajoutant les indications spatiales implicites (ISI), pour désigner une information spatiale qui peut être facilement inférée grâce aux connaissances *a priori* de l'utilisateur et au contexte visuel de la scène. Par exemple, si la scène est composée d'une carte d'Europe sur laquelle on a situé géographiquement des monuments célèbres et si la cible associée est un icone de la Tour Eiffel, alors le message sonore "La Tour Eiffel" peut être considéré comme de type ISI. Il désigne la cible par le mot "tour" et en donne implicitement la position sur la carte, par le nom propre "La Tour Eiffel", dont on sait qu'elle se trouve en France.

Nous obtenons ainsi des messages sonores contenant la désignation de la cible ainsi que sa position dans la scène sous forme d'indication spatiale absolue (ISA), relative (ISR) ou implicite (ISI). Nous montrons dans le paragraphe suivant comment les formes d'indications spatiales ont été combinées de façon à faire varier davantage les types de messages sonores.

Combinaisons d'indications orales

Un message sonore peut ne contenir aucune indication spatiale. C'est le cas si, et seulement si, l'image comporte des indices visuels suffisants pour localiser correctement la cible, grâce uniquement à sa désignation. Ce type de message sera codifié par le mot ABSENCE.

Chaque message peut contenir deux indications spatiales de même type (ISA, ISR ou ISI). Par exemple, dans le groupe nominal "En bas, à droite", il y a deux indications spatiales absolues juxtaposées, à savoir "En bas" et "À droite". Cette indication spatiale est du type ISA+ISA. Nous l'avons simplifiée en ISA. Nous avons simplifié de la même façon les messages sonores du type ISR+ISR et ISI+ISI en messages sonores de type ISR et ISI, respectivement.

Chaque message sonore peut contenir deux indications spatiales de type différent. Par exemple, dans le groupe nominal "En bas, à droite de...", il y a deux indications spatiales différentes juxtaposées, à savoir "En bas" qui est de type ISA et "À droite de..." qui est de type ISR. Cette indication spatiale est du type ISA+ISR. De la même façon, on peut obtenir des messages de type :

- ISA+ISI, en combinant indication spatiale absolue et indication spatiale implicite ; par exemple, "À l'ouest de Paris, le bus" ;
- ISR+ISI, en combinant indication spatiale relative et indication spatiale implicite ; par exemple, "À l'ouest de l'hôtel de ville, le bus".

Structure syntaxique des messages sonores

Afin que l'information contenue dans les messages soit la seule variable, tous les messages avaient la même structure syntaxique. Cette structure a été choisie de façon à mettre en relief l'information spatiale en la situant en début d'énoncé :

[indication spatiale] + nom de la cible (désignation)

Nous avons choisi cet ordre de présentation car il fournit les informations dans l'ordre où elles seront utilisées : le regard se fixe d'abord sur la zone indiquée puis identifie la cible grâce à sa caractérisation verbale.

Cette structure syntaxique a été adoptée pour tous les messages, sauf ceux de type ISI. Comme dans l'exemple du paragraphe précédent, les messages de type ISI contiennent un groupe nominal qui, à la fois, désigne et situe la cible dans la scène. Il n'est donc pas possible d'appliquer la structure type décrite ci-dessus.

4.2 Protocole expérimental

4.2.1 Généralités

Pour tester les hypothèses de travail A, B et C[30], nous avons proposé aux sujets de sélectionner des cibles visuelles dans les trois situations expérimentales PV, PO et PM. Plus précisément, pour tester les hypothèses A et B, nous avons comparé les performances des sujets dans les trois conditions de présentation des cibles. Pour tester l'hypothèse C, nous avons comparé leurs performances dans les trois conditions, pour les trois groupes d'images suivants : celui dont les images sont associées à des messages sonores de type ISA, celui dont les images sont associées à des messages sonores de type ISR et celui dont les images sont associées à des messages sonores de type ISI.

Les images auxquelles nous avons associé un message sonore de type ISI comportent des indices visuels susceptibles de faciliter la localisation de la cible dans les trois conditions. Par exemple, une des images est constituée par des photographies d'animaux présentées sur un planisphère. Le message oral correspondant se réduit alors à l'énoncé "Le roi des animaux". Il contient implicitement, à la fois, la désignation de la cible (à savoir, un lion) et l'indication spatiale de sa localisation dans la scène (à savoir, l'Afrique).

Les mêmes messages oraux sont utilisés à la fois dans les situations PO et PM. Ceci nous permet de recueillir des données comparables entre les deux situations PO et PM. La durée de la présentation de la cible, notée n au paragraphe 4.1 à la page 27, a été fixée empiriquement à 3 secondes[31] pour concilier les exigences contradictoires suivantes :

- d'une part, une durée importante présentait le risque de déconcentrer les sujets ;

[30] cf. paragraphe 4.1.1 page 28.

[31] Soit n=3000 ms, où la précision est en millièmes de secondes.

- d'autre part, la durée de la présentation devait être suffisante pour permettre aux sujets d'appréhender la totalité des propriétés visuelles caractéristiques de la cible ;
- enfin, les durées de la présentation visuelle de la cible et du message oral associé devaient être voisines[32].

En ce qui concerne le fond d'écran sur lequel les cibles isolées apparaissent, il varie d'une cible à l'autre. Il est déterminé de façon à faciliter la perception de la cible (contraste, couleur, etc.). Ce fond peut subsister pendant la présentation de l'image. En effet, les images choisies sur le Web peuvent avoir une résolution ne permettant pas toutes les modifications. Par exemple, il peut s'avérer impossible de conserver une précision suffisante lors de leur agrandissement pour qu'elles couvrent la totalité de la surface de l'écran. Dans ce cas, nous avons préféré limiter la taille de l'image pour conserver une précision suffisante.

4.2.2 Scénario d'interaction Homme-Machine

La tâche expérimentale proposée aux sujets était de sélectionner à la souris, le plus rapidement possible, des objets graphiques ou cibles, dans des affichages denses. Le mode opératoire est le suivant. Chaque cible est d'abord présentée, soit visuellement, soit oralement, soit de manière multimodale (i.e., visuellement et oralement). Lorsque la présentation de la cible s'achève, le message "Pour commencer, cliquer sur OK" apparaît. Le bouton OK est toujours placé au centre de l'écran. Enfin, l'image où figure la cible, s'affiche. Le sujet la repère puis la désigne à la souris, aussi rapidement que possible. La figure 4.1, page 35 donne un exemple de chaque phase du traitement dans la situation de présentation de la cible PV, avec une image de l'expérience. Il s'agit de l'image 14 de la classe 1 qui contient les images abstraites.

Le repositionnement de la souris au centre de l'écran permet de fixer la position initiale de la souris avant l'affichage des scènes. De cette façon, pour une scène donnée, nous pouvons comparer les réactions motrices des sujets (cf. la boucle sensori motrice dans [Rasmussen, 1986]) en mesurant le temps nécessaire au repérage d'une cible dans une scène. Dans le contexte du protocole, le temps de repérage visuel d'une cible contient le temps de recherche de la cible d'une part, et le temps de sélection à la souris de la cible d'autre part.

D'après [Drury, 1992], pour une zone donnée de l'image, les temps de recherche d'une cible dépendent de la taille du lobe visuel (cf. les travaux de [Gramopadhye et Madhani, 2001]), des temps de fixation oculaire et des stratégies de recherche visuelle.

Le scénario envisagé, la sélection d'un objet à la souris, est réaliste car c'est l'une des tâches élémentaires réalisées le plus fréquemment en situation d'interaction avec les GUIs (cf. la manipulation directe).

4.2.3 Les variables de l'expérience

Les variables libres pour notre problème sont :

- le mode de présentation de la cible : visuel (PV), oral (PO), ou multimodal (PM) ;

[32] Dans la situation PM, nous avons choisi de faire coïncider la disparition de la cible avec la fin du message oral.

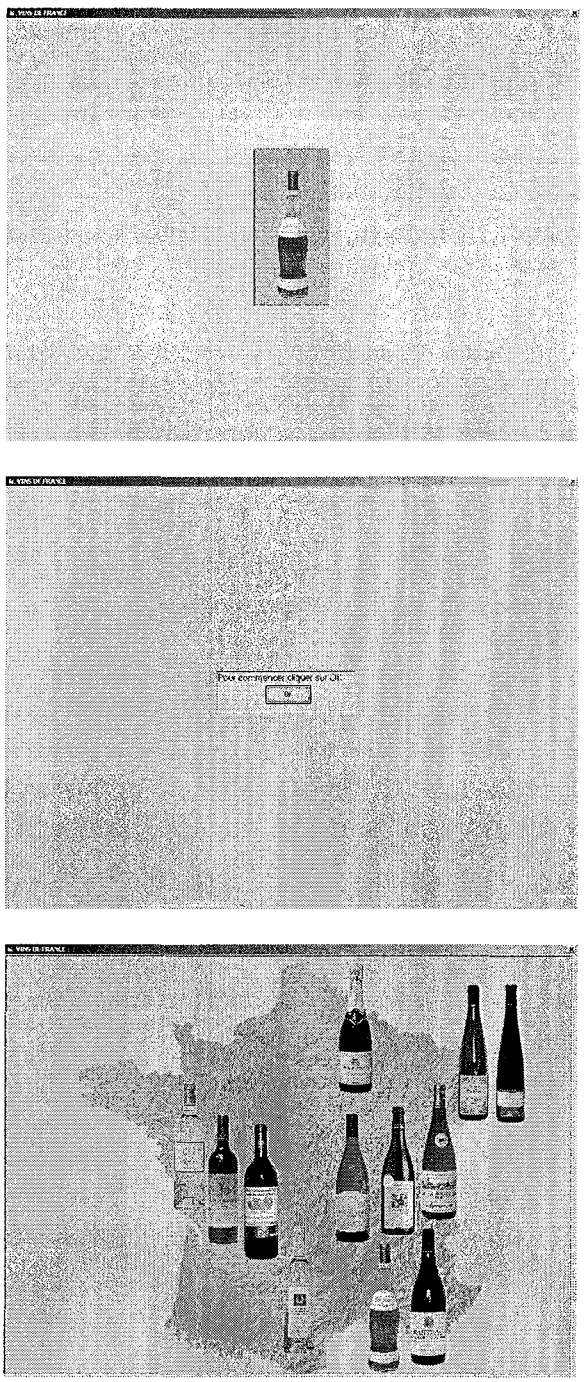


FIG. 4.1 – Déroulement d'une tâche de repérage.
La première image est la présentation de la cible, la deuxième est le repositionnement de la souris au centre de l'écran grâce à la sélection obligatoire du bouton OK, la troisième est l'affichage de la scène.

- le type de scène affichée à l'écran (abstraite *versus* réaliste, photographie *versus* dessin, etc.) ;
- le type de message sonore utilisé pour les présentations orales et multimodales (indication spatiale absolue *versus* relative, etc.) ;
- la position de la cible dans chaque scène (en haut/bas, à gauche/droite, etc.).

Les variables liées pour notre problème sont :

- le temps de sélection de la cible en millisecondes ;
- la précision de la sélection de la cible, i.e. dans la cible ou en dehors.

Pour assurer la pertinence, par rapport aux objectifs visés, du scénario, des images et des cibles proposées aux sujets nous avons d'abord défini un ensemble de critères de choix du matériel visuel. Nous avons ensuite appliqué ces critères pour sélectionner images et cibles. Nous avons enfin validé empiriquement notre sélection en soumettant les images et les cibles au jugement de diverses personnes qui ne faisaient pas partie du groupe de sujets qui ont participé à l'expérience. Ces critères sont présentés dans le paragraphe suivant.

4.2.4 Le matériel visuel

Critères de sélection des scènes visuelles

Pour mémoire, le choix des scènes complexes présentées aux sujets s'appuie sur la version simplifiée de la classification des modalités mettant en jeu la perception visuelle proposée par Bernsen dans [Bernsen, 1994]. Dans cette classification, nous avons distingué trois type d'images à proposer aux sujets : les schémas ou graphiques, les images réalistes et les graphiques arbitraires[33].

Nous avons finalement regroupé graphiques et formes arbitraires. Nous distinguons ainsi deux classes principales d'images :

- la classe 1, qui regroupe les lignes 1 et 3 du tableau. Elle contient les images abstraites ;
- la classe 2, qui correspond à la ligne 2 du tableau. Elle contient les images réalistes ou représentations simplifiées de la réalité.

La classe 1 contient les collections structurées ou non structurées d'objets graphiques symboliques ou arbitraires et les représentations/objets graphiques symboliques dont la sémiotique est une construction culturelle. On considère par exemple, des collections de drapeaux ou de panneaux de signalisation disposés en matrice. La cible associée à ce type d'images est un objet de la collection. La classe 1 contient également les formes arbitraires sans valeur sémiotique, comme les formes géométriques. La cible associée à ce type d'images est une forme géométrique. Enfin, la classe 1 contient les structures graphiques symboliques, comme les cartes. La cible associée à ce type d'images est un icone de type photographie ou dessin.

La classe 2 contient les représentations réalistes ou simplifiées de la réalité, sous forme de photographies d'une part, et de dessins ou caricatures d'autre part. Elle distingue les objets complexes isolés, les paysages ou scènes d'intérieur et les groupes de personnages. Les objets complexes isolés imposent de choisir comme cible un élément de l'objet et donc offrent la possibilité de tester l'intérêt, dans ce contexte, des indications orales exprimant une relation de type

[33] cf. paragraphe 4.1.2 page 28.

partie-tout. Les paysages et les scènes d'intérieur permettent aux sujets d'inférer, à partir de leur expérience antérieure, la position de la cible dans la scène. Enfin, les photographies ou dessins de groupes permettent d'imposer aux sujets la mise en œuvre d'inférences complexes mettant en jeu à la fois des indices visuels, des relations spatiales, leur expérience antérieure et des relations de type partie-tout.

Répartition du matériel visuel entre les classes d'images

Pour améliorer la lisibilité de note étude, nous numérotons chaque image des classes 1 et 2. Puis, pour faire référence à une image, un groupe d'images d'une classe, ou la classe entière, nous utilisons la codification suivante : numéro de l'image ou intitulé du groupe d'images, puis classe de l'image, entre parenthèses.

Par exemple, pour nommer l'image 12 de la classe 1, nous écrivons : 12(1). De même, pour coder la sous classe d'objets symboliques de la classe 1, nous écrivons : objets symboliques (1). Enfin, pour coder la classe 1 d'images, nous écrivons : images (1). Cette codification est utilisée dans toute la suite du travail.

Le matériel visuel expérimental comprend k = 36 images en tout, où k est le nombre de scènes visuelles introduit au paragraphe 4.1 à la page 27. Elles sont réparties équitablement entre les images de type (1) et (2), soit 18 images pour la classe 1 et 18 images pour la classe 2. Le nombre limité de scènes à traiter s'explique de la façon suivante. Étant donné le caractère exploratoire et préliminaire de cette étude, nous nous sommes bornés à faire en sorte que chaque sous-classe ou groupe d'images comporte un nombre d'images au moins égal à trois. Ainsi, les sous-classes objets symboliques (1), formes géométriques (1), objets réels à connotation symboliques (1) et objets complexes (2) contiennent chacune six images en tout. La sous-classe paysages (2) contient douze scènes et la sous-classe groupes de personnages (2) contient trois scènes. Il aurait fallu augmenter considérablement le nombre total d'images pour pouvoir obtenir un nombre égal d'images par sous-classe d'images retenue.

Unicité : critère de sélection des cibles

Le critère indispensable pour le choix d'une cible dans une scène est la contrainte d'unicité. Nous avons considéré comme valides en tant que cibles, les objets graphiques, ou constituants d'objets, qui vérifiaient les conditions suivantes :

- elles devaient être facilement identifiables et uniques, en termes des objets, ou constituants, qu'elles représentaient, ceci pour la situation PV ;
- elles devaient pouvoir être désignées oralement, de façon simple et non ambiguë, en termes de désignation de l'objet et de désignation spatiale, ceci pour la situation PO.

Bien que chaque cible soit unique par ses propriétés visuelles, il existe des confusions possibles avec d'autres constituants de la scène dans la situation PV. Notamment, des objets graphiques de même taille, ou de même couleur, ou de même forme que la cible peuvent apparaître dans les collections d'objets graphiques (1) ou les paysages (2). C'est le cas pour les images 1, 2, 7 et 9 (1) et l'image 7 (2).

En outre, il peut exister des ressemblances entre la cible et d'autres éléments non-cibles de la scène lorsqu'elle est du type représentation de la réalité (2) ; notamment lorsque les non-cibles appartiennent à la même classe d'objets ou constituants d'objets que la cible. Ainsi, les cibles des images 3, 4, 5, 11, 12 et 18 (2) sont peut-être moins faciles à identifier que les autres. L'uniformité du contenu de l'image est peut-être moins favorable au repérage de détails tels que :

- une fenêtre parmi seize, cf. l'image 3 (2) ;
- un canot de sauvetage d'un paquebot, cf. l'image 4 (2) ;
- un clocheton du Sacré Cœur, cf. l'image 5 (2) ;
- un porche parmi six, cf. l'image 11 (2) ;
- une cheminée parmi six, cf. l'image 12 (2) ;
- un mouton d'un troupeau, cf. l'image 18 (2).

Enfin, il convient de différencier les cibles qui sont des éléments graphiques à part entière, des cibles qui sont une partie d'un élément d'une scène. Les cibles des images (1), de même que les cibles des images 7 à 16 (2) sont des objets graphiques à part entière. Ainsi les sujets les identifient, puis les sélectionnent grâce à leurs propriétés visuelles, telles que couleur, forme, taille, ou encore caractéristiques géométriques. Quant aux cibles des images 1 à 6, 17 et 18 (2) qui sont des parties d'un élément de la scène, elles imposent aux sujets d'utiliser lors du repérage des relations de type partie-tout, c'est-à-dire des relations existant entre la cible et les objets graphiques l'entourant.

Position et saillance de la cible

Les autres critères adoptés pour la définition/sélection des cibles sont la position de la cible à l'écran et la saillance visuelle de la cible.

La position de la cible dans la scène est choisie selon un découpage de la scène en neuf parties définies comme suit : "en haut à gauche", "en haut", "en haut à droite", "à gauche", "au centre", "à droite", "en bas à gauche", "en bas, "en bas à droite". En choisissant des cibles dans ces neuf zones, on peut faire varier les indications orales des situations PO et PM, et utiliser notamment les prépositions projectives comme "en haut/bas", "à gauche/droite", "au centre", etc.

La saillance visuelle de la cible est définie, d'une part, par des critères visuels comme sa forme, sa couleur, sa taille et, d'autre part, par ses propriétés géométriques, par exemple sa position dans la scène, en premier/arrière plan. La saillance visuelle permet de tester l'efficacité des messages sonores sur des cibles plus ou moins saillantes.

N.B. La saillance d'un élément d'une image peut tenir à sa couleur, sa position (par exemple au centre de l'image, au premier plan, etc.), sa taille, son unicité ou celle de l'une de ses propriétés.

4.2.5 Critères d'élaboration des messages sonores

La définition des messages sonores s'appuie sur la caractérisation des messages sonores du paragraphe 4.1.3 à la page 31 dans laquelle nous avons défini sept types d'indications spatiales, à savoir ISA, ISR, ISI, ISA+ISR, ISA+ISI, ISR+ISI et ABSENCE. Nous avons supprimé finalement le type ISR+ISI à cause de la complexité syntaxique qu'il introduit. Nous souhaitons,

en effet, utiliser des messages sonores aussi directs et simples que possible, pour être faciles à assimiler.

Nous distinguons ainsi six types de messages sonores, selon que la position de la cible dans la scène est donnée :

- sous forme d'indication spatiale absolue (ISA) ; c'est le cas pour les images 1, 4, 5, 8, 10 (1) et les images 11, 14, 15, 16 (2) ;
- sous forme d'indication spatiale relative (ISR) ; c'est le cas pour les images 11 (1) et 1, 3, 4, 5, 7, 8, 9, 10, 12, 13, 18 (2) ;
- Sous forme d'indication spatiale implicite (ISI) ; c'est le cas pour les images 3, 14, 15, 16, 17 (1) ;
- sous forme d'indication spatiale absolue et relative (ISA+ISR) ; c'est le cas pour les images 2, 7, 9, 12 (1) et les images 2, 17 (2) ;
- sous forme d'indication spatiale absolue et implicite (ISA+ISI) ; c'est le cas pour les images 13, 18 (1) ;
- sans indication spatiale (ABSENCE) ; c'est le cas pour l'image 6 (1) et l'image 6 (2).

Les messages sont présentés en annexe A.

Sur douze messages ISR, un seul est associé à une image (1). Ceci s'explique par le fait que les messages de type ISR utilisent les relations[34] entre les objets de la scène et la cible. Ce type de relation est difficile à mettre en évidence pour les images (1). On n'observe aucune relation de type partie-tout pour les collections structurées d'objets arbitraires (1), les formes géométriques (1) ou les objets réels à connotation symbolique (1). De plus, le choix d'un objet de référence uniquement saillant visuellement ne suffit pas, le plus souvent, pour que le message sonore résultant nous ait paru efficace, c'est-à-dire non ambigu, sauf pour l'image 11(1). Pour éviter les ambiguïtés, nous utilisons plutôt des messages ISA+ISI. C'est le cas pour les images 2, 7, 9, 12 (1).

Quant aux messages ISI, aucun n'est associé à une image (2). Les messages ISI permettent aux utilisateurs d'inférer la position de la cible dans la scène, mais ne la verbalisent pas. Il est difficile d'introduire ce type de mobilisation des connaissances *a priori* des utilisateurs pour les images (2) qui sont des représentations de la réalité, puisque les cibles doivent être désignées par leur nom précis. En revanche, dans les scènes d'objets réels à connotation symbolique (1), ce type de message sonore nous a semblé efficace. C'est le cas pour les images 3, 14, 15, 16, 17 (1).

À noter que, si les messages ISA sont équitablement répartis entre les images des classes (1) et (2), ce n'est le cas ni pour les messages ISR, ni pour les messages ISI. À noter également que les messages ne sont pas équitablement répartis entre tous les types de messages sonores utilisés. On recense 9 messages ISA, 12 messages ISR, 5 messages ISI, 6 messages ISA+ISR, 2 messages ISA+ISI et 2 messages ABSENCE. Ceci s'explique par le fait que certains types de messages sonores comme ISI ou ISR ne sont pas adaptés à tous les types de scènes visuelles. Les types hybrides ISA+ISR et ISA+ISI ont été introduits essentiellement pour pallier l'inefficacité des indications ISR et ISI dans ces contextes ; c'est le cas pour les scènes 2, 7, 9, 12 (1) et 2, 17 (2) qui ont été couplées à des messages de type ISA+ISR et pour les scènes 13, 18 (1) auxquelles correspondent des messages de type ISA+ISI.

[34]Notion de distance, relation partie-tout, position de la cible par rapport à un objet de référence, etc.

Comme nous l'avons souligné au paragraphe 4.1.3 à la page 31, les cibles doivent toutes pouvoir être désignées oralement, de façon simple et non ambiguë, en termes de désignation de l'objet (DO) et de désignation spatiale (DS). Toutefois, pour l'image 18 (2), le critère de non-ambiguïté n'a pas été respecté. En effet, d'autres éléments, similaires visuellement, étaient situés dans la zone écran décrite par l'indication de localisation de type ISR "au fond". Cette cible est donc moins facile à identifier que les autres dans les situations PO et PM, en raison de l'ambiguïté de la désignation spatiale contenue dans le message sonore. Dans l'expérimentation à venir, nous éliminerons cette difficulté en n'utilisant des messages de type ISR que lorsqu'il n'y a aucune ambiguïté possible entraînée par la désignation spatiale.

4.2.6 Description des images, des cibles et des messages sonores associés

Les images de la classe 1

Les 18 images de la classe 1, qui comprend des représentations symboliques ou arbitraires de collections d'objets graphiques, se répartissent comme suit dans les trois sous-classes :

- d'objets symboliques, pour les images 1 à 6 (1) ;
- de formes géométriques arbitraires, pour les images 7 à 12 (1) ;
- d'objets réels dans des environnements symboliques (carte, plan, arborescence de photographies), pour les images 13 à 18 (1).

Dans les images 1, 2, 4 à 12, la représentation de l'ensemble d'objets ne fournit aucun indice visuel faisant appel aux connaissances ou à l'expérience antérieures de l'utilisateur, susceptible de faciliter/guider la recherche de la cible. Inversement, dans les images 3, 13 à 18, la structure spatiale de la représentation de l'ensemble d'objets (carte, plan ou arborescence) facilite le repérage de la cible, grâce à la mobilisation des connaissances et des expériences antérieures de l'utilisateur.

Nous avons varié la structure spatiale de la représentation des collections d'objets comme le présente le tableau ci-dessous.

Groupes d'images	**Non structurées**	**Matrices**	**Cercles**	**Cartes**
Objets symboliques	6	1,2,4,5		3
Formes géométriques	8,10,11,12		7,9	
Objets réels				13 à 18

TAB. 4.3 – Structures spatiales pour les images (1).
Dans la quatrième colonne, le mot "cercle" signifie structure circulaire comme les ellipses, les couronnes, les ovales, etc. Le tableau présente la répartition des images en fonction de leur structure : non structurées *versus* structurées comme matrices, cercles et cartes.

Les images de la classe 2

Les 18 images de cette classe, qui comprend des représentations réalistes ou simplifiées de scènes ou d'objets réels, se répartissent comme suit dans les trois sous-classes (en fonction de la nature de la cible et de l'image) :

- objets complexes isolés avec comme cible une partie de l'objet, pour les images 1 à 6 ;
- paysages ou scènes d'intérieur avec comme cible un élément du paysage (respectivement de la scène), pour les images 7 à 12 et, respectivement 13 à 15 ;
- groupes de personnages, qui constituent une sous-classe hybride au sens où un personnage peut être considéré comme un objet complexe (cf. la première sous-classe) et l'image comme une scène complexe (cf. la seconde sous-classe) ; ce sont les images 15 à 18.

Les images 1, 3, 7 et 8 (2) sont des dessins avec contour. Les images 2, 9 et 17 (2) sont des dessins sans contour. Les images 4 à 6, 10 à 12, 13 à 16 et 18 (2) sont des photographies.

4.2.7 Exemples d'images

Objets symboliques

La figure 4.2 à la page 42 donne un exemple de collection d'objets symboliques. Il s'agit de l'image 2 (1) dans laquelle tous les objets ont la même forme rectangulaire, la même taille et la même couleur bleu blanc rouge. La cible est un drapeau de la collection et est donc non saillante par sa forme, sa couleur, sa taille. Le message associé est : "Dans la première ligne, le drapeau à droite du drapeau Français". Il est du type ISA+ISR. Comme à la fois, la désignation de l'objet et l'indication de localisation spatiale sont non ambiguës, ce message est non ambigu.

L'image 1(1) est également une collection de drapeaux. Les images 4 et 5 (1) se composent d'une collection de panneaux routiers. En revanche, l'image 6 (1) est une collection non structurée d'objets symboliques : il s'agit d'une reproduction d'une œuvre de Miro.

Formes géométriques

La figure 4.3 à la page 42 donne un exemple de formes géométriques arbitraires 3D. Il s'agit de l'image 8 (1). Bien qu'elle soit très excentrée, la cible est saillante par sa taille par rapport aux autres éléments de la scène. Il y a 3 confusions possibles. Le message associé est : "En haut à droite, la forme en bois jaune". Il est du type ISA. Comme à la fois, la désignation de l'objet et l'indication de localisation spatiale sont non ambiguës, ce message est non ambigu.

L'image 7(1) est une photographie du génome, l'image 9(1) un dessin où des boules sont disposées en couronne. Les images 10 et 12 (1) sont des formes géométriques arbitraires 2D.

Objets symboliques sur une carte

La figure 4.4 à la page 43 donne un exemple de collection d'objets réels dans un environnement symbolique (carte). Il s'agit de l'image 16 (1). Malgré sa position au centre de l'écran, la cible est peu saillante (taille moyenne et couleurs discrètes). Le message associé est : "Le roi des animaux".

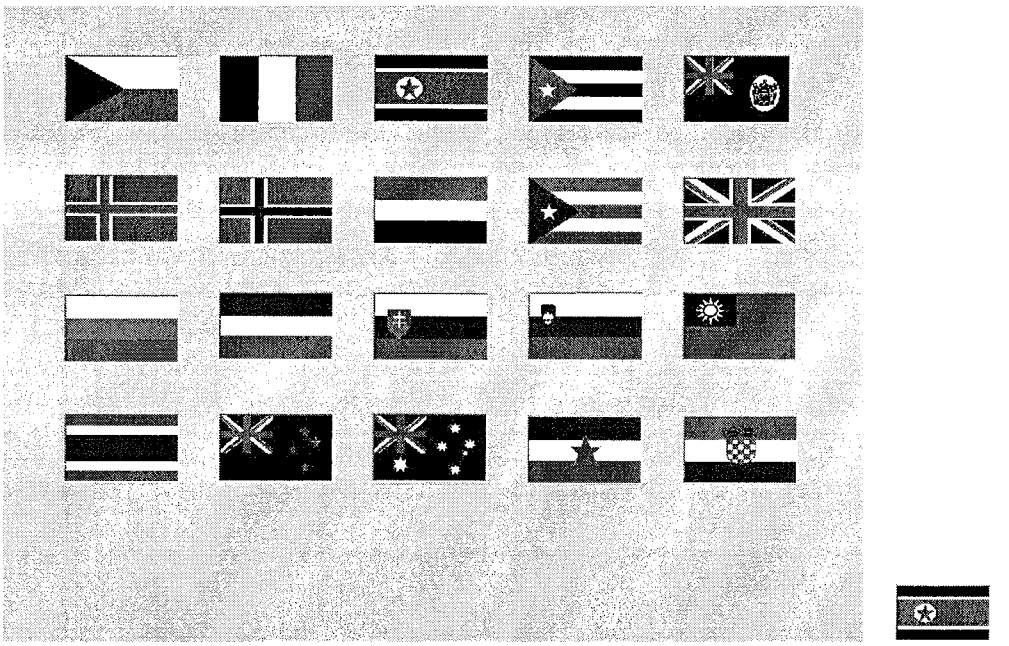


FIG. 4.2 – Exemple de collection d'objets symboliques : image 2 (1).
Caractéristiques : dessin 2D à structure matricielle peu dense, constitué de 20 éléments. Cible : non familière (drapeau de la Corée).

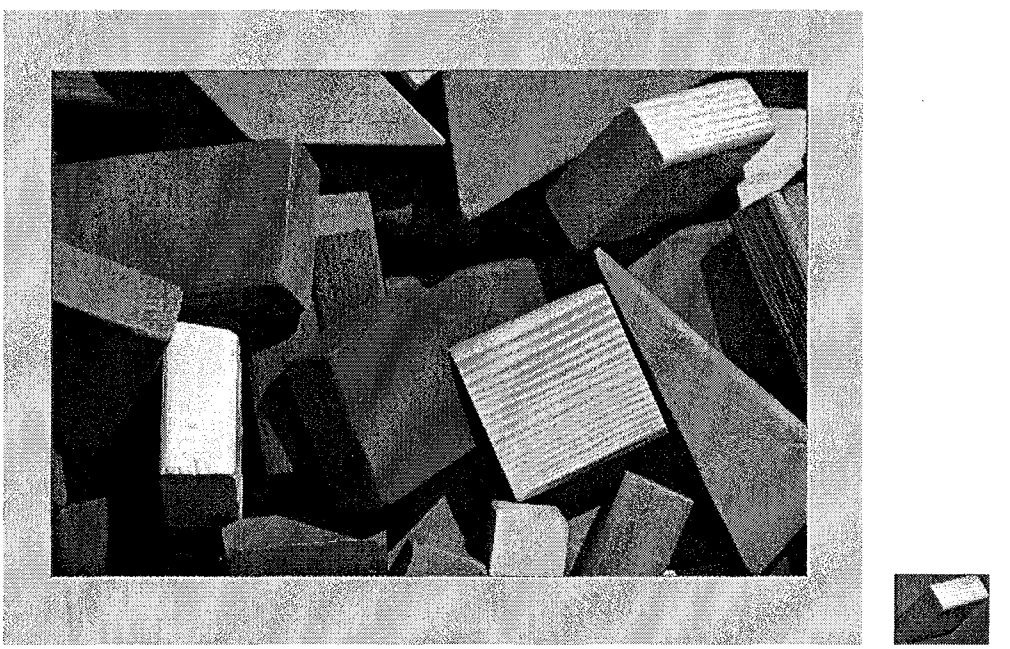


FIG. 4.3 – Exemple de formes géométriques arbitraires : image 8 (1).
Caractéristiques : dessin 3D non structuré, très dense, constitué de 30 éléments environ. Cible unique, de familiarité neutre (cube).

Il est du type ISI. Comme à la fois, la désignation de l'objet et l'indication de localisation spatiale sont non ambiguës, ce message est non ambigu.

L'image 13(1) est un arbre généalogique des rois de France, avec 12 visages de rois sous forme de médaillon. Les images 14, 15 et 17 (1) sont également des cartes sur lesquelles sont disposées des photographies d'objets réels : vins de France pour l'image 14 (1), monuments français pour l'image 15 (1) et monuments parisiens pour l'image 17 (1). L'image 18 (1) est un plan du réseau RATP de Paris.

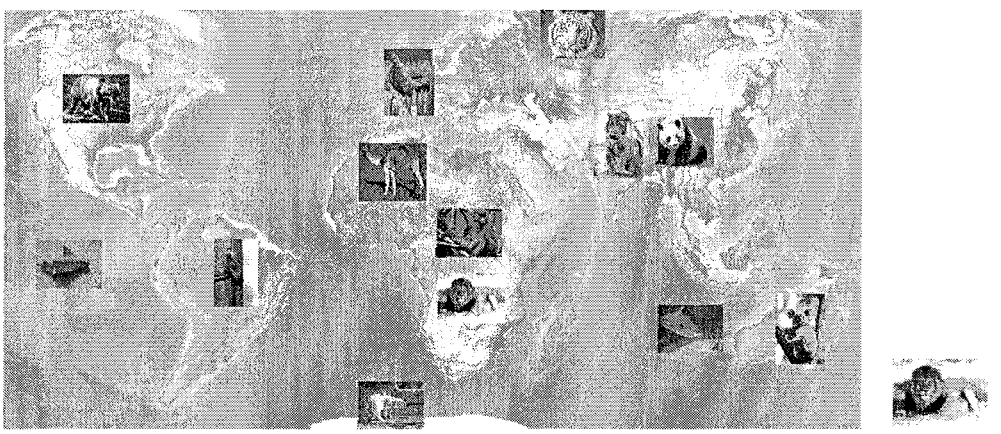


FIG. 4.4 – Exemple de collection d'objets réels sur une carte : image 16 (1).
Description : 13 photographies d'animaux du monde, disposées sur un planisphère selon leur origine géographique. Le planisphère étant un dessin, nous l'avons classée sous le type dessin, malgré les photographies d'animaux. Caractéristiques : structure géométrique peu dense et de couleur très claire. Cible associée : unique, très familière (photographie d'un lion en Afrique).

Objets complexes

La figure 4.5 à la page 44 donne un exemple d'objet complexe isolé : un bâtiment, dont toutes les fenêtres sont uniques grâce à leur position. Il s'agit de l'image 3 (2). La cible est une fenêtre. Elle est unique grâce au contexte visuel conservé : une partie du toit et une partie du ciel. Elle n'est ni saillante, car sa position est excentrée et sa forme identique à environ 20 autres fenêtres, ni familière, car ce n'est pas la façon classique dont on représente une fenêtre. Le message associé est : "Au premier étage, la fenêtre la plus à gauche". Il est du type ISR. Comme à la fois, la désignation de l'objet et l'indication de localisation spatiale sont non ambiguës, ce message est non ambigu.

L'image 1 (2) est le dessin d'un panier de fruits. L'image 2 (2) est un dessin d'immeubles. L'image 4 (2) est la photographie d'un bateau. L'image 5 (2) est une photographie du Sacré Cœur. Enfin, l'image 6 (2) est la photographie d'un pistolet à peinture.

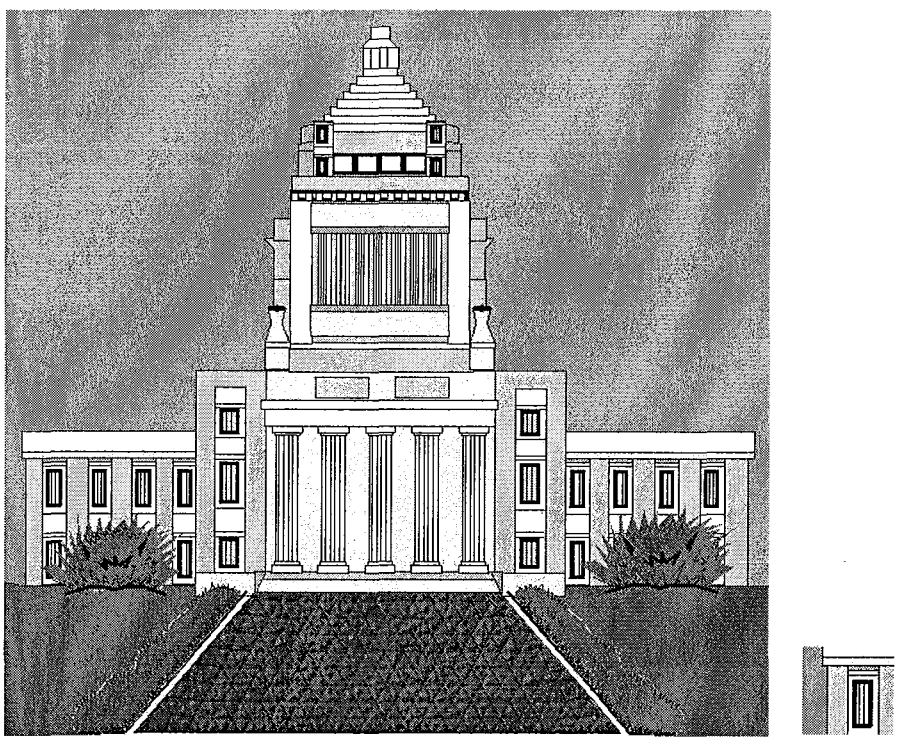


FIG. 4.5 – Exemple d'objet complexe : image 3 (2).
Caractéristiques : dessin 2D avec contours, à structure symétrique (axes horizontal et vertical du bâtiment) dense, constitué de 50 éléments. Cible associée : une fenêtre.

Paysages et scènes d'intérieurs

La figure 4.6 à la page 45 donne un exemple de paysage. Il s'agit de l'image 10 (2). La cible n'est pas saillante, puisqu'elle est en arrière-plan, loin de la zone saillante à droite, qui est au premier plan, et de petite taille par rapport aux autres éléments de la scène. Le message associé est : "A l'horizon, le plus gros des dômes". Il est de type ISR. Comme à la fois, la désignation de l'objet et l'indication de localisation spatiale sont non ambiguës, ce message est non ambigu.

Les images 7 à 9 (2) sont des dessins de paysages extérieurs, les images 11 et 12 (2) des photographies de paysages extérieurs et les images 13 à 15 (2) des photographies de scènes d'intérieur.

FIG. 4.6 – Paysage : image 10 (2).
Caractéristiques : photographie, très dense, constituée de 100 éléments environ. Cible unique, de familiarité neutre (dôme).

Groupes de personnages

La figure 4.7 à la page 46 donne un exemple de groupe de personnages. Il s'agit de l'image 18 (2). Elle représente un troupeau de moutons. La cible, une tête de mouton, est non saillante. Le message associé est : "Au fond, la tête du mouton". Il est du type ISR. Comme à la fois, la désignation de l'objet et l'indication de localisation spatiale sont ambiguës, ce message est ambigu.

Les images 16 et 17 (2) représentent également des groupes de personnages : une photographie peu dense de badauds en ville pour l'image 16(2) et un dessin représentant dix personnages pour l'image 17(2).

FIG. 4.7 – Groupe de personnages : image 10 (2).
Caractéristiques : photographie très dense et non structurée. Cible : une tête de mouton. Cible très ambiguë (5 confusions possibles), non saillante (position excentrée), non familière.

4.2.8 Contrebalancement des conditions expérimentales

Lors de l'expérience, les sujets avaient pour consigne de localiser puis de sélectionner à la souris une cible visuelle pour chacune des 36 images qui leur étaient présentées. Il leur était en outre demandé d'effectuer les 36 sélections d'objets graphiques, ou de parties d'objets graphiques, le plus rapidement possible. Ils ne pouvaient cliquer qu'une seule fois sur chaque image.

Chaque sujet effectue la tâche de repérage dans les trois situations PV, PO et PM pour 36 images en tout. Chaque sujet traite donc 12 couples de stimuli pour chaque ensemble PV, PO et PM (cf. le paragraphe 4.1 à la page 27). Plus précisément, chaque sujet effectue la tâche de repérage pour une série ou paquet de 12 images en situation PV, 12 images en situation PO et 12 images en situation PM. Cette contrainte définit trois paquets d'images P1, P2 et P3 et 3 groupes de six sujets G1, G2 et G3 comme le montre le tableau 4.4.

De façon à neutraliser les effets d'une familiarisation éventuelle avec l'activité de repérage, nous avons permuté l'ordre de passation des situations PV et PO. La moitié des sujets traite les trois paquets d'images dans l'ordre PV-PO-PM, l'autre moitié dans l'ordre PO-PV-PM. Cette contrainte divise chaque groupe de sujets G1, G2, G3 en deux. On obtient alors 6 sous-groupes de 3 sujets, comme le montre le tableau 4.5 : G11 et G12 pour le groupe G1, G21 et G22 pour le groupe G2, G31 et G32 pour le groupe G3. Les sous-groupes G11, G21 et G31 ont traité les images dans l'ordre PV-PO-PM, les sous-groupes G21, G22 et G23 dans l'ordre PO-PV-PM. Les images dans la situation PM sont toujours traitées en fin de session.

La condition PM a été effectuée en dernier par l'ensemble des sujets pour que ceux-ci l'abordent avec une expérience identique des présentations monomodales, orales et visuelles,

de la cible ; ces deux modalités étant combinées dans la condition PM. Autrement, l'interprétation des performances des sujets dans la condition PM aurait été difficile à réaliser en raison des facteurs potentiels supplémentaires introduits (i.e., expérience des présentations visuelles, expérience des présentations orales, ou expérience des présentations visuelles et orales).

Groupe de sujets	**Associations entre paquets d'images et situations de présentation**
Groupe G1	(P1 ;PV) (P2 ;PO) (P3 ;PM)
Groupe G2	(P3 ;PV) (P1 ;PO) (P2 ;PM)
Groupe G3	(P2 ;PV) (P3 ;PO) (P1 ;PM)

TAB. 4.4 – Répartition des paquets d'images entre les 3 groupes de sujets G1, G2 et G3. Chaque paquet d'images P1, P2 et P3 est traité dans les trois situations PV, PO et PM, par un groupe de sujets différent.

Sous-Groupe	**PV**	**PO**	**PM**	**Sous-Groupe**	**PO**	**PV**	**PM**
G11	P1	P2	P3	G12	P2	P1	P3
G21	P3	P1	P2	G22	P1	P3	P2
G31	P2	P3	P1	G32	P3	P2	P1

TAB. 4.5 – Répartition des paquets d'images entre les 6 sous-groupes de sujets G11, G12, G21, G22, G31 et G32.
Chaque paquet d'images est affecté au même couple (groupe, situation), mais au sein d'un couple, on balance l'ordre PV-PO-PM PO-PV-PM.(cf. le tableau 4.4).

4.2.9 Profil des sujets

Pour que chaque image soit traitée dans chacune des situations (PV, PO, PM), le même nombre de fois, il faut que le nombre de sujets soit un multiple de trois. Pour balancer l'ordre des conditions PV et PO, il faut que le nombre de sujets soit pair.

Nous avons choisi 18, car 12 était trop juste pour obtenir un volume de données suffisant pour valider nos hypothèses. Ce nombre restreint de sujets se justifie par le caractère exploratoire de notre étude. En effet, 18 sujets pour 36 images nous ont fourni une quantité de données suffisante[35] pour tester nos hypothèses de travail A, B et C.

En outre, des différences interindividuelles trop importantes n'auraient pas permis d'exploiter et interpréter les données de façon significative (temps et précision des sélections). C'est pourquoi, nous avons sélectionné les sujets de sorte qu'ils forment un groupe homogène. Non seulement ils devaient tous être en mesure de réaliser les tâches demandées avec succès, mais surtout ils devaient avoir une expérience comparable de la manipulation de la souris.

[35]soit 648 données en tout, échecs et succès confondus.

Ainsi, nous avons choisi les 18 sujets parmi les chercheurs et étudiants en informatique du LORIA. Ils avaient tous entre 18 et 29 ans et une vue normale (excepté pour un sujet daltonien). Tous les participants étaient utilisateurs experts de la souris avec des réactions motrices similaires, compte tenu de la tranche d'âge choisie[36].

4.2.10 Répartition du matériel visuel entre les groupes de sujets

La moitié des images appartient à la classe 1, qui regroupe les collections d'objets symboliques, les formes géométriques et les objets réels sur carte. L'autre moitié des images appartient à la classe 2, qui regroupe les objets complexes, les paysages ou scènes d'intérieur et les groupes de personnages.

Pour éviter tout phénomène d'apprentissage à la tâche de repérage de cibles visuelles, donc, pour éviter les effets d'accoutumance à un type d'image, nous avons ordonné de manière aléatoire les 12 images des séries P1, P2 et P3 comme suit :

- la série P1 contient, dans l'ordre, les images : 13(2), 3(1), 12(1), 16(1), 7(2), 7(1), 9(2), 5(2), 16(2), 1(2), 14(1), 6(2) ;
- la série P2 contient, dans l'ordre, les images : 15(2), 17(2), 18(2), 1(1), 4(2), 6(1), 3(2), 10(1), 12(2), 9(1), 5(1), 11(2) ;
- la série P3 contient, dans l'ordre, les images : 13(1), 15(1), 10(2), 18(1), 8(2), 4(1), 2(1), 14(2), 17(1), 2(2), 8(1), 11(1).

4.2.11 Déroulement de l'expérience

Consignes et entraînement des sujets

Un entraînement du sujet sur six images succède à la lecture des consignes par l'expérimentateur. Cet entraînement comprend deux images par situation, dans l'ordre PV-PO-PM, respectivement PO-PV-PM, si le sujet passe l'expérimentation dans l'ordre PV-PO-PM, respectivement PO-PV-PM. L'expérimentateur commente la consigne et présente les tâches. Il assiste à l'entraînement du sujet, puis répond à ses éventuelles questions. Il quitte la salle pendant la passation, mais est disponible par téléphone. Le sujet dispose en permanence d'une version écrite de la consigne et du numéro où il peut contacter l'expérimentateur.

Nous n'avons recueilli, pour les tâches d'entraînement, ni les échecs, ni les temps de repérage des cibles, car le seul objectif de cet entraînement était de s'assurer que les sujets avaient bien compris les consignes et étaient en mesure de les appliquer.

Passation

La durée globale d'une passation est d'environ 20 minutes et comprend :

- la présentation des consignes et l'entraînement initial (environ 5 minutes) ;

[36] cf. par exemple, les travaux de [Dollinger et Hoyer, 1996] concernant les effets de l'âge sur les performances motrices des utilisateurs dans les tâches visuelles.

- la réalisation des tâches de repérage (entre 4 et 5 minutes) ;
- le remplissage du questionnaire post-session et l'entretien final (environ 10 minutes).

Questionnaire post-session

À la fin de la passation, les sujets remplissent deux questionnaires, l'un portant sur la difficulté des différentes tâches traitées, l'autre sur l'évaluation de l'apport des messages oraux.

Dans le premier questionnaire, nous avons demandé aux sujets d'évaluer la difficulté de chaque image, sur une échelle de 1 (très facile) à 6 (très difficile). Il serait peut-être préférable que ce questionnaire soit rempli au fur et à mesure de la réalisation des tâches afin d'obtenir leurs réactions immédiates plutôt que leur reconstruction plus ou moins fidèle *a posteriori*, mais cela risquerait de distraire les sujets dans leurs repérages visuels. Pour pallier cet inconvénient, nous avons mis à la disposition des sujets un jeu de reproductions en couleur sur support papier des 36 images, rangées dans l'ordre où ils les avaient traitées.

Dans le second questionnaire, nous avons demandé aux sujets d'évaluer l'apport des messages oraux en termes d'efficacité et de confort du repérage dans les situations PO et PM par rapport à la situation de référence PV. L'efficacité est évaluée en terme de rapidité de l'identification des cibles. Le confort est évalué en termes de facilité de repérage, de charge de travail et de fatigue. Le second questionnaire comprenait également des questions sur l'aide éventuelle apportée et la gêne éventuelle introduite par les messages oraux dans les situations PO et PM et sur la pertinence des informations orales fournies dans les conditions PO et PM. Il était enfin demandé aux sujets de classer les trois situations de présentation des cibles PV, PO et PM par ordre décroissant de préférence et d'efficacité[37].

Le remplissage des questionnaires est suivi d'un bref entretien semi-directif, visant à développer les thèmes abordés dans le second questionnaire. Ces entretiens n'ont pas été enregistrés pour que les sujets s'expriment aussi spontanément que possible ; ils ont fait l'objet d'une prise de notes à la volée par l'expérimentateur.

4.2.12 Mesures

Pour mesurer l'apport de la parole, nous nous sommes bornés à enregistrer, pour chaque tâche, d'une part, le temps en millisecondes mis par le sujet pour sélectionner (cliquer sur) la cible et, d'autre part, la précision de la sélection (dans la cible ou en dehors). Pour obtenir des mesures comparables concernant les temps de sélection, la souris est repositionnée au centre de l'écran avant la présentation de chaque image (cf. le paragraphe 4.2.2 à la page 34).

Nous avons choisi, dans un premier temps, cette mesure plutôt que le temps nécessaire au sujet pour repérer visuellement la cible, par analyse des fixations oculaires. En effet, en situation d'IHM, le repérage s'inscrit le plus souvent dans une perspective de sélection à la souris de la cible. On peut donc considérer le temps de sélection de la cible à la souris comme une mesure globale de la boucle perception-action contenue dans la tâche du repérage visuel de cibles [Shneiderman, 1983].

[37] cf. annexe A.

Au chapitre 6, nous décrivons une vérification expérimentale de la cohérence des résultats fournis par la mesure des temps de sélection des cibles par rapport à ceux fournis par l'analyse des fixations oculaires, avec un eye-tracker. En effet, d'un point de vue théorique, mesurer le repérage visuel d'un point de vue perceptif exclusivement consiste à décomposer artificiellement la boucle perception-action [Rasmussen, 1986]. Les fixations oculaires sont un indice pertinent et fiable du temps requis pour repérer la cible. Par leur analyse, nous souhaitons en outre obtenir des indications sur les stratégies d'exploration visuelle des scènes.

4.3 Analyse des données : méthodologie

Notre analyse des données a été réalisée selon les deux axes suivants : analyse quantitative et analyse qualitative.

Dans l'analyse quantitative des données, nous montrons l'influence de la parole sur les performances des sujets dans les tâches visuelles. Cette analyse est présentée dans le paragraphe 4.4 à la page 50. Elle porte sur :

- les performances globales des sujets dans les trois situations de présentation des cibles ;
- la comparaison inter-modalités de ces résultats globaux ;
- les performances des sujets selon le type d'images traitées (classe 1 *versus* classe 2) ;
- les performances des sujets selon le type d'indication orale contenue dans les messages sonores (ISA *versus* ISR, etc.).

Dans l'analyse qualitative des données, nous montrons l'influence des messages sonores sur l'efficacité de l'interaction pendant la réalisation des tâches visuelles (étude comparative), d'une part, et évaluons la satisfaction des utilisateurs (étude subjective), d'autre part. Cette analyse est présentée dans le paragraphe 4.5 à la page 59. Elle porte sur :

- la comparaison des erreurs des sujets dans les situations de présentation PO et PM par rapport à la situation PV ;
- le dépouillement des questionnaires utilisateurs et la comparaison des résultats globaux relatifs à chacune des modalités ;
- les performances des sujets selon le type d'images traitées (classe 1 *versus* classe 2) ;
- les performances des sujets selon le type d'indication orale contenue dans les messages sonores (ISA *versus* ISR, etc.).

À noter que nous avons dû écarter deux scènes (toutes deux appartenant à la classe 1 d'images) des analyses, en raison d'incidents techniques. Il s'agit des images 8(1) et 15(1).

4.4 Étude quantitative

L'exploitation des données pour l'analyse quantitative est basée sur les mesures suivantes :

- le calcul de la moyenne arithmétique des temps de repérage visuel des cibles : il s'agit des temps nécessaires à la localisation et à la sélection des cibles visuelles ;

- le nombre d'erreurs commises par les sujets : il s'agit des sélections d'objets graphiques en dehors de la cible associée à la scène.

L'objectif de cette étude est de valider les hypothèses de travail A, B et C[38].

4.4.1 Résultats globaux

Pour tester la validité des hypothèses de travail A et B, nous avons mené une étude comparative entre les trois types de présentation des cibles, à savoir les situations PV, PO et PM. Les mesures ont été réalisées de la manière suivante. Nous avons regroupé les données recueillies lors de l'expérience par type de présentation des cibles, autrement dit par condition expérimentale. Nous obtenons 612 données par modalité PV, PO et PM : chaque donnée est soit un temps de sélection de la cible, soit une erreur.

Rappelons que la situation PV est considérée comme situation de référence. Nous avons donc comparé les temps et la précision des sélections observés en situation PV, à ceux observés en situation PO (validation de l'hypothèse B), et à ceux observés en situation PM (validation de l'hypothèse A)[39].

Présentation orale *versus* présentation visuelle

En considérant la précision des sélections, les messages oraux semblent plus efficaces que les présentations visuelles des cibles. C'est ce qui ressort des comparaisons entre les situations PV et PO. Néanmoins, les sélections sont moins rapides en l'absence de présentation visuelle des cibles. Le nombre total des erreurs produites en situation PO est inférieur de de 55% par rapport à la situation PV, tandis que le temps moyen de sélection des cibles en situation PO est supérieur de 28% par rapport à la situation PV. Ces différences sont statistiquement significatives.

Le temps moyen plus lent dans la situation PO, associé à un écart type plus élevé, peut être expliqué par le caractère inhabituel des tâches de recherche visuelle en situation PO ; cette situation leur est bien moins familière que les situations PV et PM qui se produisent dans la vie de tous les jours. Par conséquent, la variabilité élevée des temps de sélection dans la situation PO peut refléter la diversité interindividuelle des capacités et processus cognitifs impliqués dans l'apprentissage.

Présentation multimodale *versus* présentation visuelle

En considérant la précision des sélections, les messages multimodaux semblent plus efficaces que les présentations visuelles des cibles, d'après les comparaisons entre les situations PV et PM. En effet, on observe 75% d'erreurs en moins dans la situation PM par rapport à la situation PV. Cette différence est statistiquement très significative (cf. le tableau 4.6, t=-3,94 ; p<0,0001).

[38] cf. supra paragraphe 4.1.1 page 28.
[39] cf. infra tableau 4.6 page 56.

En outre, les sélections sont plus rapides dans la situation PM par rapport à la situation PV : 2,7 secondes dans la situation PM *versus* 2,83 secondes dans la situation PV. Cette différence de l'ordre du dixième de seconde n'est pas statistiquement significative.

Présentation multimodale *versus* présentation orale

En considérant la précision des sélections, les messages multimodaux semblent plus efficaces que les présentations orales des cibles, d'après les comparaisons entre les situations PM et PO. En effet, on observe 57% d'erreurs en moins dans la situation PM par rapport à la situation PO. Au sens statistique, cette différence indique une tendance (cf. le tableau 4.6, t=-1,31 ;p<0,189).

En outre, les sélections sont plus rapides dans la situation PM que dans la situation PO : 2,7 secondes dans la situation PM *versus* 3,92 secondes dans la situation PO. Cette différence est statistiquement très significative (cf. le tableau 4.6, t=-4,2 ;p<0,0001).

Test de nullité des différences

Nous avons réalisé un test statistique de nullité des différences des temps de sélection des cibles entre les trois situations PV, PO et PM. On définit trois variables a, b et c telles que :

- la variable a contienne, pour les 34 images, la différence entre les temps moyens par image observés entre les situations PO et PV ;
- la variable b contienne, pour les 34 images, la différence entre les temps moyens par image observés entre les situations PO et PM ;
- la variable c contienne, pour les 34 images, la différence entre les temps moyens par image observés entre les situations PV et PM.

La nullité des différences consiste à tester la nullité de chaque variable et montre que :

- il existe une tendance statistique selon laquelle la situation PV réduit les temps de sélection des cibles par rapport à la situation PO (a ; t=1,89 et p=0,067) ;
- il existe une différence statistique très significative, en terme de temps de sélection des cibles, entre les situations PO et PM (b ; t=2,83 et p=0,008) : la situation PM réduit les temps de sélections des cibles par rapport à la situation PO ;
- il n'existe pas de différence significative en terme de temps de sélection des cibles entre les situations PV et PM (c ; t=1,49 et p=0,145).

Interprétation des résultats

Ces résultats suggèrent que, dans la situation PM, les sujets tirent avantage de l'information spécifique apportée par chaque modalité, qu'elle soit visuelle ou orale. Autrement dit, la pauvreté relative des informations visuelles, respectivement orales, peut être compensée par la richesse des informations orales, respectivement visuelles. Plus précisément :

- la modalité orale fournit aux sujets une indication, à la fois, spatiale et dénominative (cf. le paragraphe 4.1.3 à la page 31). Les sujets cherchent à localiser un objet inconnu visuellement, mais dont ils connaissent la localisation et le nom. Ces deux informations

permettent de pallier les éventuelles ambiguïtés visuelles entre la cible et d'autres objets de la scène ;

- la modalité visuelle "montre" la cible aux sujets. Ils disposent alors de caractéristiques visuelles sur l'objet graphique à sélectionner, telles que son type, ses propriétés graphiques (forme, taille, couleur), ses propriétés géométriques, etc. Ces indications leur permettent alors d'identifier rapidement la cible dans la scène, impliquant des processus complexes de prise de décision.

Nous avons été surpris que l'indication spatiale exprimée oralement dans les messages multimodaux, n'améliore de façon statistiquement significative, ni le temps de sélection des cibles par rapport à la situation PV, ni la précision des sélections des cibles par rapport à la situation PO. D'une part, ce résultat peut être expliqué dans le cadre des modèles de perception visuelle qui supposent que les mouvement oculaires sont moins influencés par les processus cognitifs (descendants) que par les stimuli visuels (ascendants), lors des tâches d'exploration visuelle. Voir par exemple, [Henderson et Hollingworth, 1998]. D'autre part, ce résultat est compatible avec les modèles cognitifs de traitement des stimuli multimodaux. Ces modèles, tel que celui proposé dans [Engelkamp, 1992], montrent l'existence d'interactions de haut niveau entre perception et interprétation qui résultent d'interférences ou de collaborations entre les processus visuels et auditifs de bas niveau.

Toutefois, il convient de noter les temps de sélection des cibles plus courts dans la situation PM que dans la situation PO (résultat statistiquement significatif). Ce résultat semble tenir au fait que la tâche de détection de cible est familière dans les situations PV et PM, contrairement à la situation PO. En effet, la détection de cibles dans la situation PV, d'ailleurs situation de référence du protocole expérimental, ressemble aux tâches de repérage visuel auxquelles sont habitués les utilisateurs. Il en est de même pour la situation PM puisqu'elle comporte, outre un message sonore, l'affichage de la cible à détecter comme dans la situation PV. En revanche, la tâche de détection de cibles dans la situation PO est nouvelle en raison de l'absence de présentation visuelle des cibles.

À noter également, que les sujets effectuent les tâches de repérage dans la situation PO avant de les effectuer dans la situation PM. L'apprentissage de la tâche, qui tient à l'ordre des situations PV, PO et PM, peut également être une interprétation possible des résultats observés sur les temps de sélection des cibles dans les situations PO et PM.

D'autres études mettant en œuvre un protocole expérimental similaire d'interaction[40] entre perceptions visuelle et auditive améliore le repérage de la cible dans la situation expérimentale PM. Il est nécessaire de raffiner les caractérisations visuelles et sémiotiques des scènes et des cibles, afin de disposer d'un plus vaste ensemble d'images et de couples (scène + cible) de difficulté équivalente. En effet, la "difficulté" de la tâche de repérage varie d'une scène à l'autre. Par exemple, dans la condition PV, tous les sujets (6) ont échoué sur l'image 5(2). De plus, les 31 erreurs observées dans la situation PV se sont produites sur 13 scènes seulement, soit à peine 40% du matériel visuel. Un ensemble d'images homogènes est indispensable afin de réaliser des comparaisons intra- et interindividuelle, puisque le même couple (scène + cible) ne peut être traité par le même sujet dans les trois situations PV, PO et PM.

[40] À savoir, compétition, synergie, ou complémentarité.

Conclusion

En conclusion, ces résultats valident l'hypothèse A, mais ne confirment que partiellement l'hypothèse B. Toutefois, si notre interprétation des temps de sélection plus longs en situation PO est correcte, l'hypothèse B semblerait validée pour les utilisateurs familiarisés avec les tâches visuelles de sélection de cibles à partir des seules indications orales.

Ces résultats globaux quantitatifs suggèrent également des recommandations utiles à fournir aux concepteurs d'interfaces graphiques. Dans le but de faciliter et d'améliorer l'efficacité de la recherche visuelle sur des écrans encombrés, deux formes d'aide utilisateur devraient s'avérer utiles :

- si l'objectif visé est la seule précision de la localisation de la cible, alors un message oral contenant la désignation verbale non ambiguë de l'objet graphique et l'indication spatiale de sa localisation dans la scène sont suffisants ;
- si les objectifs visés sont, à la fois, précision et rapidité de localisation, alors un message multimodal est plus approprié. Il s'agit de messages incluant une présentation visuelle isolée de la cible et un message oral contenant les mêmes informations que celles décrites au point précédent.

En fait, les trois situations de présentation de la cible PV, PO et PM induisent deux types de recherche visuelle dans la scène :

- soit la cible est connue visuellement à l'affichage de la scène, ce qui correspond aux situations PV et PM. Dans ce cas, le sujet reconnaît un objet graphique dont les propriétés visuelles lui ont été présentées antérieurement ;
- soit la cible est inconnue visuellement à l'affichage de la scène, ce qui correspond à la situation PO. Dans ce cas, seul le message sonore permet au sujet d'identifier et de localiser la cible, par localisation dans la scène, puis désignation.

Toutefois, des recherches expérimentales supplémentaires sont nécessaires pour la confirmation de ces recommandations. Elles ont, en effet, été inférées à partir d'un échantillon relativement faible de données et de mesures expérimentales.

4.4.2 Étude détaillée

Résultats par type d'image

Les performances des sujets, regroupées par classe de scènes et type de présentation des cibles, sont présentées dans le tableau 4.7 à la page 56. Nous avons calculé le pourcentage d'erreurs pour 96 instances de la classe 1 (deux images ayant été retirées, cf. le paragraphe 4.3 à la page 50) et 108 instances de la classe 2.

Les messages multimodaux s'avèrent être plus efficaces, en particulier pour les scènes de la classe (1) que les présentations visuelles isolées de la cible ou les présentations orales[41].

Pour les images de la classe 1 :

[41] La classe 1 d'images représente des objets graphiques symboliques ou arbitraires.

- les temps moyens de sélection des cibles dans la situation PM sont plus courts de 7%, respectivement 30%, que ceux observés dans la situation PV, respectivement PO ;
- le nombre d'erreurs moyen dans la situation PM est inférieur de 86%, respectivement 73%, au nombre d'erreurs moyen observé dans la situation PV, respectivement PO.

Pour les images de la classe 2 :

- les temps moyens de sélection des cibles dans la situation PM sont similaires à ceux observés dans la situation PV et inférieurs de 33% à ceux observés dans la situation PO ;
- le nombre d'erreurs moyen dans la situation PM est inférieur de 35% au nombre d'erreurs moyen observé dans la situation PV et similaire à celui observé dans la situation PO.

Ces résultats renforcent l'interprétation selon laquelle, dans la situation PM, les sujets tirent avantage de l'information spécifique apportée par chaque modalité visuelle ou orale (cf. la première interprétation à la page 52).

Enfin, les temps de sélection moyens plus longs pour la classe 1 que pour la classe 2 peuvent s'expliquer de la façon suivante : si la cible est un objet réel familier (tel qu'un téléphone) dans une scène réaliste familière (telle qu'un bureau), alors l'exploration visuelle de la scène est facilitée par les connaissances *a priori* sur la structure standard de la scène et la position usuelle de la cible dans celle-ci.

Une telle connaissance n'est pas disponible dans le cas de scènes non réalistes, comme celles de la classe 1. La structure de la scène ainsi que les positions possibles de la cible dans celle-ci ne peuvent être prévues grâce aux connaissances *a priori*. Donc, une recherche plus minutieuse dans la scène, ou même une exploration exhaustive de la scène, est nécessaire pour localiser la cible. Ces hypothèses pourraient également expliquer pourquoi les présentations multimodales des cibles s'avèrent plus efficaces pour les scènes appartenant à la classe 1 que pour celles appartenant à la classe 2 : à la fois, l'information visuelle et l'information orale compensent le manque de connaissances *a priori*.

Résultats par type de message

Pour tester la validité de l'hypothèse de travail C, nous avons mené une étude comparative entre les cinq types de messages sonores expérimentés (cf. le paragraphe 4.1.3 à la page 31). Nous avons regroupé les données (i.e., temps et précision de sélection des cibles visuelles) selon le type de message sonore, à savoir : ISA, ISR, ISI, ISA+ISR, ISA+ISI. Les performances des sujets, regroupées selon le type de message, sont présentées dans le tableau 4.8 à la page 57. Comme le nombre d'images par type de message varie d'un type à l'autre, les pourcentages d'erreurs ont été calculés pour : 48 instances du type ISA, 72 instances du type ISR, 24 instances du type ISI, 36 instances du type ISA+ISR et 12 instances du type ISA+ISI, soit 192 instances en tout[42].

Pour chaque catégorie de messages, nous avons comparé les performances par sujet et par situation (PV, PO et PM). Les résultats montrent, d'une part, que les messages donnant une indication spatiale absolue et/ou relative améliorent remarquablement la précision de la sélection

[42] Au lieu $6 \times 36 = 204$ instances. Pour mémoire, deux scènes ont été retirées de l'exploitation des données (cf. supra 4.3 page 50). Deux autres scènes n'ont pas été prises en compte, en raison de l'absence d'information spatiale (cf. les messages sonores de type ABSENCE).

Présentation des cibles	Nombre d'erreurs	Temps moyens de sélection (sec.)	Écart type (sec.)
PV	31	2,83	1,70
PO	14	3,92	3,50
PM	**8**	**2,70**	1,93

Présentation des cibles	Nombre d'erreurs		Temps moyens de sélection des cibles (sec.)	
PO *versus* PV	t=-2,70	**p=0,007**	t=3,79	**p=0,0002**
PM *versus* PV	t=-3,94	**p<0,0001**	t=-0,70	p=0,4852
PM *versus* PO	t=-1,31	p=0,189	t=-4,2	**p<0,0001**

TAB. 4.6 – Résultats par type de présentation des cibles.
Les meilleurs résultats sont en gras, les plus mauvais sont soulignés. Les résultats significatifs issus des tests statistiques sont en gras.

Présentation des cibles	Pourcentage d'erreurs	Temps moyens de sélection (sec.)	Écart type (sec.)
PV-CLasse 1	14,6	3,27	1,94
PV-Classe 2	15,7	2,43	1,39
PO-CLasse 1	7,3	4,3	4,09
PO-Classe 2	6,5	3,58	2,87
PM-CLasse 1	2	3,03	2,36
PM-Classe 2	5,5	2,40	1,36

TAB. 4.7 – Résultats par type de présentation des cibles et type d'image.
Nous n'avons pas réalisé de tests statistiques sur les données ainsi regroupées en raison du nombre restreint de données par groupe.

Situation PV (référence)			
Scènes regroupées par type de message	Pourcentage d'erreurs	Temps moyens de sélection (sec.)	Écart type (sec.)
ISA	10,4	2,87	1,19
ISR	26,4	2,86	2,14
ISI	4,17	1,84	0,57
ISA+ISR	13,89	3,57	1,99
ISA+ISI	8,33	3,54	0,98

Situation PO			
Scènes regroupées par type de message	Pourcentage d'erreurs	Temps moyens de sélection (sec.)	Écart type (sec.)
ISA	**0**	2,91	3,41
ISR	**8,33**	4,03	5,94
ISI	16,67	6,12	3,78
ISA+ISR	**5,56**	3,82	3,78
ISA+ISI	16,67	5,19	3,37

Situation PM			
Scènes regroupées par type de message	Pourcentage d'erreurs	Temps moyens de sélection (sec.)	Écart type (sec.)
ISA	4,17	2,42	1,41
ISR	8,33	3,12	2,43
ISI	0	2,1	1,06
ISA+ISR	0	2,82	1,84
ISA+ISI	0	2,98	2,53

TAB. 4.8 – Résultats par type de message et de présentation des cibles.
Situation PO : les meilleurs résultats sont en gras, les plus mauvais sont soulignés.

des cibles (cf. les messages de type ISA, ISR, ISA+ISR). D'autre part, l'utilité des messages ISI semble discutable, au moins pour la situation PO. Leur efficacité dans la situation PM montre la complexité des processus cognitifs mis en jeu lors de l'interprétation de stimuli multimodaux.

La moyenne des temps de sélection avec les messages de type ISR et ISI est plus élevée dans la situation PO (4,03 sec. ; 6,12 sec.) que dans les situations PV (2,86 sec. ; 1,84 sec.) et PM (3,12 sec. ; 2,1 sec.). Nous proposons pour expliquer ces deux phénomènes l'interprétation suivante.

On suppose que l'information spatiale relative contenue dans les messages de type ISR modifie et complique la stratégie d'exploration de scènes dans la situation PO, par rapport à l'information spatiale absolue contenue dans les messages de type ISA. Vraisemblablement, la stratégie d'exploration de la scène visuelle mise en œuvre après un message sonore de type ISA consiste simplement à localiser l'objet graphique désigné oralement. En revanche, la stratégie d'exploration visuelle mise en œuvre après un message sonore du type ISR pourrait comporter deux étapes :

- la première serait la localisation de l'objet graphique auquel il est fait référence dans le message sonore ; nous l'appellerons "objet de référence" ;
- la seconde serait l'exploration du voisinage de l'objet de référence à l'aide, probablement, de la vision périphérique ; la localisation de l'objet de référence et de la cible ne nécessite ainsi qu'une fixation oculaire (cf. les travaux de van Diepen dans [Van Diepen *et al.*, 1998]).

De la même façon, l'information spatiale implicite contenue dans les messages de type ISI, entraîne des processus cognitifs susceptibles de ralentir la localisation et la sélection visuelle des cibles. La stratégie d'exploration visuelle développée suite aux messages sonores de type ISI pourrait comporter également deux étapes :

- la première serait l'interprétation du message sonore de type ISI, mobilisant les connaissances *a priori* des utilisateurs ;
- la seconde serait l'exploration de la scène pour localiser la cible, grâce à l'indication spatiale implicite.

Ces interprétations permettent d'expliquer également pourquoi les messages de type ISR et ISI impliquent des temps de sélection légèrement plus longs dans la situation PM par rapport à la situation PV. Peut-être le repérage de cibles est-il ralenti par rapport à la situation PV en raison des processus cognitifs plus compliqués qui semblent être impliqués lorsque les présentations multimodales sont composées de messages ISR ou ISI. Il convient de noter, toutefois, que les présentations multimodales composées de messages ISR ou ISI entraînent un taux d'erreurs moins important par rapport aux présentations visuelles de cibles. D'autres études spécifiques sur ces points précis sont nécessaires pour juger de l'utilité de tels messages sonores.

Globalement, i.e., si on considère à la fois temps et précision de la sélection des cibles, tous les types de messages sonores sont efficaces dans la situation PM. Les couples (temps + précision) sont meilleurs dans la situation PM par rapport à la situation PV. En revanche, seules les indications spatiales absolues sont efficaces dans la situation PO. En effet, on observe, dans la situation PO, des différences importantes de temps et de précision des sélections entre les messages contenant une indication spatiale absolue et les autres types de messages oraux. Cette conclusion valide partiellement l'hypothèse C.

4.5 Étude qualitative

L'étude qualitative porte sur l'analyse des erreurs des sujets, d'une part, et sur le dépouillement des questionnaires post-session, d'autre part. L'objectif principal de cette analyse est de comprendre comment les messages sonores aident les utilisateurs pour les tâches d'exploration visuelle de scènes et d'évaluer leur apport, en termes de confort et d'efficacité de l'interaction.

4.5.1 Méthode d'analyse

L'analyse qualitative des performances des sujets devait s'appuyer sur la caractérisation *a priori* des matériels visuel et sonore[43]. Mais, nous l'avons jugée insuffisante pour expliquer certains résultats. Nous avons alors adopté d'autres critères pour la caractérisation des matériels visuel et sonore :

- la complexité des scènes, établie selon le nombre d'objets affichés. Cette caractéristique permet d'introduire la notion de densité des images et d'expliquer d'éventuelles erreurs de sélection des cibles ;
- le type d'affichage (photographie ou dessin avec/sans contour) ;
- la structure visuelle, pour les scènes de la classe 1 seulement (i.e., affichage non structuré d'objets graphiques *versus* affichage structuré, comme les arbres, les couronnes, les matrices, les structures géographiques), les scènes réalistes de la classe 2 étant toutes structurées par nature.

On considérera pour les cibles :

- la position à l'écran (centre, haut, bas,...) ;
- la saillance visuelle ;
- le caractère familier[44] de la cible ;
- la singularité visuelle *versus* l'ambiguïté (dans le cas d'une ambiguïté, les confusions possibles).

On considérera enfin, pour les messages sonores :

- le type d'indication spatiale (absolue, relative, ...) ;
- l'ambiguïté s'il y a lieu (liée soit à la désignation de l'objet, soit à la désignation spatiale).

4.5.2 Filtrage des données

Nous avons dans un premier temps regroupé les données recueillies selon le type de présentation des cibles, soit par situation PV, PO, PM. À chaque ensemble de scènes ainsi déterminé, nous avons ensuite appliqué le même filtre, relatif au nombre d'erreurs apparues par scène. L'opération du filtre consistait à ne conserver que les images ayant entraîné plus d'une erreur par scène, dans une situation donnée. Ce pré-traitement se fonde sur l'hypothèse suivante :

[43] cf. la caractérisation issue de la taxonomie de Bernsen [Bernsen, 1994], supra paragraphes 4.1.2 et 4.1.3 pages 28 et 31.

[44] Précise si la cible présente des caractéristiques insolites, voire incongrues, ou non.

> *Lorsqu'un couple (scène + situation) n'a provoqué qu'une seule erreur, l'origine de cette erreur est de la seule responsabilité du sujet qui l'a commise. Lorsqu'un tel couple a suscité plus d'une erreur, la cause de cette erreur peut être imputée aux caractéristiques de la scène, de la cible, ou du message sonore.*

Ainsi, dans chacune des trois situations PV, PO, PM, pour chacune des 34 scènes traitées, nous obtenons six données, à savoir les résultats des six sujets qui ont traité une image donnée dans une modalité donnée. Nous écartons une scène de l'analyse si elle a entraîné moins de deux échecs. Après avoir appliqué le filtre, il reste les images :

- 7(1), 9(1), 10(1), 11(1), 3(2), 5(2), 7(2) et 18(2), soit 8 images ayant provoqué 26 erreurs, dans la situation PV ;
- 13(1), 14(1), 2(2), 5(2) et 12(2), soit 5 images ayant provoqué 12 erreurs, dans la situation PO ;
- 3(2) et 5(2), soit 2 images ayant provoqué 4 erreurs, pour la situation PM.

4.5.3 Présentation visuelle (situation PV)

Pour déterminer les facteurs plausibles à l'origine des échecs observés dans la situation PV, nous avons classé dans l'ordre décroissant le pourcentage d'échecs qu'une caractéristique visuelle de l'image ou de la cible pouvait expliquer. Les pourcentages, présentés dans la suite, représentent le nombre d'erreurs qu'une caractéristique visuelle pouvait expliquer[45] sur le nombre total des erreurs filtrées dans la situation PV[46].

Comme facteurs plausibles à l'origine des échecs dans la situation PV, nous retenons les images denses (69% des échecs), non structurées (46% des échecs), ou représentant des formes géométriques (42% des échecs).

Les caractéristiques des cibles, supposées être à l'origine des échecs dans la situation PV, sont les suivantes :

- le manque de saillance de la cible dans la scène (85% des échecs) ;
- la position excentrée de la cible dans la scène (69% des échecs) ;
- l'ambiguïté possible entre la cible et d'autres objets graphiques (69% des échecs) ;
- le caractère non familier de la cible (42% des échecs).

Cette caractérisation des erreurs des sujets dans la situation PV sert de référence pour l'analyse des erreurs dans les situations PO et PM, présentée dans le paragraphe suivant.

4.5.4 Présentations orales et multimodales (situations PO et PM)

Le tableau 4.9 page 62 présente les erreurs regroupées par image, après filtrage des données. L'analyse des erreurs s'appuie sur la répartition des erreurs entre les trois conditions de présentation des cibles. On parle notamment de "correction d'erreurs" ou d'erreurs "corrigées" par les

[45] À elle seule, ou associée à d'autres facteurs.

[46] Soit 26 erreurs.

messages sonores lorsque le nombre d'erreurs dans les situations PO et PM est inférieur à celui observé pour la situation PV.

La correction d'erreurs par les messages sonores

Cinq scènes en situation PO et seulement deux en situation PM ont occasionné plus d'une erreur. On en dénombrait huit en situation PV. De plus, parmi les 26 erreurs filtrées de la situation PV, 24 ont été "corrigées" dans la situation PO, c'est-à-dire que les sujets n'ont pas échoué sur ces images dans la situation PO. Ainsi, sur les huit scènes ayant entraîné des erreurs dans la situation PV, six n'occasionnent aucune erreur en situation PO. De la même façon, parmi les 26 erreurs filtrées de la situation PV, 22 ont été "corrigées" dans la situation PM. Ainsi, sur les huit scènes ayant entraîné des erreurs dans la situation PV, six n'occasionnent aucune erreur en situation PM. Ces comparaisons mettent en évidence l'apport des messages oraux pour améliorer la précision de la sélection des cibles dans les tâches d'exploration visuelle.

Analyse des erreurs dans la situation PO

Quatre scènes sans aucune erreur dans les situations PV et PM, à savoir les scènes 13(1), 14(1), 2(2) et 12(2), ont occasionné dix erreurs, sur les 12 erreurs filtrées, dans la situation PO. Le nombre d'erreurs pour ces quatre scènes est, respectivement, 2, 4, 2, 2 (cf. infra tableau 4.9 page 62). Donc, il est probable que le facteur à l'origine de ces erreurs est la qualité discutable de l'information contenue dans les messages sonores correspondants. L'analyse détaillée de ces quatre messages sonores, associée aux informations fournies par les questionnaires et les debriefings post-session, appuie cette conclusion.

En effet, les sujets ont commis quatre erreurs sur l'image 14(1), dont le message associé est de type ISI. Dans cette image, 12 bouteilles de vin français sont positionnées sur une carte de la France en fonction de leur région d'appellation[47]. Le message associé est le suivant : "Le Côtes de Provence". Il faisait référence à une information non familière pour ces quatre sujets. Le message de type ISA+ISI associé à la scène 13(1) était trop complexe par sa structure et sa longueur. Il s'agissait du message "À droite, le Roi Soleil fils d'Anne d'Autriche", associé à un arbre généalogique de familles royales françaises (les Valois et les Bourbons). Les rois et reines apparaissaient sous forme de médaillons. Les sujets ont été déroutés par le lien de parenté mentionné dans ce message, notion historique avec laquelle ils n'étaient pas familiers. Quant aux deux autres paires d'erreurs commises sur les images 2(2) et 12(2), elles sont probablement dues aux désignations verbales des cibles. Le message utilisé pour désigner la cible de l'image 2(2) contenait les mots "premier plan". Bien que la notion de premier plan soit familière, elle peut être considérée comme ambiguë en raison de son caractère subjectif. Celui utilisé pour désigner la cible de l'image 12(2) contenait le mot "mansarde", que les sujets ont jugé lors des debriefings comme non familier.

[47] cf. infra figure 4.8 page 62.

Analyse des erreurs en nombre d'erreurs par situation			
Scène	Situation PV	Situation PO	Situation PM
7(1)	3	-	-
9(1)	2	-	-
10(1)	4	-	-
11(1)	2	-	-
13(1)	-	**2**	-
14(1)	-	**4**	-
2(2)	-	**2**	-
3(2)	3	-	2
5(2)	6	2	2
7(2)	2	-	-
12(2)	-	**2**	-
18(2)	4	-	-
Total par situation	26	12	4

TAB. 4.9 – Analyse des erreurs par image après filtrage des données.
Après filtrage des données (cf. supra 4.5.2 page 59), nous avons regroupé les erreurs par image et par condition de présentation des cibles. En gras figurent les images qui n'ont entraîné des erreurs que dans la situation PO ainsi que le nombre d'erreurs associé.

FIG. 4.8 – Image 14(1).
Quatre erreurs ont été commises par les sujets sur cette image dans la situation PO, alors qu'on n'en compte aucune dans les situations PV et PM. Le message associé était : "Le Côtes de Provence".

Analyse des erreurs dans la situation PM

Les messages multimodaux n'ont entraîné aucune erreur supplémentaire par rapport à la situation de référence PV, contrairement à la situation PO (cf. l'analyse des erreurs dans la situation PO). De plus, les quatre erreurs observées dans la situation PM se sont produites sur deux images seulement (2 erreurs sur l'image 3(2) et 2 sur l'image 5(2)). Ces observations montrent les avantages de combiner les informations visuelles et orales durant la présentation des cibles.

L'image 3(2) est un dessin de bâtiment pour lequel la cible associée est une fenêtre de ce bâtiment. Le message associé est le suivant' : "Au premier étage, la fenêtre la plus à gauche" (cf. figure 4.9 page 64). Les sujets ont commis trois erreurs sur cette image dans la situation de référence, la situation PV. En revanche, il n'en n'ont commis aucune dans la situation PO. Le message sonore associé à cette image semble donc efficace.

Par ailleurs, les deux sujets ayant commis une erreur sur cette image dans la situation PM ont cliqué "au premier étage", sur la fenêtre opposée : il s'agit de la fenêtre la plus à droite. On peut alors supposer qu'ils ont confondu leur gauche et leur droite sur cette image. Ou alors, ces deux erreurs, de même que les trois erreurs observées dans la situation PV, pourraient provenir des caractéristiques de la cible. En effet, nous l'avons jugée non saillante par la couleur, par sa forme et par sa taille. De plus, nous avons recensé entre 12 et 14 confusions possibles avec la cible.

De la même façon, nous avons porté une attention particulière à l'analyse des erreurs sur l'image 5(2) en raison :

- du nombre élevé d'erreurs observées sur cette image, 10 au total ; à noter qu'il s'agit de l'image ayant entraîné le plus d'erreurs sur toute la passation ;
- de la présence d'erreurs dans chaque condition expérimentale ; à noter qu'il s'agit de la seule image ayant entraîné des erreurs, à la fois, dans les situations PV, PO et PM.

L'image 5(2) est une photographie du Sacré Cœur. Le message associé est : "Le clocheton du petit dôme"(cf. figure 4.10 page 64). Nous avons jugé cette scène comme "difficile" pour les raisons suivantes :

- dans la situation PV, les six sujets se sont trompé sur cette image ;
- la cible est non saillante ;
- le nombre de confusions possibles avec d'autres objets de la scène est important : nous en avons recensé six au total ;
- la scène est dense en nombre d'éléments.

Sur cette même image, deux erreurs se sont produites dans la situation PO, en raison du vocabulaire technique et non familier utilisé pour désigner la cible.

Conclusions

Ces résultats montrent la complexité de l'interprétation des processus de traitement impliqués dans la perception multimodale. Le traitement de l'information multimodale semble guidé ou

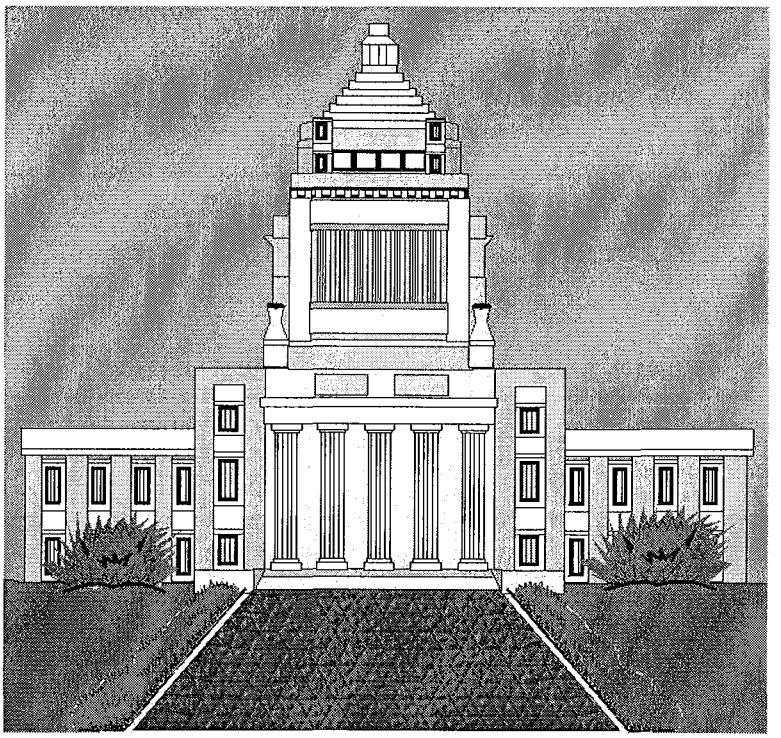


FIG. 4.9 – Image 3(2).
Dessin de bâtiment. Le message associé était : "Au premier étage, la fenêtre la plus à gauche".

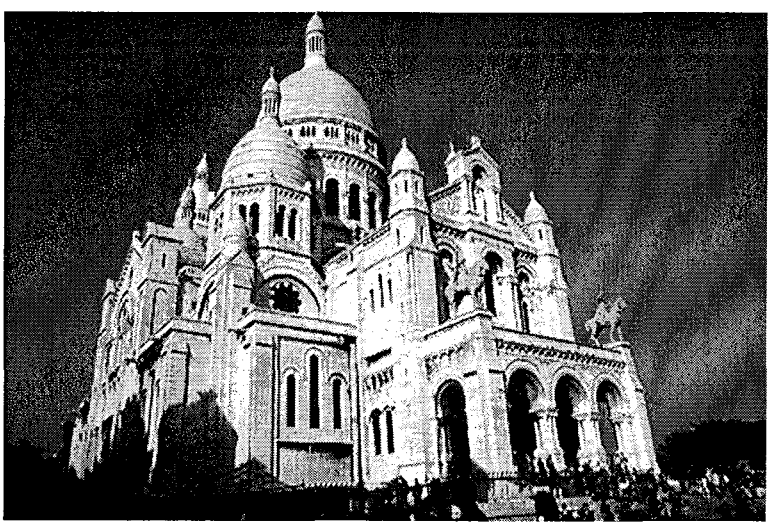


FIG. 4.10 – Image 5(2).
Photographie du Sacré Cœur. Le message associé était : "Le clocheton du petit dôme".

contrôlé davantage par les stratégies de perception visuelle que par les processus cognitifs induits par les messages sonores.

En résumé, l'analyse qualitative des erreurs a permis de confirmer l'apport des messages sonores pour améliorer le repérage visuel de cibles, pourvu que :

- les messages sonores soient courts, leur structure syntaxique simple et que le vocabulaire utilisé soit familier aux utilisateurs ;
- le contenu de l'information soit approprié et non ambigu.

4.5.5 Dépouillement des questionnaires utilisateurs

À l'issue de l'expérience, les sujets ont rempli un questionnaire portant sur la difficulté des différentes tâches de repérage dans les trois situations, l'efficacité des trois présentations en terme de rapidité du repérage de la cible, l'aide éventuelle apportée par les messages oraux dans les situations PO et PM, leur préférence entre les conditions PV, PO et PM. La difficulté du repérage a été évaluée de la façon suivante : les sujets ont attribué un niveau de difficulté allant de 1 à 4 à chaque condition. Le récapitulatif, sous forme de pourcentages, est présenté dans le tableau 4.10.

Présentation de la cible	"Très facile"	"Facile"	"Difficile"	"Très difficile"
Situation PV	22%	28%	39%	11%
Situation PO	22%	61%	17%	0%
Situation PM	72%	17%	11%	0%

TAB. 4.10 – Difficulté du repérage par type de présentation des cibles.

Plus des deux tiers (soit 72% des sujets) ont jugé la tâche de repérage très facile dans la situation PM, en lui attribuant un niveau de difficulté égal à 1. Aucun sujet n'a attribué un niveau de difficulté égal à 4 au repérage dans cette situation, c'est le cas, également dans la situation PO. En revanche, les tâches de repérage dans la situation PO sont jugées moins faciles que dans la situation PM. En effet, en comparaison des 72% de sujets ayant évalué à 1 le niveau de difficulté dans la situation PM, seulement 22% des sujets ont attribué le niveau 1 de difficulté dans la situation PO. Enfin, pour la situation PV, on observe une répartition plus équilibrée des sujets entre les différents niveaux. En effet, la moitié des sujets considère les tâches de repérage dans la situation PV comme faciles, avec 22% qui les classent de niveau 1 et 28% de niveau 2. L'autre moitié des sujets considère les tâches de repérage comme difficiles, avec 39% qui les classent de niveau 3 et 11% de niveau 4.

L'efficacité de l'interaction multimodale a été évaluée essentiellement en termes de rapidité et de facilité/difficulté du repérage. Nous avons demandé aux sujets de classer les trois situations PV, PO et PM par ordre décroissant d'efficacité. Seulement quatre sujets sur les dix-huit n'ont pas classé la situation PM comme étant la plus efficace. Deux d'entre eux ont ordonné les situations dans l'ordre PO-PM-PV, les deux autres dans l'ordre PV-PO-PM. Sur les quatorze autres sujets, dix ont choisi l'ordre PM-PO-PV et quatre ont choisi l'ordre PM-PV-PO.

Un seul sujet a déclaré ne pas avoir été aidé par les messages sonores "lorsqu'ils n'indiquent pas la position de la cible dans la scène", faisant sans doute référence aux messages de type ISI. En outre, trois sujets ont déclaré avoir été gênés par les messages sonores :

- le premier a été gêné par le vocabulaire du message de l'image 12(2) qui utilise le mot "mansarde" et par le caractère implicite de l'indication orale associée à l'image 14(1) "Le Côtes de Provence" ;
- le deuxième a été gêné par les messages trop longs, en particulier ceux de type ISR ;
- le troisième a été gêné par les messages sonores dans la situation PM, et a déclaré se servir des indications orales uniquement dans le cas où l'information visuelle ne suffisait pas.

Enfin, 66% des sujets préfèrent la situation PM, 17% la situation PO et 17% la situation PV.

Ces résultats mettent en évidence l'apport de la parole pour les tâches d'exploration visuelle comme le repérage de cibles. Ils appuient la conclusion selon laquelle les messages multimodaux facilitent le repérage de cibles dans les affichages denses et complexes dans la mesure où ils facilitent les tâches de repérage aux sujets et améliorent l'efficacité de l'interaction.

4.6 Conclusions générales

Cette étude préliminaire et exploratoire visait à mettre en évidence la contribution de la parole en tant que modalité d'expression du système complémentaire du graphique, donc dans un contexte d'interaction Homme-Machine multimodale. Plus précisément, il s'agissait de déterminer expérimentalement l'influence d'indications orales, de nature spatiale notamment, pour les tâches d'exploration visuelle, et en particulier, pour le repérage visuel de cibles dans un affichage dense et complexe. Nous avons mené une étude comparative entre trois types de présentation des cibles, à savoir les présentations visuelle, orale et multimodale. Les résultats des analyses quantitatives et qualitatives effectuées suggèrent que des messages sonores appropriés peuvent améliorer, à la fois, les temps et la précision de la sélection d'objets graphiques.

En particulier, les présentations multimodales[48], composées de la présentation visuelle isolée de la cible et d'une indication spatiale absolue de sa localisation dans la scène, sont les plus efficaces. En effet, les comparaisons entre les type de présentation des cibles ont montré que la présentation multimodale améliore les performances des utilisateurs, en terme de temps de recherche et de sélection des cibles, par rapport aux présentations exclusivement visuelles des cibles. Les présentations multimodales des cibles ont obtenu également le plus haut taux de satisfaction des sujets participant à l'expérience.

La préférence marquée des sujets pour la multimodalité parole + graphique est un résultat d'importance capitale, étant donné le nombre d'applications potentielles de cette forme d'assistance à la navigation ou au repérage dans des espaces d'informations graphiques. On peut citer l'exploration de visualisations interactives de grands ensembles d'informations comme les banques d'images, ou la recherche visuelle d'informations dans les affichages complexes comme le Web. De plus, l'assistance de la parole à la navigation au sein d'affichages réduits comme les P.D.A.

[48] À noter que les présentations orales des cibles peuvent suffire dans beaucoup d'applications où l'utilisateur a déjà vu la cible.

est une application potentielle de la multimodalité parole + graphique en sortie du système. La réalité virtuelle ou augmentée, ainsi que les environnements 3D sont également concernés. En effet, les messages multimodaux d'indication spatiale, associant un message sonore et un affichage graphique, semblent être une forme d'assistance appropriée et efficace à la recherche visuelle d'informations (icones, images, fichiers, etc.) au sein d'affichages graphiques complexes :

- appropriés, car non seulement ils ne gênent pas les sujets dans leur tâche de repérage, mais en plus, ils recueillent le plus grand taux de satisfaction des utilisateurs entre les autres formes d'assistance testées, à savoir les présentations monomodales visuelles (considérées comme la situation d'interaction de référence) et les présentations exclusivement orales ;
- efficaces, car ils améliorent et facilitent les performances des utilisateurs dans leur tâche de repérage visuel.

Par ailleurs, la recherche visuelle d'informations fait partie de nombreuses activités visuelles, outre la localisation d'objets graphiques et la navigation par *Go To* (cf. [Byrne *et al.*, 1999]) au sein des interfaces utilisateur. On peut citer :

- la manipulation directe des GUIs[49] pour réaliser des opérations sur des données, des ensembles de données ou des représentations graphiques (créer, modifier, dupliquer, supprimer, imprimer, etc.) [Chuah et Roth, 1996] ;
- la discrimination [Drury, 1992] ;
- le classement, le tri, etc.

Par conséquent, faciliter les tâches de recherche visuelle pourrait améliorer l'efficacité et l'utilisabilité des interfaces utilisateur en général au sens ergonomique du terme. Cependant, bien que prometteurs, ces résultats restent préliminaires en raison :

- du nombre limité de sujets participant à l'expérience (18) ;
- du nombre limité de scènes à traiter (36) ;
- de la grossièreté des mesures, basées sur la sélection, réussie ou non, à la souris des cibles ;
- de la caractérisation insuffisante des images présentées aux sujets.

De plus, les analyses qualitatives basées sur les erreurs de sujets, suggèrent l'influence possible, sur leurs performances, de facteurs dont nous n'avons pas tenu compte dans la conception du protocole expérimental. Concernant les images présentées aux sujets, il s'agit notamment de leur complexité en termes du nombre d'éléments affichés simultanément à l'écran et de leur structure visuelle pour la classe d'images abstraites seulement, qui contient les collections, structurées ou non, d'objets symboliques ou arbitraires comme les cartes, les formes géométriques ou encore les collections de panneaux de circulation. Concernant les cibles, il s'agit notamment de leur position à l'écran (e.g. centrée *versus* excentrée) et de leur saillance visuelle au sein de l'affichage. Plus particulièrement, pour les cibles associées aux images abstraites, nous n'avons pas tenu compte de l'homogénéité *versus* l'hétérogénéité des collections d'objets, et du nombre de confusions possibles entre la cible et les non-cibles dans le cas où la collection est homogène. Pourtant, ces paramètres sont déterminants si l'on veut juger de la complexité des couples (scène + cible) [Chelazzi, 1999].

Ces observations constituent les directions des deux études expérimentales suivantes (cf. infra les chapitres 5 et 6). Elles visent à mettre en évidence l'influence possible, sur la précision et les temps de sélection d'objets graphiques, des caractéristiques visuelles des scènes et des cibles.

[49] La manipulation directe est le paradigme de conception d'interfaces utilisateur le plus répandu actuellement.

Par exemple, nous supposons que des critères graphiques, tels que la structure de la scène ou la position de la cible dans la scène, sont susceptibles de faciliter le repérage visuel d'une cible dans une scène dense et complexe. En particulier, les structures géographiques semblent avoir permis de faciliter l'exploration visuelle des images abstraites [Cribbin et Chen, 2001a]. En revanche, l'efficacité des informations spatiales relatives semble être influencée par la proximité ou non de la cible par rapport à l'objet de référence [Gramopadhye et Madhani, 2001].

Ces expériences sont conçues selon la même approche que celle présentée ici, et implémentées en utilisant un protocole expérimental similaire à celui-ci. Cependant, ce protocole sera conçu pour aborder des questions spécifiques, afin de raffiner et enrichir les résultats de l'étude préliminaire. En particulier, nous augmentons significativement les nombre de sujets par condition de présentation des cibles ainsi que la quantité de tâches de repérage à effectuer, afin d'être en mesure de procéder à une analyse statistique (tests t) des données recueillies. Nous avons pour objectif d'analyser les stratégies de recherche visuelle adoptées par les sujets au sein d'organisations spatiales 2D comme les matrices ou les structures spatiales circulaires. Nous supposons qu'une meilleure compréhension des stratégies de recherche visuelle s'avérerait utile pour améliorer la conception des messages oraux [Cribbin et Chen, 2001b].

Chapitre 5

Deuxième étude

Cette deuxième étude, comme l'étude préliminaire, est conçue dans le cadre d'une approche expérimentale. Elle est centrée sur l'étude de l'influence de la structure ou organisation spatiale 2D des scènes graphiques sur l'efficacité de leur exploration, avec ou sans l'assistance d'indications orales à caractère spatial. Quatre structures 2D sont testées : la structure aléatoire, la structure matricielle, la structure en ellipse et la structure radiale, que nous allons définir. Comme dans l'expérience préliminaire, les performances des sujets sont évaluées en termes de temps et de précision de sélection des cibles.

Cette deuxième étude porte donc sur l'influence des ces quatre structures spatiales 2D sur les performances des sujets pour la même tâche de repérage que la première, avec ou sans l'assistance d'indications orales à caractère spatial. Elle a pour objectif, non seulement, de déterminer l'influence des organisations spatiales testées sur l'efficacité des messages oraux d'assistance au repérage de cibles, mais aussi, de déterminer si une structure émerge comme étant plus efficace que les autres pour améliorer et/ou faciliter les tâches de repérage de cibles.

Ce chapitre suit le même plan que celui adopté dans le chapitre précédent. Dans un premier temps, nous présentons la méthodologie utilisée dans la conception du plan expérimental ainsi que le protocole. Puis, après la description de la base d'images réalisée lors de la conception du matériel visuel utilisé lors de l'expérimentation, nous présentons les développements logiciels nécessaires à la mise en œuvre du protocole. Enfin, après la passation, nous présentons l'analyse des données recueillies lors de l'expérimentation, puis, nous commentons les résultats.

5.1 Méthodologie

Les principales conclusions de l'analyse quantitative des données recueillies au cours de l'étude préliminaire sur l'assistance de messages sonores à caractère spatial pour la détection de cibles sont les suivantes :

- une présentation uniquement visuelle de la cible conduit à des temps de réponse courts mais accompagnés d'un taux élevé d'erreurs ;

- une présentation exclusivement orale de la cible augmente les temps de réponse mais diminue les taux d'erreurs par rapport aux présentations visuelles de cibles ; ces deux résultats sont statistiquement significatifs ;
- une présentation multimodale conduit à des temps comparables à ceux obtenus pour les présentations visuelles, mais réduit de façon très significative les taux d'erreurs par rapport aux présentations visuelles.

Compte tenu de ces résultats, la première question de recherche de cette deuxième étude ergonomique sur la multimodalité parole + graphique en tant qu'assistance aux tâches visuelles, notamment la détection de cibles, est d'asseoir les conclusions de l'étude préliminaire. En effet, les premiers résultats peuvent être discutés en raison des limites du protocole expérimental de cette première étude : nombre limité de sujets (18) associé à un nombre limité de scènes à traiter (36). Le premier objectif de cette seconde étude peut donc se résumer comme suit : concevoir et réaliser une étude expérimentale qui validerait ces premiers résultats[50].

L'analyse qualitative des erreurs observées au cours de l'étude préliminaire sur l'assistance de messages sonores à caractère spatial pour la détection de cibles montre que :

- les erreurs croissent avec la complexité visuelle de la scène (densité en nombre d'éléments), le manque de saillance de la cible dans la scène, les confusions possibles avec les non cibles ;
- la structure visuelle des scènes a une influence sur le nombre d'erreurs observées puisque 46% des échecs sont dus à l'absence de structure visuelle.

Compte tenu de ces résultats qualitatifs, nous avons choisi de centrer nos travaux sur l'étude de l'influence de la structure visuelle des scènes sur l'efficacité de la recherche d'informations. En particulier, le deuxième objectif de cette étude est de mettre en évidence l'influence de l'organisation spatiale des scènes sur le repérage visuel de cibles, avec ou sans l'assistance de messages sonores d'indications spatiales.

Ce problème présente un double intérêt :

- sur le plan scientifique, d'une part, dans la mesure où il n'a pas encore été étudié de façon systématique, surtout dans des situations d'interaction Homme-Machine réalistes ;
- sur le plan des applications potentielles, d'autre part, en raison du développement des techniques de visualisation de grands ensembles d'informations qui mettent en œuvre des organisations spatiales variées.

Cette étude de l'influence de la structure spatiale des affichages sur l'activité de repérage des cibles est conçue également dans le cadre d'une approche expérimentale. Dans la suite de cette section, nous décrivons d'abord les modifications que nous avons apportées à la méthodologie adoptée lors de l'étude préliminaire[51]. Nous décrivons ensuite les hypothèses de travail, puis la caractérisation du matériel visuel et des messages sonores.

[50] À noter que les premiers résultats de cette étude préliminaire ont étés acceptés et présentés à un Workshop international [Carbonell et Kieffer, 2002]. Les résultats complets ont fait l'objet d'un chapitre de 25 pages dans un ouvrage scientifique collectif à paraître [Carbonell et Kieffer, To appear] et d'une communication acceptée à IHM'03 [Kieffer et Carbonell, 2003].

[51] cf. supra chapitre 4 section 4.1 page 27.

5.1.1 Présentation générale

Le nombre de conditions expérimentales a été réduit à deux. Nous conservons les modes de présentation des cibles suivants :

- présentation visuelle de la cible isolée au centre de l'écran (PV) ;
- présentation multimodale de la cible (PM).

Il est inutile de maintenir la condition orale, i.e., la présentation exclusivement orale de la cible, puisque les résultats obtenus dans cette condition sont voisins de ceux obtenus dans la condition multimodale pour la précision, et très inférieurs à ceux obtenus dans les deux autres conditions pour la rapidité (cf. supra chapitre 4 tableau 4.6 page 56.).

Comme dans l'étude préliminaire, la tâche est la détection de cibles connues visuellement *a priori* et le repérage est effectué pour chacun des deux modes de présentation, définissant ainsi deux situations expérimentales distinctes, à savoir les situations PV et PM.

Afin d'évaluer l'influence de la structure spatiale des scènes sur le repérage visuel de cibles, avec ou sans l'assistance de messages sonores d'indications spatiales, les scènes présentées aux sujets peuvent prendre les structures suivantes :

- la structure matricielle, où les éléments de la scène sont disposés en matrice ;
- la structure radiale, où les éléments de la scène sont disposés en étoile, i.e., sur huit branches à partir du centre vers les bords de la scène ;
- la structure elliptique, où les éléments de la scène sont disposés sur deux ellipses concentriques ;
- la structure aléatoire, où les éléments de la scène sont disposés aléatoirement.

Il convient de noter le caractère symétrique des structures matricielle, radiale, ou encore elliptique, par opposition à la structure aléatoire, qui elle représente dans le cadre de notre étude un cas de structure asymétrique. Les structures asymétriques nécessitent une étude particulière assez longue. On peut citer, à titre d'exemple, les structures arborescentes qui sont les structures asymétriques les plus utilisées, et ont fait l'objet de nombreux travaux en visualisation de masses de données ; voir, notamment, les cartes arborescentes [Plaisant *et al.*, 2002].

Le choix de tester dans cette expérimentation des structures symétriques simples telles que les matrices, les structures radiales ou encore les ellipses se justifie de la manière suivante : ces trois formes de structure se rapprochent d'organisations souvent utilisées par les techniques de visualisation de grands ensembles d'informations. On peut considérer les structures radiale et elliptique comme des simplifications des arbres hyperboliques [Lamping *et al.*, 1995]. Nous avons inclus la structure elliptique à notre deuxième étude, également, en raison de la forme de guidage circulaire que cette structure est susceptible d'induire chez les sujets.

De la même façon, la structure matricielle est proche de l'organisation spatiale d'icones dans les systèmes d'exploitation les plus courants tels que Unix ou Windows. Les images 1, 2, 4, 5 (1) de l'étude préliminaire étaient organisées en matrice 4×5 d'objets symboliques. Il s'agissait de drapeaux pour les images 1 et 2 (1)[52] et de panneaux de signalisation routière pour les images 4 et 5 (1). Nous avons observé pour ces quatre images des temps de sélection moyens inférieurs à la moyenne des temps observés pour l'ensemble des images de la classe 1.

[52] cf. supra figure 4.2 page 42.

Enfin, la détection visuelle d'une cible dans un affichage non structuré peut se rapporter à la recherche d'un icone, sur le bureau d'un ordinateur par exemple (cf. infra figure 5.2 page 73). De plus, dans l'étude préliminaire, les images 10, 11, 12 (1) étaient constituées de formes géométriques disposées aléatoirement au sein de l'affichage (cf. l'image 11(1) infra figure 5.1 page 72). Les temps de sélection observés sur ces trois images sont respectivement 3,4 secondes, 6,3 secondes et 4,5 secondes, soit une moyenne de 4,7 secondes, tandis que le temps moyen observé pour l'ensemble des images de la classe 1 est de 3,4 secondes. Le nombre d'erreurs, observé sur ces trois images est 5 erreurs sur l'image 10(1), 3 erreurs sur l'image 11(1), aucune erreur sur l'image 12(1). Ces 8 erreurs représentent plus d'un tiers des erreurs qui ont été commises par les sujets sur les images de la classe 1.

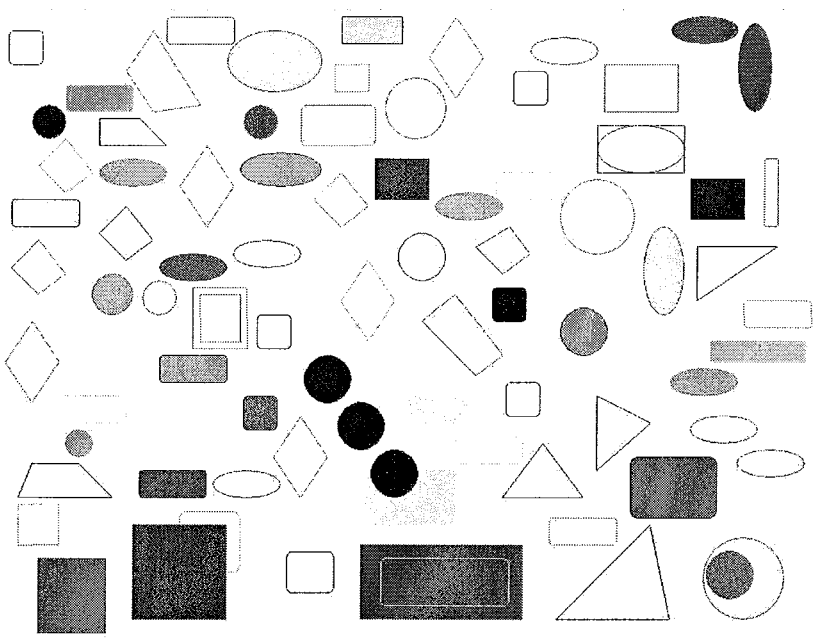


FIG. 5.1 – Image 11(1).
Collection de formes géométriques 2D non structurée, constituée d'environ 70 éléments.

5.1.2 Hypothèses de travail

L'expérience réalisée porte, comme dans l'étude préliminaire, sur la détection visuelle de cibles et est destinée à tester la validité des hypothèses suivantes :

- hypothèse A : les messages multimodaux, c'est-à-dire ceux qui associent une indication orale de localisation spatiale absolue à la présentation visuelle de la cible à l'écran, devraient améliorer de façon significative les performances des sujets par rapport à la présentation visuelle de la cible isolée ;
- hypothèse B : l'apport des messages sonores de localisation spatiale absolue au sein d'une présentation multimodale de la cible devrait varier en fonction de la structure visuelle de la scène ; en particulier, on devrait observer des différences de temps de réponse et de taux d'erreurs entre les différentes structures visuelles présentées aux sujets ;

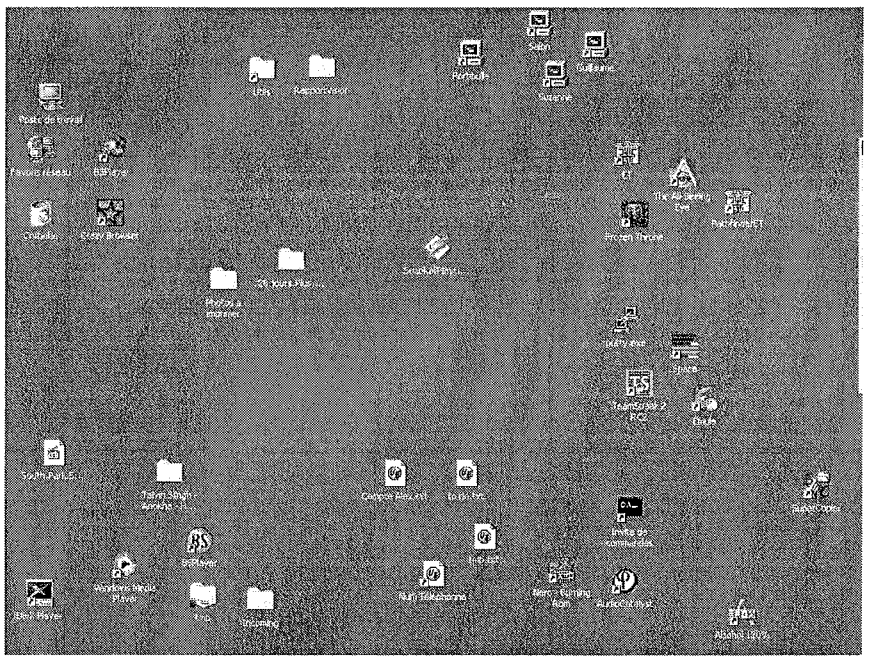


FIG. 5.2 – Exemple de bureau non structuré.
Caractéristiques : affichage 2D non structuré constitué de 37 éléments.

- hypothèse C : parmi les structures testées, il devrait exister une structure pour laquelle l'apport des messages oraux est plus sensible ;
- hypothèse D : la structure visuelle des scènes présentées aux sujets devrait avoir une influence sur leurs performances, même en l'absence d'une présentation multimodale de la cible ;
- hypothèse E : une des structures visuelles testées devrait émerger comme étant la meilleure en l'absence d'indication orale de localisation de la cible dans la scène.

5.1.3 Caractérisation du matériel visuel

La caractérisation du matériel visuel de l'expérience préliminaire s'appuyait sur la taxonomie des modalités graphiques en sortie de Bernsen [Bernsen, 1994] (cf. supra paragraphe 4.1.2 page 28). Nous avons éliminé les images symboliques et arbitraires du champ de cette étude, soit la classe 1 d'images, et nous avons retenu, dans la classe 2 des images réalistes, des photographies à l'exclusion de dessins. Nous avons fait ces choix en vue d'obtenir des résultats et conclusions directement utilisables pour la conception des futures interfaces de visualisation et de recherche d'informations multimédias, tout en gardant la complexité de cette étude dans des limites raisonnables.

Concernant l'élimination des dessins, précisons que les études expérimentales portant sur les fréquences spatiales, en particulier sur la notion de fréquence spatiale diagnostic, obtiennent des résultats différents selon la nature du matériel visuel, lettres de l'alphabet, photographies, etc. (voir dans [Giraudet, 2000], le tableau page 84).

Afin de rendre la tâche de repérage la plus réaliste possible, nous avons choisi de centrer notre deuxième expérimentation sur la recherche d'items. La tâche consiste donc à rechercher un item dans une collection d'items. Plus précisément, il s'agit pour les sujets de sélectionner à la souris, le plus rapidement possible, une photographie dans une scène constituée d'un ensemble ou d'une collection de photographies. Dans toute la suite du travail, nous utiliserons le terme "collection" pour désigner les photographies qui constituent une scène. L'objectif est de se rapprocher le plus possible de la recherche visuelle au sein d'une banque d'images. Ce travail s'inscrit en outre, grâce à cette tâche expérimentale, dans l'étude théorique qui sera réalisée dans le cadre du projet pluridisciplinaire Micromégas, "Approche multiéchelles pour la navigation dans les masses de données familières" [Micromégas, 2003][53].

De plus, il est apparu dans l'analyse qualitative des données recueillies lors de l'étude préliminaire que plusieurs caractéristiques des scènes et des cibles pouvaient être assimilées aux facteurs expliquant les taux d'erreurs observés sur certaines scènes. Il s'agissait notamment :

- de la complexité ou densité visuelle des scènes exprimée en terme du nombre d'éléments composant la scène ;
- de la qualité des affichages (contrastes, contours, etc.) ;
- de l'absence de structure des scènes (pour les images 10, 11, 12 (1) représentant des formes géométriques 2D organisées de façon aléatoire) ;
- de la position excentrée de la cible ;
- du manque de saillance visuelle de la cible au sein de l'affichage ;
- de l'ambiguïté entre la cible et d'autres éléments non cibles de la scène, etc.

Caractéristiques globales de la scène

Nous avons retenu, comme caractéristiques globales de la scène sa densité visuelle en terme de nombre de photographies qui la constituent ainsi que sa structure visuelle (absence de structure, ellipse, matrice ou structure radiale). Pour assurer la pertinence des comparaisons statistiques entre les quatre structures testées, nous avons choisi de fixer le nombre N de photographies contenues dans chaque scène, quelle que soit sa structure visuelle.

Caractéristiques des collections d'images

Pour étudier l'influence de la structure spatiale des affichages sur les performances de repérage visuel, il est nécessaire de contrôler avec précision les autres paramètres visuels qui caractérisent une scène, et tout particulièrement, les caractéristiques de la collection de photographies.

Chaque scène est donc composée de N photographies couleur, toutes au format 4/3, toutes de même taille, disposées spatialement selon l'une des structures retenues. De plus, pour que les photographies d'une même scène ne présentent pas entre elles des différences de saillance visuelle et de contenu sémiotique trop marquées, nous avons choisi qu'elles portent toutes sur le même thème.

[53] Soumis en avril 2003 à l'ACI "Masses de données" qui lui a accordé un soutien de trois ans.

Concernant le contenu des affichages, nous avons choisi de distinguer les photographies dites d'objets (objets complexes, personnages, ou groupes de personnages) des photographies de paysages ou scènes d'intérieur.

Les catégories créées au sein des deux types de collections d'images (i.e., objets *versus* paysages) sont au nombre de trois. On distingue :

- **les collections de photographies hétérogènes** : il s'agit de collections de photographies très différentes visuellement les unes des autres, soit par leurs couleurs, la nature des éléments représentés, la profondeur de champ (i.e., gros plan *versus* arrière-plan) ;
- **les collections de photographies homogènes** : il s'agit de collections de photographies proches visuellement les unes des autres par leurs couleurs, la nature des éléments représentés, la profondeur de champ ;
- **les collections de photographies homogènes complexes** : il s'agit des collections de photographies proches visuellement les unes des autres mais avec, en plus, une complexité de détail au sein de chaque photographie.

Dans son article [Chelazzi, 1999], Chelazzi fait une revue exhaustive des publications sur la recherche visuelle, tant du point de vue des neurosciences, que du point de vue de la psychologie cognitive. Ces travaux ont particulièrement retenu notre attention, car ils montrent que les temps nécessaires à l'analyse perceptuelle de chaque item d'une scène pourraient croître lorsque la complexité de chaque item augmente ou lorsque la cible est similaire aux items non-cibles (cf. les travaux de [Treisman et Gormican, 1988] et [Duncan et Humphreys, 1989]). Autrement dit, les items non-cibles très homogènes et facilement différentiables de la cible peuvent être rejetés plus rapidement que les items qui partagent une, voire plusieurs propriétés avec la cible.

Ces différentes considérations nous ont permis d'établir un niveau de difficulté de la tâche de repérage des cibles selon la nature de la collection de photographies.

En effet, d'après [Chelazzi, 1999], nous distinguons :

- les collections de photographies hétérogènes, qui devraient permettre aux sujets de repérer la cible facilement, grâce aux mécanismes parallèles pré-attentifs qui permettent de "rejeter en bloc" un, voire plusieurs items non-cibles de la scène. Ce type de collection définit le niveau de difficulté 1 dit "facile".
- les collections de photographies homogènes, qui devraient accroître les temps de recherche des cibles. Ce type de collection définit le niveau de difficulté 2 dit "moyen".
- et les collections de photographies homogènes complexes, qui devraient accroître davantage les temps de recherche des cibles. Ce type de collection définit le niveau de difficulté 3 dit "difficile".

En ce qui concerne le niveau de détail des photographies, nos hypothèses sont compatibles avec le modèle temporel "coarse to fine" de perception des fréquences spatiales [Huges *et al.*, 1996]. Concernant la qualité des affichages, nous avons choisi de ne sélectionner pour notre matériel visuel que des photographies de qualité visuelle similaire en termes de contraste, luminosité, définition. Toutes les photographies ont en outre été choisies sans contours ni bordures.

Caractéristiques des cibles

Nous avons retenu, comme caractéristiques de la cible, sa position au sein de l'affichage et sa saillance dans la scène. Nous avons systématiquement varié la position des cibles dans la scène, car nous ne disposons pas, à l'heure actuelle, d'hypothèses suffisamment pertinentes relatives à l'influence de la position de la cible dans la scène sur le repérage visuel. D'après [Cadet *et al.*, 2002] chapitre "Techniques", l'exploration d'une photographie suit les lignes de forces pour se terminer au bas de la photographie, ignorant quasiment le centre de l'image. Les lignes de force d'une photographie, comme le montre la figure 5.3 page 76, divisent la photographie en neuf zones : les zones haut-gauche, haut, haut-droit, gauche, centre, droit, bas-gauche, bas, bas-droit. Aussi, nous définissons la position des cibles en fonction de ces neuf zones.

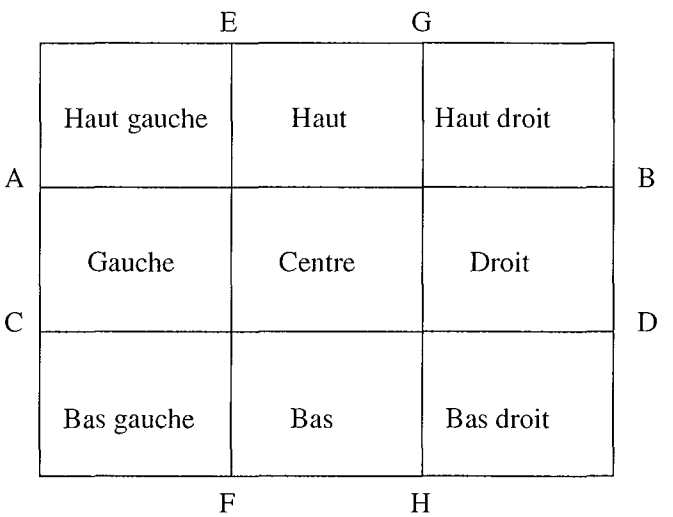


FIG. 5.3 – Découpage de l'écran pour fixer la position des cibles.
Les lignes de forces (AB), (CD), (EF), (GH) découpent l'image en tiers, définissant ainsi quatre points de force à leur intersection.

La saillance relative de la cible au sein de la collection à laquelle elle appartient doit, en outre, être contrôlée. Nous avons remarqué, lors de l'analyse de l'étude préliminaire, que la saillance d'une cible tient aux facteurs suivants : sa forme, sa taille, sa couleur, l'éventuelle ambiguïté qui existe pour les sujets entre la cible et les non-cibles de la collection, et enfin, le degré de familiarité de la cible pour les sujets. Toutes les photographies d'une collection, y compris la cible, ont même forme (format 4/3) et même taille. Concernant la couleur de la cible, nous avons élaboré trois niveaux de difficulté pour les collections de photographies tenant compte, à la fois, du degré d'homogénéité et de la complexité de détail des photographies d'une collection. Les photographies de chaque collection relèvent d'un même thème. Par conséquent, c'est essentiellement par la couleur que se distinguent les unes des autres les photographies composant chacune des collections hétérogènes (cf. niveau 1 de difficulté). Pour les collections homogènes, complexes ou non (cf. respectivement, niveaux 3 et 2 de difficulté), nous avons choisi de regrouper, dans une même collection, des photographies dont les couleurs dominantes sont voisines.

Pour résumer, la différence de couleur entre les éléments d'une collection décroît lorsque le niveau de difficulté attribué à la collection augmente. En d'autres termes, pour contrôler la saillance par la couleur de la cible au sein d'une collection nous appliquons l'algorithme suivant :

- si la collection est de niveau de difficulté 1, alors les éléments la constituant diffèrent par la couleur ; ce cas de figure implique de choisir une cible de couleur neutre, par opposition aux couleurs vives telles que le rouge, ou le bleu électrique (cf. chapitre 4 dans [Ware, 2004]) ;
- si la collection est de niveau de difficulté 2 ou 3, alors les éléments la constituant ne différent pas par la couleur ; ce cas de figure permet de sélectionner la cible, uniquement en fonction du nombre de confusions possibles avec d'autres éléments de la scène.

À noter que la saillance visuelle de la cible par la couleur est contrôlée globalement par les niveaux de difficulté définis pour caractériser les collections de photographies. Il en est de même pour le facteur d'ambiguïté éventuelle entre la cible et les non-cibles de la collection. En effet, plus le niveau de difficulté de la collection augmente, plus le nombre de confusions possibles entre la cible et les non-cibles croît, en raison du degré d'homogénéité de la collection.

Pour contrôler le degré de familiarité des sujets avec les cibles, nous avons choisi de ne faire porter les collections que sur des thèmes simples : par exemple, les animaux, les avions, les portraits ou encore les groupes de personnages en ce qui concerne les photographies de type "objets complexes". En ce qui concerne les photographies de type "paysages", les thèmes retenus sont, par exemple, les bâtiments, les montagnes, les couchers de soleil ou encore les scènes d'intérieur.

Comparaison avec un modèle d'attention visuelle basé sur la saillance

Dans leurs travaux [Itti *et al.*, 1998; Itti et Koch, 1999; Itti et Koch, 2000], Itti et Koch présentent un modèle de l'attention visuelle basé sur des processus visuels ascendants, ou saillance visuelle, utilisés par les singes dans la détection de cibles manifestes ou visuellement "frappantes" au sein d'environnements visuels denses. Leur modèle est construit de la manière suivante. Des caractéristiques multiéchelles des images sont combinées au sein d'une unique carte topographique de saillance. Un réseau dynamique de neurones sélectionne une suite de localisations classées par ordre de saillance décroissante. Les caractéristiques multiéchelles retenues sont la couleur, l'orientation, l'intensité lumineuse et le mouvement. Dans [Itti *et al.*, 1998; Itti et Koch, 1999], le modèle est appliqué à la détection de cibles visuelles lors de la conduite automobile. Dans [Itti et Koch, 2000], il est appliqué à une tâche artificielle de recherche visuelle courante en psychologie ainsi qu'à l'acquisition de cibles militaires.

Bien que très efficace dans les contextes de la conduite automobile, ou plus généralement des scènes naturelles en deux dimensions [Itti, 2000], le modèle d'Itti et Koch n'a pas été utilisé pour contrôler le degré de saillance des cibles retenues au sein des collections de photographies. Nous aurions pu envisager d'utiliser le modèle sur chaque scène présentée aux sujets et de choisir une cible de saillance neutre grâce au logiciel d'Itti *et al.* Nos scènes contenant toutes le même nombre de photographies N, il suffisait d'en choisir une apparaissant dans la suite de saillance autour de la position $\frac{N}{2}$.

Nous n'avons pas retenu cette solution en raison des développements logiciels importants qu'il aurait fallu réaliser pour parvenir à utiliser le modèle. D'une part, nous avons jugé le coût

de ces développements trop lourd par rapport aux bénéfices minimes qui en auraient découlé. En effet, la seule caractéristique des images utilisée aurait été la couleur car :

- l'orientation des éléments de la collection, qui se réduit au cas particulier à la forme des photographies constituant la collection, est uniforme[54] ;
- l'intensité lumineuse des éléments d'une collection est contrôlée[55] ;
- le mouvement est absent[56].

D'autre part, dans la mesure où les caractéristiques des images à tester se réduisent à la couleur, le modèle n'est pas adapté pour traiter les collections de difficulté 2 et 3, car les photographies contenues dans ces deux types de collections sont homogènes par rapport à la couleur. Le modèle appliqué à notre étude n'aurait permis de traiter que les collections de niveau de difficulté 1.

5.1.4 Caractérisation du matériel sonore

Afin de faciliter la détection de la cible dans la scène, les messages sonores utilisés au cours de l'expérience préliminaire contenaient la désignation de la cible visuelle et des indications spatiales sur sa localisation dans la scène (cf. supra paragraphe 4.1.3 page 31). La structure syntaxique des messages sonores était la suivante :

[indication spatiale] + nom de la cible (désignation)

Les indications spatiales étaient absolues (ISA), relatives (ISR), implicites (ISI), absolues et implicites (ISA+ISI), relatives et implicites (ISR+ISI).

Nous avons supprimé les messages de type ISR, ISI, ISA+ISR, ISA+ISI pour ne conserver que les messages de type ISA, en raison des résultats obtenus lors de l'étude préliminaire (cf. supra tableau 4.8 page 57). En effet, parmi les 5 types de messages utilisés, les messages contenant des indications spatiales absolues se sont révélés les messages sonores les plus efficaces en termes de temps et de précision de sélection des cibles, que ce soit lors d'une présentation orale (situation PO) ou lors d'une présentation multimodale (situation PM) de la cible.

Nous avons fait ce choix en vue d'obtenir des résultats et conclusions directement utilisables pour la conception des futures interfaces de visualisation et de recherche d'informations multimédias, tout en gardant la complexité de cette étude dans des limites raisonnables.

En outre, puisque nous avons limité le nombre des conditions expérimentales aux modes de présentation visuelle et multimodale des cibles, en supprimant de notre deuxième étude les présentations exclusivement orales, il n'est pas nécessaire de maintenir la désignation verbale de la cible au sein des messages sonores. En effet, l'affichage de la cible inclus à la fois dans la présentation visuelle et dans la présentation multimodale, permet aux utilisateurs d'identifier la cible, notamment sa nature (type d'objet ou type de paysage) ainsi que ses propriétés visuelles, comme par exemple, sa couleur ou sa forme. La structure syntaxique des messages sonores devient donc la suivante :

[54] Les photographies d'une collection sont toujours au format 4/3.
[55] Toutes les images appartenant à une collection sont d'intensité lumineuse similaire.
[56] Toutes les images appartenant à une collection sont statiques.

[indication spatiale absolue de la position de la cible dans la scène]

Les différents messages utilisés sont : 'en haut, à gauche", "en haut", "en haut, à droite", "à gauche", "au centre", "à droite", "en bas, à gauche", "en bas" et "en bas, à droite".

Les messages oraux ont été enregistrés au format WAV en utilisant le magnétophone de Windows. Ils ont été énoncés par une locutrice expérimentée s'exprimant avec une prosodie neutre.

5.2 Protocole expérimental

5.2.1 Généralités

Le protocole expérimental est semblable à celui adopté pour l'étude préliminaire. Le scénario d'interaction est le même (cf. supra paragraphe 4.2.2 page 34) avec :

- la présentation visuelle ou multimodale de la cible, i.e., la présentation d'une des N photographies d'une collection au cas particulier ;
- le repositionnement de la souris au centre de l'écran grâce à la sélection obligatoire du bouton OK ;
- l'affichage de la scène.

Comme dans l'étude préliminaire, le temps de présentation des cibles est de 3 secondes (cf. supra paragraphe 4.2.1 page 33). La taille de la cible ne varie pas entre sa présentation visuelle[57] et son affichage au sein de la scène. Une fois que la scène apparaît, les sujets ne disposent que d'un seul clic pour sélectionner la cible. Au clic, on passe à la présentation, visuelle ou multimodale, de la prochaine cible.

En outre, pour tester les hypothèses de travail de cette deuxième étude, nous avons comparé les performances des sujets en regroupant les données :

- par condition expérimentale pour tester la validité de l'hypothèse A ;
- par condition expérimentale, puis par structure, pour tester la validité des hypothèses B, C, D et E.

Dans la suite, nous présentons la terminologie utilisée concernant le matériel expérimental, puis les variables de l'expérience, et enfin les différences de protocole par rapport à l'étude préliminaire. Il s'agit notamment du travail qui a été réalisé afin d'éliminer les difficultés rencontrées lors de l'étude préliminaire en raison du nombre limité de sujets participant à l'expérimentation (18), du nombre limité de scènes présentées aux sujets (36) et de leur caractérisation visuelle insuffisante.

5.2.2 Terminologie

Chacune des scènes, ou images, présentées aux sujets contient N photographies appartenant à une collection de photographies. Nous désignerons par séquence l'ensemble des scènes présentées aux sujets lors de l'expérimentation. Nous parlerons de "séquence visuelle" pour désigner

[57] Chacune des deux conditions expérimentales contient l'affichage de la cible isolée au centre de l'écran

l'ensemble des scènes présentées aux sujets dans la situation PV, et de "séquence multimodale" pour désigner l'ensemble des scènes présentées aux sujets dans la situation PM.

5.2.3 Les variables de l'expérience

Parmi les caractéristiques du matériel visuel de la deuxième étude, nous distinguons les caractéristiques de la scène des caractéristiques de la cible associée à une scène. Les caractéristiques globales de la scène comprennent sa structure spatiale et sa densité en nombre d'éléments (N) qu'elle contient. Les caractéristiques de la collection de photographies composant la scène sont au nombre de trois. Il s'agit du format des photographies de la collection (orientation, taille), du type de contenu de la collection (objet ou paysage) et du niveau de difficulté attribué à la collection. Les caractéristiques de la cible associée à une scène, au nombre de deux, sont : la position de la cible dans la scène et sa saillance visuelle dans la scène. Nous présentons ces caractéristiques dans le tableau 5.1 ci-dessous.

Caractéristiques globales des scènes	**Structure spatiale**	**matricielle, radiale, elliptique, aléatoire**
	Densité	N éléments
Caractéristiques des collections de photographies	Format des photos	format 4/3, même taille
	Type de contenu	**objet, paysage**
	Niveau de difficulté	**1, 2, 3**
Caractéristiques des cibles	**Position**	**HG, H, HD, G, C, D, BG, B, BD**
	Saillance visuelle	forme, couleurs, complexité (détail), nombre de confusions possibles (ambiguïté), familiarité

TAB. 5.1 – Récapitulatif des caractéristiques du matériel visuel.
Les variables libres sont en gras. Les autres caractéristiques sont fixées, i.e., toutes les scènes contiennent N photographies (N fixé), les N photographies d'une scène ont même format (4/3) et même taille. La saillance visuelle de chaque cible est contrôlée en fonction du niveau de difficulté de la collection de photographies, à laquelle appartient la cible.

Les messages oraux de localisation de la cible dans la scène sont tous de type indication spatiale absolue (ISA). Nous considérons les deux conditions expérimentales suivantes : présentation visuelle de la cible (situation PV) et présentation multimodale de la cible (situation PM). Il en résulte les variables libres suivantes :

- le mode de présentation de la cible : visuel (PV) ou multimodal (PM) ;
- la structure de la scène affichée à l'écran : matricielle, radiale, elliptique, ou aléatoire ;
- le type de contenu des photographies : objet ou paysage ;
- le niveau de difficulté attribué aux collections de photographies : 1 (facile), 2 (moyen) ou 3 (difficile) ;
- la position de la cible dans la scène, à savoir, dans les zones haut-gauche, haut, haut-droit, gauche, centre, droit, bas-gauche, bas, bas-droit.

Les variables liées sont, comme dans l'étude préliminaire :

- le temps de sélection de la cible en millisecondes ;

– la précision de la sélection de la cible, i.e., dans la cible ou en-dehors.

Pour assurer la pertinence des traitements statistiques sur les données quantitatives recueillies, i.e. temps de sélection et précision de la sélection de la cibles, nous avons d'abord modifié le nombre de sujets participant à l'expérimentation par rapport à la première étude, le nombre de scènes par condition expérimentale, ainsi que la répartition du matériel visuel entre les groupes de sujets. Nous avons élaboré, en nous appuyant sur l'ouvrage collectif [Bouchard et Cyr, 2000], un protocole expérimental qui assure, à la fois, la validité interne et externe de la recherche. Les paragraphes 5.2.4 page 81 et 5.2.5 page 83 portent sur les facteurs d'invalidité et de biais susceptibles d'intervenir dans notre étude. Ils sont basés sur une étude approfondie du chapitre 2 de [Bouchard et Cyr, 2000].

La validité interne d'un protocole expérimental est le contrôle des variables nuisibles et correspond à la capacité d'une procédure de pouvoir dissocier clairement les effets dus aux variables d'intérêt des effets que pourraient générer des variables qui ne sont pas directement considérées dans la recherche. La validité externe détermine le degré de généralité des résultats d'une étude. Sa valeur incite ou non à conduire de nouvelles recherches de façon à prouver la robustesse des résultats obtenus.

5.2.4 La validité interne

Le contrôle des variables nuisibles peut signifier leur élimination complète de l'expérience. Cependant, il est impossible d'éliminer complètement l'influence de la majorité des variables qui peuvent affecter les résultats d'une expérience (l'intelligence, les expériences antérieures, ou encore la classe professionnelle des participants). Des techniques de contrôle au sein du plan expérimental, comme l'appariement, le contrebalancement, l'affectation aléatoire, l'automatisation de la collecte des données et un questionnaire post-expérimental, vont nous permettre de minimiser l'influence de ces variables.

Appariement des conditions expérimentales

Afin d'obtenir des mesures comparables entre les deux conditions expérimentales PV et PM, les sujets effectuent la tâche de repérage de cibles dans la condition PV et dans la condition PM. Il s'agit de la technique de l'appariement des conditions expérimentales qui fait en sorte qu'on retrouve dans chacune des conditions exactement le même niveau de variables nuisibles. L'appariement permet de contrôler les changements qui s'opèrent chez les individus, dont les effets sont par exemple, la fatigue, la faim, la motivation, ou encore les émotions.

Contrebalancement

La réactivité de la mesure est l'expérience que les participants ont acquise de l'instrument[58] de mesure dans le cadre d'épreuves répétées. En somme, les performances des sujets lors de la

[58] Au sens large ; on doit considérer comme instrument tout dispositif physique ou abstrait utilisé dans l'expérience.

passation de la deuxième condition sont affectées par différents facteurs tels que la pratique ou la familiarité avec l'instrument de mesure. Le fait que les participants soient soumis à plus d'une mesure peut entraîner deux effets différents sur le rendement postérieur à une situation initiale : un effet de sensibilisation ou un effet d'inoculation [Robert, 1988].

On parle d'effet de sensibilisation lorsque la première situation de mesure rend les participants plus réceptifs à la seconde situation. Des phénomènes comme la familiarisation, l'apprentissage ou une augmentation de l'intérêt ou de la motivation par rapport aux situations de mesure peuvent être reliés à ce genre d'effet. On parle d'effet d'inoculation lorsque les situations antérieures auxquelles a participé l'individu affectent à la baisse le rendement aux situations présentées ultérieurement. À ce moment, des facteurs comme l'ennui, la fatigue peuvent être responsables, en partie, du déclin du rendement des participants à ces mesures.

Le contrebalancement des conditions expérimentales permet le contrôle de la réactivité de la mesure dans les expériences où les sujets participent à plusieurs conditions ou à plusieurs mesures. Pour éviter l'effet de séquence, il s'agit de faire varier l'ordre de présentation des conditions. En d'autres termes, la moitié des sujets est affectée à l'ordre de passation PV-PM, l'autre moitié à l'ordre de passation PM-PV.

Affectation aléatoire des conditions

Nous avons ajouté à l'appariement et au contrebalancement, l'affectation aléatoire de l'ordre de présentation des deux conditions PV et PM. Cette distribution aléatoire permet l'obtention de conditions équivalentes pour les différentes variables de l'expérimentation[59].

Automatisation de la collecte des données

L'expérimentateur lui-même peut avoir une influence sur les résultats obtenus. On peut identifier deux ensembles de caractéristiques des expérimentateurs qui risquent de menacer la validité interne de la recherche : les attentes et les attributs de l'expérimentateur.

Les attentes de l'expérimentateur peuvent amener les participants à se comporter de façon à conforter les attentes de l'expérimentateur (cf. les études portant sur l'*effet Pygmalion* [Rosenthal et Jacobson, 1968; Rosenthal, 1976]). En outre, il existe toujours des sujets à la limite des définitions/catégorisations de la manipulation expérimentale. Il est à craindre que, dans ce cas, l'expérimentateur ait tendance à classer ces sujets dans le sens de son hypothèse de travail. Enfin, les attributs de l'expérimentateur, comme ses caractéristiques biologiques, sociales, individuelles ou encore, son anxiété, affectent le comportement des participants. C'est pour éviter les erreurs d'enregistrement, contrôler les attentes de l'expérimentateur, que nous avons automatisé la collecte des données.

[59] cf. infra 5.2.3 page 80.

Questionnaire post-expérimental

Les participants peuvent développer des attentes à l'égard de l'expérience. Ces activités cognitives peuvent interagir avec les procédures expérimentales et fausser les résultats. Les exigences implicites du protocole, comme la présentation de l'expérience, le rôle qu'on feint de lui donner (ici, évaluer une interface 2D interactive), le dispositif ou encore, les tâches à effectuer, définissent l'expérience selon le point de vue du participant. Cette perception peut fausser ses réponses, de même que la motivation de présenter une image positive de soi. Néanmoins, il existe de nombreuses techniques de contrôle parmi lesquelles nous avons choisi le questionnaire post-expérimental. Dans le questionnaire post-expérimental, les participants peuvent livrer leurs perceptions des objectifs de l'expérimentation, de ce qu'on attendait d'eux, du comportement qu'ils devaient adopter.

5.2.5 La validité externe

La validité externe est le degré de généralisation possible des résultats d'une étude. Il s'agit, dans notre cas, de pouvoir ou non appliquer les résultats obtenus aux activités de recherche visuelle en général. Assurer cette généralisation a consisté à assurer la validité échantillonnale de notre expérimentation, d'une part, et sa validité écologique, d'autre part.

Validité échantillonnale

La validité échantillonnale consiste à être en mesure d'affirmer que l'échantillon, utilisé lors de l'expérimentation, est représentatif des individus auxquels on veut étendre les résultats. En théorie, pour assurer la validité échantillonnale d'une expérimentation, on utilise des échantillons probabilistes (i.e., larges et constitués au hasard parmi la population cible). En pratique, les échantillons ne sont que très rarement probabilistes. D'une part, la constitution de tels échantillons nécessite des ressources énormes et présente le risque d'inclure dans l'étude des individus pour lesquels l'expérience est inadaptée (hétérogénéité de l'échantillon). D'autre part, on ne peut forcer les individus à participer.

Assurer la validité échantillonnale de notre expérimentation a consisté à tenir compte à la fois du nombre de sujets participant à l'expérience, mais aussi du nombre de scènes qui leur étaient présentées. En effet, dans l'étude préliminaire, la difficulté à généraliser les résultats ne résidait pas tant dans le nombre de sujets (18) participant à l'expérience ou dans le nombre de scènes présentées (36)[60], mais plutôt dans le fait de ne pouvoir comparer les performances de chacun des sujets sur la même image dans chaque condition expérimentale, en raison du nombre limité d'images par condition expérimentale (12) et du nombre réduit de sujets par image et par condition.

Nous avons donc augmenté le nombre de sujets participant à l'expérience ainsi que le nombre de scènes dans chacune des deux conditions expérimentales, en tenant compte du nombre de

[60]Des travaux au protocole expérimental similaire (18 sujets, 3 conditions expérimentales, 7 tâches par condition) ont été validés par leur publication dans les actes de la conférence internationale CHI'92 [Ahlberg *et al.*, 1992].

structures testées (4). Pour étudier l'influence de la structure spatiale des affichages sur les performances de repérage, il est nécessaire de contrôler avec précision les autres paramètres visuels qui caractérisent une scène comme :

- le type de contenu des photographies (objet ou paysage) : le nombre de scènes par condition doit donc être un multiple de 2 ;
- le niveau de difficulté attribué aux collections de photographies (facile, moyen, difficile) : le nombre de scènes par condition doit donc être un multiple de 3 ;
- la densité N de la scène en nombre de pastilles par scène (nous avons fixé N à 30, définissant ainsi 30 positions possibles pour la cible au sein des zones haut-gauche, haut, haut-droit, gauche, centre, droit, bas-gauche, bas, bas-droit) : le nombre de scènes par condition doit donc être un multiple de 30.

Le nombre de sujets est désormais fixé à 24. Comme dans l'étude préliminaire, nous avons sélectionné un groupe de sujets homogène[61]. La plupart des sujets a été choisie au LORIA (21 étudiants, ingénieurs et chercheurs en informatique). L'un des sujets venait du LITA[62] (étudiant en informatique). Les deux autres sujets sont employés d'entreprises nancéiennes. Tous les sujets sélectionnés sont utilisateurs experts de la souris compte tenu de leur occupation et de la tranche d'âge choisie, entre 24 et 29 ans. En effet, dans leur étude [Dollinger et Hoyer, 1996], Dollinger et Hoyer ont comparé les performances, en termes de précision et de temps de réponse, de deux groupes de sujets. Le premier groupe, dont l'âge moyen était de 26,5 ans, s'est avéré plus précis et plus rapide que le deuxième groupe, dont l'âge moyen était de 45,7 ans. La tâche expérimentale proposée aux participants était l'inspection visuelle au sein de reproductions biologiques.

Nous avons fixé le nombre de scènes par condition à 120, obtenant ainsi 30 scènes par structure, les 30 positions possibles de la cible par structure pouvant être testées. Sur les 30 scènes d'une même structure, 15 sont des objets complexes, 15 sont des paysages. Parmi les 15 objets ou paysages d'une même structure, 5 sont de niveau de difficulté 1, 5 sont de niveau de difficulté 2 et 5 sont de niveau de difficulté 3. La répartition du matériel visuel entre les quatre structures, les deux types de photographies, les trois niveaux de difficulté des scènes est présentée dans le tableau 5.2 ci-dessous.

Type \ Structure	Matrice	Radiale	Ellipse	Aléatoire	Total
Objets complexes	15	15	15	15	60
Paysages	15	15	15	15	60
Total	30	30	30	30	**120**

TAB. 5.2 – Répartition du matériel visuel selon les différents critères caractérisant les scènes. Ce tableau à double entrée présente le nombre de scènes par structure et par type de photographie. Concernant les niveaux de difficulté attribués aux scènes, considérons par exemple, les 15 scènes matricielles de type paysage. Alors parmi ces 15 images, 5 sont de difficulté 1, 5 de difficulté 2 et 5 de difficulté 3.

[61] cf. supra 4.2.8 page 46.

[62] Laboratoire d'Informatique Théorique et Appliquée, Metz (FRANCE).

La réalisation d'un plan expérimental permettant de comparer en toute rigueur les performances des sujets vis-à-vis des quatre structures d'images et des deux modes de présentation de la cible aurait imposé de définir 30 couples scène et cible et de les présenter huit fois aux sujets[63], ce qui est la solution adoptée dans [Van Diepen *et al.*, 1999]. Mais cette solution rend l'activité de repérage artificielle et fastidieuse. Elle risque donc de créer chez les sujets des effets de lassitude et de démotivation en cours d'expérimentation. En outre, elle ne permet pas, même en augmentant considérablement le nombre de sujets, de contrôler les phénomènes de réactivité à la mesure, ici l'apprentissage de la tâche.

Nous avons donc choisi le compromis suivant, qui tient compte des objectifs principaux de cette étude en privilégiant les comparaisons entre les deux modes de présentation : les mêmes images associées aux mêmes cibles sont présentées dans les deux conditions, mais dans des ordres différents définis aléatoirement pour chaque sujet. L'ordre de passation est contrebalancé, formant ainsi deux groupes de 12 sujets. Un groupe traite les conditions expérimentales dans l'ordre PV-PM, l'autre dans l'ordre PM-PV, chaque condition comprenant 120 images en tout.

Par ailleurs, pour que l'ordre de présentation des scènes n'influence pas les résultats en formant des séries d'images plus faciles/difficiles que d'autres, nous avons fait le choix de définir aléatoirement l'ordre de présentation des scènes pour chaque sujet. Autrement dit, les sujets ne voient pas les images dans le même ordre, même s'ils effectuent les conditions expérimentales dans le même ordre. Enfin, les 120 images, illustrant les quatre structures sont toutes différentes (i.e., les photographies qui les composent sont toutes différentes).

Validité écologique

Cette notion fait référence à la possibilité de généraliser les résultats à d'autres situations que celle dans laquelle s'est faite la collecte des données. Il existe plusieurs menaces à la validité écologique : les caractéristiques des stimuli, les caractéristiques du contexte et les caractéristiques de la mesure.

Un stimulus est caractérisé par la nature du milieu (laboratoire *versus* terrain), l'expérimentateur et le matériel utilisé. Dans notre cas, la salle dans laquelle se déroulaient les passations, est considérée comme un environnement naturel pour notre expérience, en raison du profil des sujets, tous informaticiens. De plus, la tâche demandant aux participants de cliquer le plus rapidement possible sur la cible à l'écran est un bon indicateur de l'efficacité des présentations graphiques et orales des scènes dans un contexte de repérage visuel de cibles. Parmi les caractéristiques contextuelles qui peuvent affecter la validité écologique d'une expérience, on peut citer les aspects suivants : la réactivité au contexte expérimental, l'interférence d'interventions multiples et l'effet de nouveauté.

La réactivité au contexte expérimental fait référence au fait de prendre conscience de participer à une expérience. Les participants peuvent alors vouloir plaire à l'expérimentateur, faire attention à ne pas donner des réponses jugées négativement par l'expérimentateur, être plus assidus qu'à l'accoutumée. La solution est de leur donner un rôle fictif naturel (cf. la présentation de l'expérience). Les interventions multiples, telles que les différentes conditions d'une expérience

[63] 4 structures × 2 modes de présentation

ou les séquences, peuvent nuire à l'efficacité des conditions. La solution, comme pour vérifier la validité interne, est le contrebalancement. Enfin, l'effet de nouveauté peut également nuire à la validité externe. L'étude de Brownell (1966 ; cité dans [Christensen, 1997]) fournit un excellent exemple de l'influence de la nouveauté sur les résultats observés. Il faut tenir compte de cet effet dans le cadre de l'interprétation des résultats. Concernant les caractéristiques de la mesure, il s'agit, parmi les caractéristiques les plus courantes, soit de la réactivité de la mesure, soit de la sensibilisation au test. La question est la suivante : la réactivité à la mesure consiste à savoir si les changements observés à partir de l'instrument (ordinateur, papier, test de QI) se transposent dans le cadre de la vie quotidienne des individus. Dans le contexte de notre étude, la performance des sujets est évaluée grâce aux temps et à la précision des sélections. Donc les sujets ne peuvent pas répondre de façon différente de celle qu'ils auraient adoptée dans la vie courante.

5.3 Création de la base d'images

Le nombre de scènes différentes présentées aux sujets est de 120 : les mêmes scènes sont présentées aux sujets dans les deux conditions expérimentales PV et PM. Chaque scène contient 30 pastilles, toutes différentes les unes des autres. La banque d'images doit permettre de générer automatiquement les 120 scènes. La même pastille ne peut être contenue dans plusieurs scènes à la fois. Ainsi, la banque d'images doit contenir au moins $120 \times 30 = 3600$ photographies.

Nous avons créé une banque d'images contenant environ 6000 photographies en tenant compte de leur qualité graphique (contraste, luminosité, définition), et de façon à pourvoir équitablement les deux types de photographies considérés, à savoir objets complexes (objets, animaux, personnages) et paysages (montagnes, volcans, paysages urbains, scènes d'intérieur). Toutes les photographies contenues dans la banque proviennent du Web. Nous avons ensuite constitué 60 collections d'objets complexes et 60 collections de paysages, contenant chacune au moins 30 photographies, en veillant à respecter les contraintes suivantes :

- une photographie ne peut appartenir qu'à une seule collection ;
- les 60 collections d'un même type (objet ou paysage) doivent être équitablement réparties entre les trois niveaux de difficultés, facile (1), moyen (2) et difficile (3), soit 20 collections de niveau 1, 20 collections de niveau 2 et 20 collections de niveau 3.

Nous illustrons la façon dont nous avons réparti les photographies en niveaux de difficulté. Considérons par exemple le thème des animaux. Nous avons constitué quatre collections d'animaux par niveau de difficulté. Nous avons utilisé la méthode suivante. D'après les caractéristiques des collections du paragraphe 5.1.3 page 73, le niveau de difficulté 1 comprend toutes les collections hétérogènes, le niveau de difficulté 2 toutes les collection homogènes et le niveau de difficulté 3 toutes les collection homogènes avec en plus une complexité de détail. Nous avons amélioré ces critères en ajoutant la notion d'échelle des plans : gros plan, plan moyen, plan général, etc. En effet, plus le sujet est photographié en gros plan, plus il est facile d'en percevoir les détails. Plus il est photographié en plan large, plus il est difficile d'en percevoir les détails.

Ainsi, pour constituer les collections d'animaux de niveau 1, nous avons regroupé des photographies d'animaux qui non seulement formaient une collection hétérogène, mais aussi présentaient les animaux en gros plan. Pour constituer les collections d'animaux de niveau 2, nous

avons regroupé des photographies d'animaux qui non seulement formaient une collection homogène, mais aussi présentaient les animaux en plan moyen ou avec vue en pied[64]. Pour constituer les collections d'animaux de niveau 3, nous avons regroupé des photographies d'animaux qui non seulement formaient une collection homogène, mais aussi présentaient les animaux en plan large ou en plan général[65]. Nous donnons un exemple de collection pour chaque niveau de difficulté (cf. infra les figures 5.4, 5.5 et 5.6).

FIG. 5.4 – Exemple de collection de niveau 1.

Nous avons fixé le nombre de scènes par condition à 120, obtenant ainsi 30 scènes par structure, les 30 positions possibles de la cible par structure pouvant être testées. Sur les 30 scènes d'une même structure, 15 sont des objets complexes, 15 sont des paysages. Parmi les 15 objets

[64]Par exemple, on trouve des plans moyens dans les portraits laissant à l'attitude et au costume un rôle dans la signification. La vue en pied correspond à la représentation intégrale du personnage qui remplit le cadre de l'image ; on la trouve dans les portraits officiels ou dans les vignettes de B.D au moment où l'action privilégie un personnage (cf. [Cadet *et al.*, 2002] page 18).

[65]Le plan large permet d'évoquer globalement l'action et de suggérer le contexte sans lui accorder une place particulière. Le plan général est utilisé pour photographier les paysages (cf. [Cadet *et al.*, 2002] page 18).

FIG. 5.5 – Exemple de collection de niveau 2.

FIG. 5.6 – Exemple de collection de niveau 3.

ou paysages d'une même structure, 5 sont de niveau de difficulté 1, 5 sont de niveau de difficulté 2 et 5 sont de niveau de difficulté 3.

Nous avons numéroté les collections de la manière suivante : chaque répertoire contenant une collection est nommé DiffXXX, où -XXX est un numéro de collection attribué en tenant compte à la fois du niveau de difficulté (1, 2 ou 3) et du type de photographies contenu dans la collection. Le premier chiffre donne le niveau de difficulté de la collection, les deux suivants le numéro de la collection en respectant l'ordre suivant. Les objets complexes sont numérotés de 0 à 19 et les paysages sont numérotés de 20 à 39. Par exemple, la collection de la figure 5.4 page 87 est contenue dans le répertoire Diff110, celle de la figure 5.5 page 88 dans le répertoire Diff211, et celle de la figure 5.6 page 89 dans le répertoire Diff314.

Nous avons réparti les collections équitablement entre les quatre structures testées, créant ainsi une arborescence à quatre niveaux. Le premier niveau est le répertoire "Images" qui contient quatre sous-répertoires "Absence", "Ellipse", "Matrice" et "Radian". Ces quatre sous-répertoires de deuxième niveau contiennent chacun deux sous-répertoires "Objets" et "Paysages" qui, eux, contiennent les collections (cf. l'exemple ci-dessous figure 5.7).

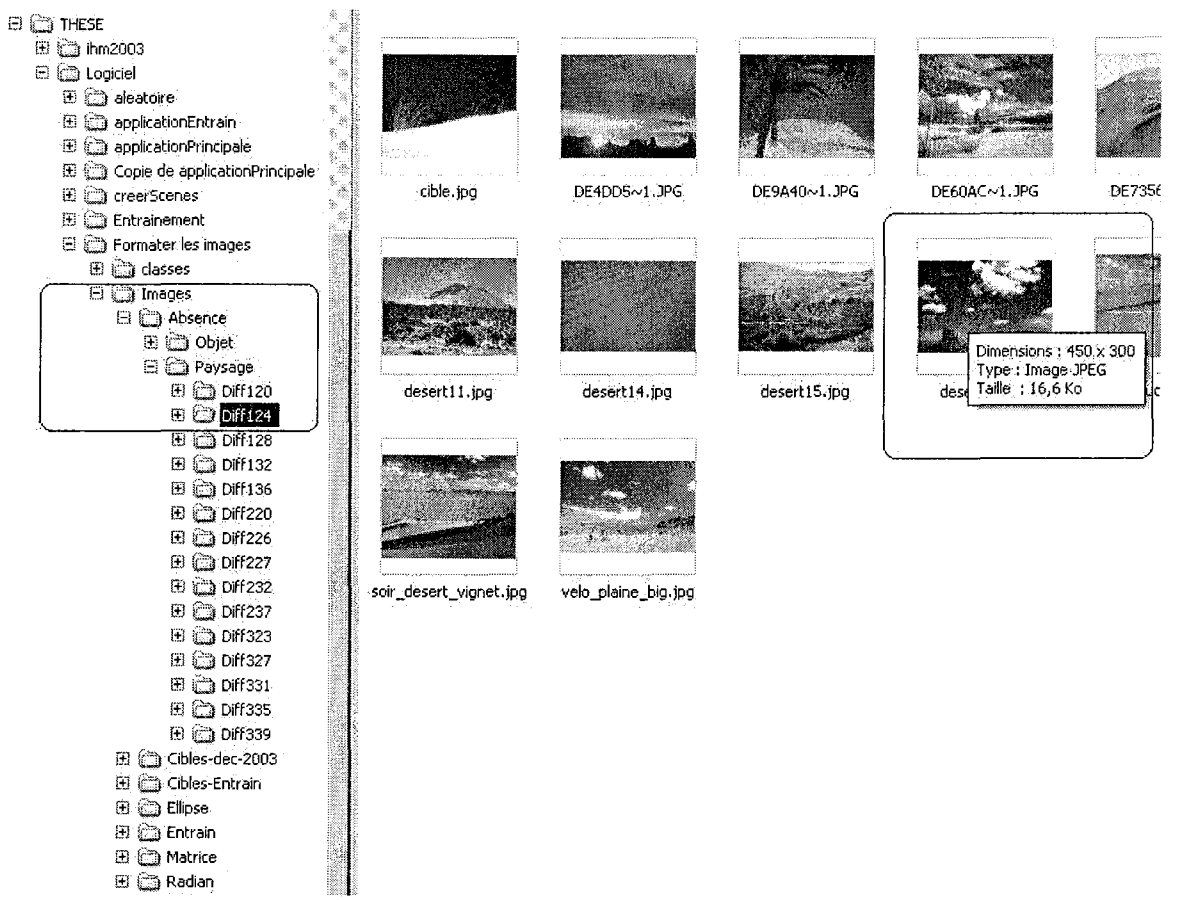


FIG. 5.7 – Arborescence de la base d'images.
Encadré gauche : chemin d'accès à la collection Diff124. Encadré droit : les images récupérées sur le Web n'ont pas toutes la même taille et ne sont pas toutes exactement au format 4/3.

5.4 Développements logiciels réalisés

La mise œuvre du protocole expérimental adopté imposait, pour être réalisable facilement, l'informatisation de la création des 120 images présentées aux sujets (3600 photographies) à l'exception de la constitution des ensembles de photographies correspondant à chaque image (base d'image), ainsi que celle du déroulement des passations des 24 sujets et le recueil des mesures de performance.

Après avoir constitué une base de données de photographies, nous avons donc automatisé la construction des scènes présentées aux sujets, en particulier la disposition des photographies dans les différentes structures, ainsi que l'ordre d'affichage des scènes pour les différents sujets. Rappelons que l'ordre des scènes est aléatoire pour chaque sujet ; de même, pour chaque sujet, l'ordre des scènes varie d'une condition expérimentale à l'autre.

Le logiciel calcule et fournit également les temps de réponse et la position exacte des sélections. Tous les développements logiciels ont été réalisés en Java.

5.4.1 Création des images

Nous avons formulé précédemment des contraintes sur les photographies d'une collection[66]. Les photographies doivent être toutes de même taille et au format 4/3 au sein d'une scène pour neutraliser la saillance visuelle d'un ou plusieurs éléments dans la scène. La résolution de l'écran de passation étant 1280 × 1024, nous avons choisi comme taille pour les photographies 120 × 90 pixels.

Pour créer les collections respectant ces contraintes, nous avons développé une application qui parcourt toutes les collections contenues dans l'arborescence de la base d'images. Chaque image de la base est parcourue et redimensionnée au format 120 × 90 pixels.

Pour créer les scènes à structure elliptique, matricielle et radiale, nous avons créé manuellement des "structures à trous" au sein desquelles le logiciel place de façon automatique et aléatoire les différentes photographies d'une collection. Il a fallu calculer les positions de chaque photographie de façon à ce qu'elles soient parfaitement symétriques et soient ajustées le mieux possible à la taille de l'écran[67]. La disposition des photographies d'une collection au sein des "structures à trous" est aléatoire. En outre, le logiciel est conçu de sorte que les cibles visitent les 30 positions possibles au sein d'une même structure.

La création des scènes à structure aléatoire a nécessité un algorithme différent. En effet, les scènes dont la structure est aléatoire devant toutes être différentes, nous avons créé automatiquement autant de "structures aléatoires à trous" que de scènes aléatoires nécessaires (30). En outre, ces 30 scènes étant toutes différentes, on ne peut considérer 30 positions à visiter. La cible a donc été placée, automatiquement, à la première position de chaque structure aléatoire à trous.

[66] cf. supra 5.1.3 page 73.

[67] Ces positions ne sont pas paramétrables puisqu'elles sont déterminées en fonction de la résolution de l'écran de passation.

Le choix des cibles a été effectué manuellement par deux juges pour garantir leur saillance visuelle neutre au sein de la collection. Chaque structure est illustrée par un exemple (cf. infra les figures 5.8 à 5.11).

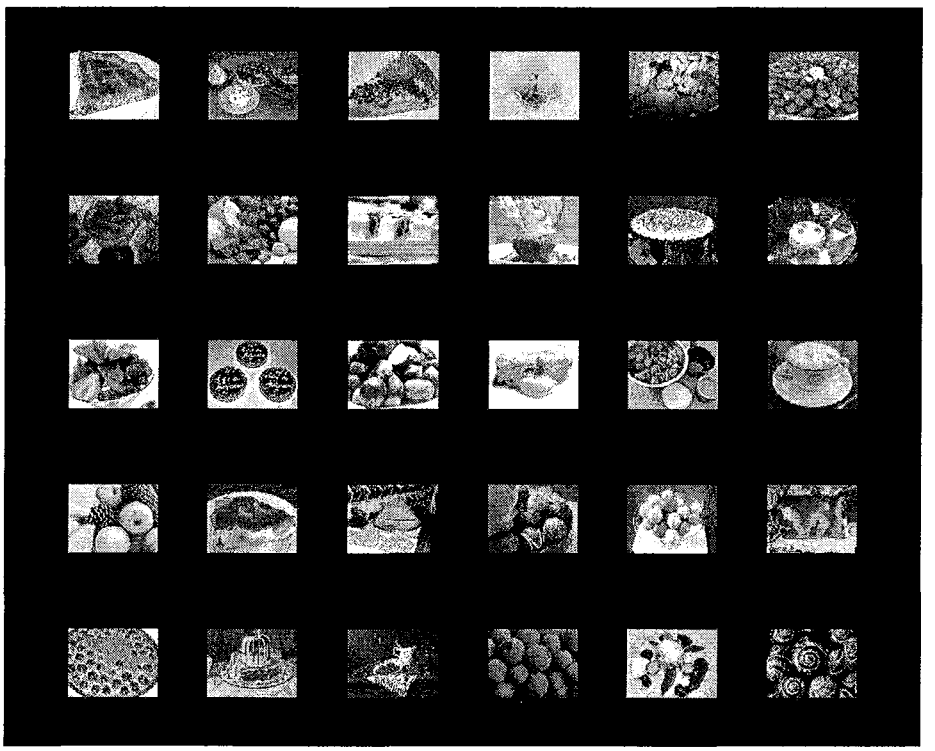


FIG. 5.8 – Exemple de scène matricielle.
Thème de la collection : Desserts. Niveau de difficulté : 1.

5.4.2 Automatisation du déroulement des passations

Outre la création des couples (scène + cible), nous avons automatisé le déroulement des passations. Pour ce faire, le logiciel crée pour chacun des sujets (24) un ordre aléatoire de présentation des couples (120) pour chacune des conditions (PV et PM). Le logiciel crée en fait deux fichiers exécutables par sujet, un par condition expérimentale, chacun des exécutables contenant la liste aléatoire de présentation des scènes. Pour minimiser l'effet de mémorisation des scènes entre les deux conditions, nous avons ajouté la contrainte suivante : toute scène entre la $1^{ère}$ et la $60^{ème}$ position de la liste aléatoire correspondant à la première condition expérimentale, doit figurer entre la $1^{ère}$ et la $60^{ème}$ position de la liste aléatoire correspondant à la seconde condition expérimentale ; idem pour les 60 scènes entre la $60^{ème}$ et la $120^{ème}$ position.

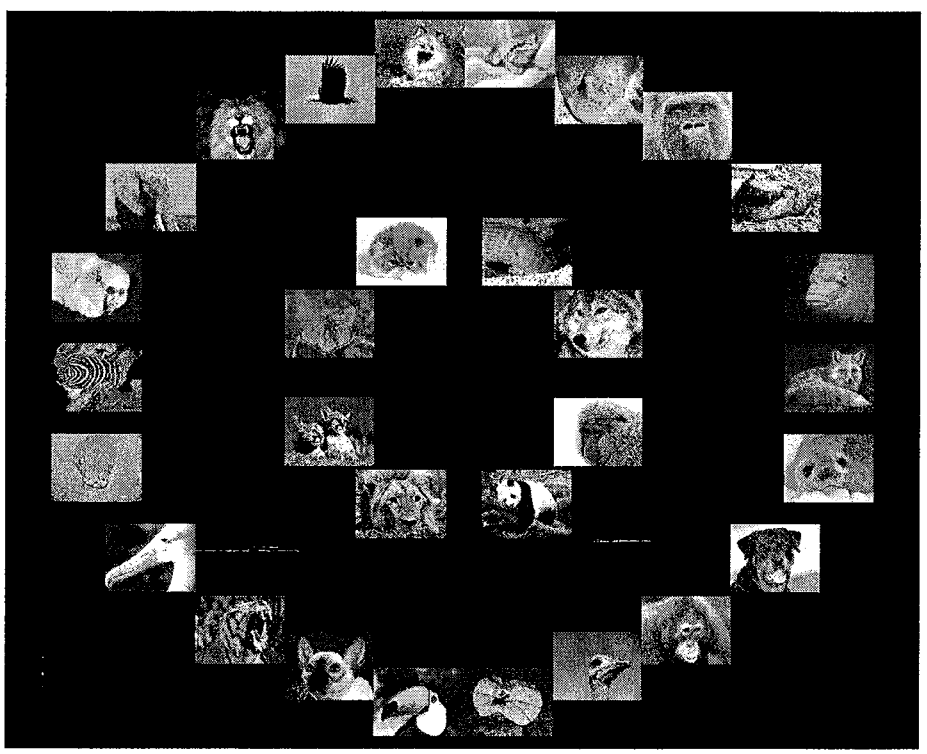


FIG. 5.9 – Exemple de scène elliptique.
Thème de la collection : Animaux. Niveau de difficulté : 1.

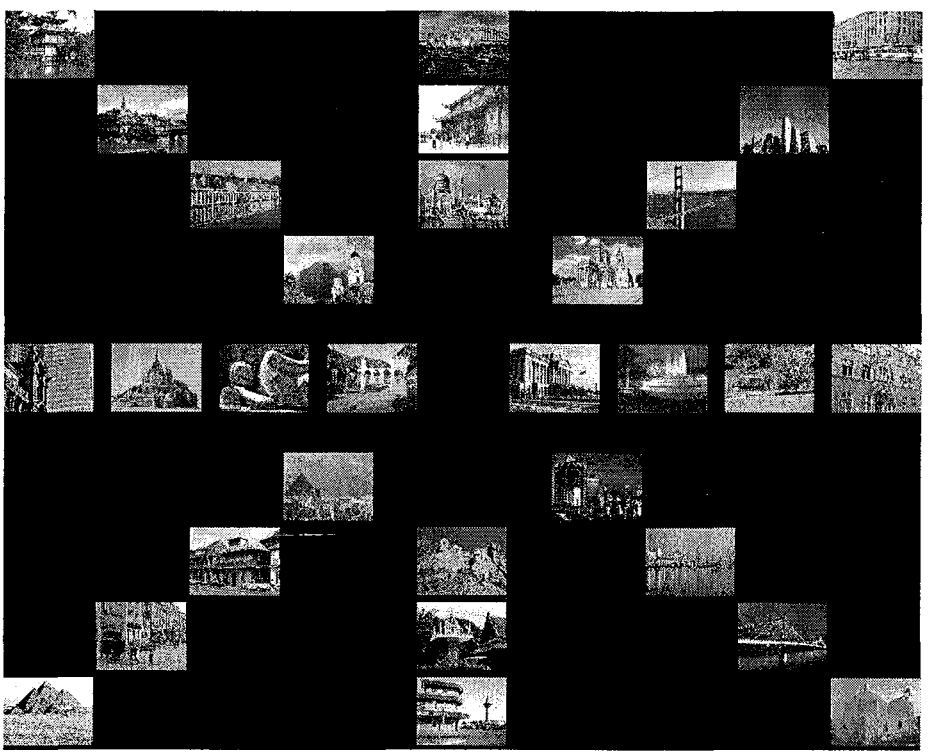


FIG. 5.10 – Exemple de scène radiale.
Thème de la collection : Bâtiments. Niveau de difficulté : 1.

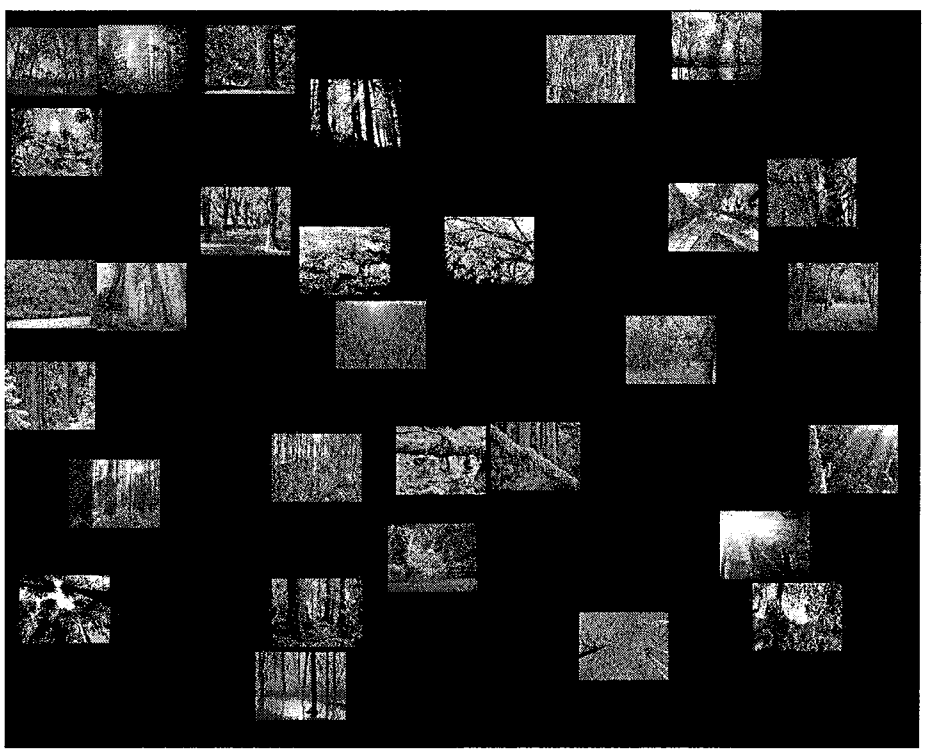


FIG. 5.11 – Exemple de scène non structurée, ou dont la structure est aléatoire. Thème de la collection : Forêt. Niveau de difficulté : 1.

5.4.3 Recueil des données

Le logiciel calcule en temps réel les temps de réponse des sujets ainsi que la position exacte des sélections. Il crée automatiquement, pour chaque sujet, un fichier de ses résultats qui contient pour chaque scène, dans l'ordre :

- le numéro de la scène ;
- la zone de localisation de la cible (haut-gauche, haut, haut-droit, etc.), ainsi que ses coordonnées (x,y) en pixels ;
- le chemin d'accès à la scène qui permet d'obtenir directement sa structure, son type, son niveau de difficulté grâce à l'arborescence de la base d'images (cf. supra figure 5.7 page 90) ;
- le mode de présentation de la cible ;
- la position du clic de sélection de la cible (x,y) en pixels ;
- l'intervalle de temps, en millisecondes, entre le clic qui déclenche l'affichage de la scène et celui de sélection de cible.

5.5 Déroulement de la passation

Le déroulement d'une passation est le même pour tous les sujets, à l'exception de l'ordre de passation des conditions expérimentales PV et PM. D'abord, le sujet lit la présentation générale de l'expérimentation et répond à un rapide questionnaire qui concerne son statut, ses compétences et ses activités informatiques. L'expérimentateur lui fait ensuite passer des tests de vision. Après la lecture des consignes et l'entraînement, le sujet réalise les tâches de repérage dans les deux conditions. Il remplit un questionnaire après chaque condition. Le remplissage du troisième questionnaire est suivi d'un entretien.

La durée globale d'une passation est d'environ 1H30 et comprend :

- le remplissage du premier questionnaire (environ 5 minutes) ;
- les tests de vision (environ 15 minutes) ;
- la présentation des consignes et l'entraînement initial (environ 5 minutes) ;
- la réalisation des tâches de repérage dans la première condition (environ 20 minutes) ;
- le remplissage du deuxième questionnaire qui porte sur la première condition (environ 5 minutes) ;
- la réalisation des tâches de repérage dans la seconde condition (environ 20 minutes) ;
- le remplissage du troisième questionnaire qui porte sur la seconde condition et l'ensemble des deux conditions ainsi que l'entretien final (environ 20 minutes).

5.5.1 Tests de vision

Ces tests ont été réalisés avec du matériel Stereo Optical Co., Inc. Il s'agit des tests de vision BIOPTOR (TM. Reg. U.S. Pat. Off; by Stereo Optical Co., Inc).

Les tests portant sur la vision de loin sont les suivants :

- fusion verticale : ce test permet de mesurer la convergence du regard lors de la fusion verticale de deux stimuli disjoints ;

- fusion latérale : ce test permet de mesurer la convergence du regard lors de la fusion latérale de deux stimuli disjoints. On parle d'ésophorie lorsque les yeux ont tendance à converger vers le même point, d'exophorie lorsqu'ils ont tendance à diverger dans deux directions différentes ;
- fusion centrale : ce test permet d'estimer la capacité d'une petite zone centrale de l'œil à fusionner deux stimuli disjoints, sans l'assistance de la vision périphérique (contrairement aux deux tests précédents) ;
- acuité visuelle : ce test permet de mesure l'aptitude à lire des lettres (Lettres de Snellen) à différentes distances, donc de différentes tailles (effet de perspective) pour l'œil gauche, l'œil droit, les deux yeux ;
- vision stéréoscopique : ce test mesure l'aptitude à juger des distances relatives en profondeur (vision 3D) lorsque tous les indices visuels, excepté la disparité binoculaire, sont éliminés. La disparité binoculaire est un décalage horizontal sur les deux images rétiniennes d'un même objet ;
- discrimination des couleurs : ce test consiste en quatre reproductions précises des Assiettes Pseudo Isochromatiques d'Ishahari qui mesurent l'aptitude à la discrimination des couleurs. L'ordre de présentation des assiettes n'est pas fonction de la difficulté, mais permet de détecter des anomalies variées de la vision des couleurs.

Tous ces tests portent sur la vision binoculaire, excepté le test d'acuité visuelle qui porte également sur la vision monoculaire. Les tests portant sur la vision de près reprennent les tests de fusion latérale et centrale, ainsi que le test d'acuité visuelle décrit précédemment.

5.5.2 Entraînement des sujets

Un entraînement du sujet sur dix images succède à la lecture des consignes par l'expérimentateur. Cet entraînement comprend cinq images par situation, dans l'ordre PV-PM, respectivement PM-PV, si le sujet passe l'expérimentation dans l'ordre PV-PM, respectivement PM-PV. Comme dans l'étude préliminaire, l'expérimentateur commente la consigne et présente les tâches. Il assiste à l'entraînement du sujet, puis répond à ses éventuelles questions. Il quitte la salle pendant la passation, mais est disponible par téléphone. Le sujet dispose en permanence d'une version écrite de la consigne et du numéro où il peut contacter l'expérimentateur.

Questionnaires

Le premier questionnaire concerne le profil des participants (cf. annexe B). Il nous a permis de recueillir des informations générales sur les participants, telles que leur niveau général de culture (diplôme en cours de préparation ou profession) et leurs compétences en informatique (Internet, logiciels grand public, logiciels graphiques, programmation, etc.).

Les deux autres questionnaires concernent, entre autres, l'évaluation des modes de présentation de la cible et la comparaison des différentes structures testées (cf. annexe B). Le deuxième questionnaire porte sur la tâche de repérage visuel dans la première condition expérimentale. Nous avons demandé aux sujets d'évaluer notamment le niveau de difficulté de la tâche de repérage dans cette condition, leur rapidité, leur confort. Le troisième questionnaire porte sur la tâche

de repérage dans la seconde condition expérimentale. Les questions sont les mêmes que dans le deuxième questionnaire. Nous leur avons demandé d'évaluer, en plus, l'efficacité des différentes structures des affichages que comportait l'expérimentation. Enfin, nous leur avons demandé de comparer les deux modes de présentation des cibles.

5.6 Exploitation des données expérimentales : méthodologie

Pour mémoire, nous avons sollicité 24 sujets qui devaient effectuer la tâche de repérage sur 240 scènes, soit 120 dans la condition visuelle PV et 120 dans la condition multimodale PM, avec contrebalancement de l'ordre des conditions : 12 sujets ont effectué les tâches de repérage dans l'ordre PV-PM, et 12 sujets ont effectué les tâches de repérage dans l'ordre PM-PV. Les mêmes images sont présentées dans les deux conditions. Pour chaque sujet, on dispose d'un fichier de résultats. Ce fichier contient 240 entrées, soit une entrée par scène. Pour chaque entrée de ce fichier, on dispose des données relatives à la scène (son numéro, sa structure, son type, son niveau de difficulté), à la cible (la zone de l'écran où elle se situe et ses coordonnées dans la scène en pixels), à la condition expérimentale (PV ou PM) et aux performances des sujets (temps de sélection de la cible en millisecondes, et coordonnées du clic de sélection en pixels). On dispose donc de $240 \times 24 = 5760$ données. Les performances des sujets, comme dans l'étude préliminaire, sont les temps et la précision de la sélection des cibles.

5.6.1 Présentation générale

L'exploitation des données suit le plan suivant. La première partie porte sur l'évaluation de l'apport spécifique des messages multimodaux pour le repérage visuel de cibles, par rapport aux présentations visuelles. Cette partie de l'analyse comprend la comparaison des modes de présentation des cibles, l'évaluation de l'influence de l'ordre de passation PV-PM *versus* PM-PV sur les performances des sujets et la comparaison des stratégies d'exploration visuelle qu'ils adoptent en fonction de l'ordre des conditions. La deuxième partie porte sur l'influence de l'organisation spatiale des affichages sur les performances des sujets. Cette partie de l'analyse compare les temps de réponse des sujets et leur précision en fonction de l'organisation spatiale des scènes. La troisième partie porte sur l'influence du niveau de difficulté des scènes sur les performances des sujets. Cette partie de l'analyse a pour objectif, entre autres, de montrer que l'apport des présentations multimodales des cibles varie en fonction du niveau de difficulté des scènes présentées aux sujets. La quatrième et dernière partie porte, d'une part, sur l'influence de la nature des éléments (paysage *versus* objet) constituant les scènes sur les performances des sujets. Elle porte, d'autre part, sur l'influence de la position des cibles au sein des scènes sur les performances des sujets.

Il convient de noter que les analyses statistiques ont été réalisées en collaboration avec François-Xavier Jollois, Maître de Conférence à l'Université René Descartes (Paris V).

5.6.2 Résultats des tests de vision

Tous les sujets ayant participé à l'expérimentation présentaient une vue normale en termes d'acuité visuelle, de fusion latérale et de fusion centrale de près, mais aussi en termes de discrimination des couleurs. Nous avons accordé moins d'importance aux résultats observés quant à leur vision de loin, bien que nous n'ayons pas relevé d'anomalie.

5.6.3 Filtrage des données

Nous avons dénombré 229 erreurs, soit 229 sélections à la souris identifiées comme en-dehors de la cible. Ces 229 données forment l'ensemble des erreurs brutes, parmi lesquelles il convient de distinguer :

- les erreurs effectives, i.e., lorsque le sujet clique sur une photographie qui n'est pas la cible ;
- les clics-souris sur le fond noir, i.e., lorsque le sujet ne clique sur aucune photographie car il ne détecte pas la cible ;
- les clics-souris juste à côté de la cible, i.e., lorsque le sujet détecte la cible mais clique juste à côté, ou bien par manque de précision, ou bien dans un souci de rapidité de sélection.

Nous avons exclu les "dérapages", ou sélections juste à côté de la cible, des analyses qualitatives des erreurs. Nous justifions ce choix de la manière suivante : de récents travaux à l'aide d'un eye-tracker [Pelz *et al.*, 2001] portant sur la coordination des mouvements des yeux, de la tête et des mains dans des tâches naturelles, ont montré que pour des tâches de sélection à la souris, la fixation oculaire de la cible s'arrête juste avant que le curseur de la souris n'atteigne la cible.

C'est en appliquant successivement deux filtres (F1) et (F2) que nous obtenons trois catégories d'erreurs : les erreurs effectives, les clics sur le fond noir et les "dérapages". Le filtre (F1) permet d'isoler les "dérapages". Nous considérons qu'une cible associée à une scène a été détectée par un sujet, si la sélection à la souris correspondante se situe dans le voisinage proche de la cible. Nous avons fixé la tolérance pour reclasser une erreur en "dérapage" à 10 pixels, horizontalement et verticalement, car cela correspond à une augmentation d'environ 10% de la surface totale de la cible. Le filtre (F2) permet d'isoler les clics sur le fond noir. Si une sélection ne désigne pas une photographie non cible de la scène, alors il s'agit d'un clic sur fond noir. Après application de (F1) et (F2) aux données, nous obtenons :

- 145 erreurs effectives ;
- 62 clics sur le fond noir ;
- 22 dérapages.

5.7 Présentation visuelle *versus* présentation multimodale

Cette partie de l'analyse a pour objectif de valider l'hypothèse de travail A (cf. supra 5.1.2 page 72). Autrement dit, il s'agit de prouver l'apport spécifique des indications spatiales contenues dans les messages multimodaux (condition PM) pour le repérage visuel de cibles, par rapport aux présentations visuelles isolées des cibles (condition PV).

Dans cette section, l'analyse statistique a consisté, dans un premier temps, à comparer les performances des sujets dans les deux conditions expérimentales PV et PM. Elle a consisté, dans un deuxième temps, à évaluer l'influence potentielle de l'ordre de passation, PV puis PM ou PM puis PV. En outre, les analyses globales de cette section portent sur les erreurs brutes observées lors de la passation. Nous distinguons les erreurs effectives (sélections d'un élément non-cible) des dérapages et des clics sur le fond noir pour les analyses plus fines portant sur la structure des affichages ou les niveaux de difficulté des scènes.

5.7.1 Résultats globaux

Afin de comparer les performances des sujets entre les deux conditions expérimentales, nous avons regroupé les données recueillies par condition (cf. infra tableau 5.3).

Variable : temps de sélection des cibles (ms)			
Condition	Moyenne (ms)	Écart type (ms)	Nombre d'observations
Visuel (PV)	5674	5985	2880
Multimodal (PM)	1747	1552	2880

Variable : précision des sélections de cibles			
Condition	Nombre d'erreurs	Taux d'erreurs (%)	Nombre d'observations
Visuel (PV)	150	5.2	2880
Multimodal (PM)	79	2.7	2880

TAB. 5.3 – Résultats globaux.
Le premier tableau présente les résultats concernant les temps de sélection des cibles. Le deuxième présente les résultats concernant la précision des sélections.

Rapidité de la sélection des cibles

Le temps moyen de sélection des cibles observé dans la condition visuelle (5674 ms), associé à un écart type élevé (5985 ms), est plus de 3 fois supérieur à celui observé dans la condition multimodale (1747 ms), associé à un écart type faible (1552 ms). En d'autres termes, les sujets sont 3 fois plus rapides et beaucoup plus réguliers dans la condition multimodale qu'ils ne le sont dans la condition visuelle. Ce résultat est statistiquement hautement significatif (t=-34,07 ; $p<0,0001$).

Précision de la sélection des cibles

Globalement, le taux d'erreurs est peu élevé (moins de 6% dans les deux conditions expérimentales). Néanmoins, le taux d'erreurs observé dans la condition visuelle (PV) est près de 2 fois supérieur à celui observé dans la condition multimodale (PM).

Analyse des questionnaires

Les 24 sujets se jugent plus rapides dans la condition multimodale qu'ils ne pensent l'être dans la condition visuelle. Quel que soit l'ordre de passation, tous les sujets sont plus rapides dans la condition multimodale.

En revanche, en ce qui concerne la précision de la sélection des cibles, il convient de considérer l'ordre de passation. Tous les sujets (12) ayant effectué les tâches de repérage dans l'ordre PV-PM, ont été plus précis dans la condition multimodale. Dix d'entre eux l'ont exprimé dans le questionnaire. Un sujet est resté sans avis (une seule erreur dans la condition visuelle, aucune dans la condition multimodale). Un autre n'a pas observé de différence. Pourtant ce sujet a commis 16 erreurs dans la condition visuelle contre seulement 4 dans la condition multimodale.

Parmi les sujets ayant effectué les tâches de repérage dans l'ordre PM-PV (12 sujets) :
- 6 sont plus précis dans la condition multimodale (PM) ;
- 5 sont plus précis dans la condition visuelle (PV) ;
- 1 a commis autant d'erreurs dans les deux conditions (3 erreurs).

Parmi les 6 sujets plus précis dans la condition multimodale, 4 l'ont déclaré, les deux autres n'ont pas observé de différence[68]. Parmi les 5 sujets plus précis dans la condition visuelle, seul, un sujet l'a exprimé dans le questionnaire. Un sujet a déclaré ne pas observer de différence[69]. Les 3 autres sujets ont déclaré avoir été plus précis dans la condition multimodale.

En ce qui concerne la facilité du repérage, 75% des sujets (18 sujets) ont jugé les tâches de repérage plus faciles dans la condition multimodale. Un sujet a jugé le repérage plus facile dans la condition visuelle : ce sujet a également précisé avoir effectué les tâches de repérage dans l'ordre PM-PV. Quatre sujets n'ont pas ressenti de différence entre les deux conditions. Un sujet s'est déclaré sans avis sur la question. D'après les questionnaires, la condition multimodale supprime les hésitations et les clics "au hasard", plus fréquents dans la condition visuelle.

Une large majorité des sujets a préféré effectuer les tâches de repérage dans la condition multimodale (16 sujets). Les autres sujets ont déclaré avoir été plus intéressés par la condition visuelle car la tâche leur a paru plus "amusante" et/ou "stimulante" dans cette condition (cf. les debriefings post-expérimentation).

Une large majorité des sujets a décrit la tâche comme familière (16 sujets). Aucun des huit sujets pour lesquels la tâche n'a pas semblé familière n'a été gêné par son caractère inhabituel. Les messages ont été considérés comme une assistance efficace à la tâche par tous les sujets, et

[68] 3 erreurs dans la condition PV et 1 erreur dans la condition PM pour le premier, 4 erreurs dans la condition PV et 5 erreurs dans la condition PM pour le deuxième.

[69] 2 erreurs dans la condition PV et 5 erreurs dans la condition PM.

comme une aide à la mémorisation pour certains d'entre eux (environ 1/4). De plus, les sujets n'ont pas été surpris par les messages car ils étaient décrits au préalable dans la consigne.

Conclusions

D'après l'analyse globale des différences entre les deux modes de présentation des cibles, la présentation multimodale de la cible facilite et améliore le repérage visuel par rapport à la présentation visuelle de la cible, en termes du temps de sélection et de la précision de la sélection de la cible. Ce résultat est statistiquement significatif concernant les temps moyens de sélection des cibles et valide l'hypothèse A.

D'après les commentaires qui nous ont été faits lors des debriefings, il apparaît que les sujets suivent le message sonore contenu dans les présentations multimodales. Les sujets se dirigent immédiatement vers la zone indiquée oralement. Les messages sonores leur permettent ainsi de réduire efficacement la zone de recherche. En n'imposant pas un parcours systématique de l'organisation ou structure spatiale de la scène, et en réduisant le nombre de confusions possibles entre la cible et les non cibles, les messages sonores permettent aux sujets d'être plus efficaces.

Les sujets déclarent adopter deux stratégies différentes en fonction du mode de présentation des cibles. Dans la condition visuelle, ils se construisent mentalement une description verbale de la cible, d'un détail qu'ils espèrent être discriminant. Au bout de 10 scènes, ils ont observé la "cohérence thématique" de chaque scène : trouver ce détail pertinent devient plus aisé. Au moment où la cible apparaît, ils parcourent la scène en suivant, à deux exceptions près, sa structure visuelle. Dans le cas des structures aléatoires, ils se dirigent vers les zones de l'écran les plus denses. Dans la condition multimodale, ils suivent simplement l'indication contenue dans le message sonore. Cette condition ne leur impose ni d'élaborer une description mentale et verbale la cible ou d'un détail de la cible, ni de suivre la structure spatiale de la scène.

En d'autres termes, dans la condition multimodale, les sujets tirent avantage de l'information spécifique véhiculée par chacune des modalités. La présentation visuelle contenue dans la présentation multimodale "montre" la cible, ses caractéristiques, comme sa forme, sa nature, ou sa couleur, par exemple. Le message sonore de localisation spatiale de la cible permet d'indiquer précisément et de réduire la zone de recherche. Ces informations suffisent alors aux sujets pour repérer la cible :

- sans qu'ils aient à parcourir la structure de la scène dans son intégralité ;
- sans qu'ils aient recours à la description linguistique d'un détail, qui "risque d'être pertinent".

Les sujets, habitués rapidement à la corrélation entre la cible et le thème de la scène, formulent une hypothèse *a priori* sur le thème de la scène, à partir de la présentation de la cible.

En outre, les sujets ne semblent pas gênés par les messages sonores et les acceptent comme une "assistance à la tâche". Les messages sonores ne semblent pas augmenter leur charge cognitive : au contraire, les sujets ressentent moins de fatigue lors de la condition multimodale que lors de la condition visuelle. La tâche, dans cette condition, est décrite comme plus facile et plus rapide. La fusion des deux types d'informations, présentation visuelle de la cible et indication orale de sa localisation dans la scène, ne semble pas surcharger la mémoire de travail. Des résultats similaires

issus de la psychologie cognitive sont présentés dans [Kalyuga *et al.*, 1999]. Dans cette étude, les sujets devaient effectuer des exercices interactifs d'apprentissage sur le thème de la fusion en soudure dans l'une des trois conditions expérimentales suivantes, définies en fonction du format des supports :

- la condition "Visual plus Audio text" où un texte, présenté sous forme écrite et orale, accompagne un diagramme ;
- la condition "Visual text" où un texte, présenté sous forme écrite exclusivement, accompagne un diagramme ;
- la condition "Audio text" où un texte, présenté exclusivement de façon orale, accompagne un diagramme.

Les présentations orales du texte sont supérieures aux présentations visuelles (cf. "Audio text" *versus* "Visual text"), mais pas lorsque le texte est présenté sous forme orale et écrite (cf. "Visual plus Audio text") en raison de la redondance des informations qui imposent une charge cognitive qui interfère avec l'apprentissage. La condition "Audio text" correspondrait à notre condition PM.

Enfin, ces résultats tirés des commentaires des sujets recueillis à la suite de l'expérimentation correspondent à ceux présentés dans [Pelz *et al.*, 2001] sur la création d'une représentation mentale de la tâche.

5.7.2 Influence de l'ordre de passation sur les performances des sujets

Afin de mettre en évidence l'éventuelle influence de l'ordre de passation PV-PM ou PM-PV sur les performances des sujets, nous avons comparé le temps moyen de sélection ainsi que le nombre d'erreurs entre les deux groupes de sujets dans chacune des conditions expérimentales. Pour ce faire, les données recueillies lors de l'expérimentation ont été regroupées par condition et par groupe de sujets[70], formant ainsi 4 ensembles de 1440 données[71] :

- 1440 données pour le groupe 1 dans la condition PV (G1V) ;
- 1440 données pour le groupe 1 dans la condition PM (G1M) ;
- 1440 données pour le groupe 2 dans la condition PV (G2V) ;
- 1440 données pour le groupe 2 dans la condition PM (G2M).

Nous avons calculé la moyenne et l'écart type des temps de sélection ainsi que le nombre et les taux d'erreurs pour chaque ensemble de données G1V, G1M, G2V, G2M (cf. infra tableau 5.4 page 104).

Rapidité de la sélection des cibles

Dans la condition visuelle, on observe un temps moyen de sélection des cibles de 6935 ms pour le groupe 1 (associé à un écart type de 6930 ms) et de 4413 ms pour le groupe 2 (associé à un écart type de 4524 ms). Dans la condition multimodale, on observe un temps moyen de sélection des cibles de 1855 ms pour le groupe 1 (associé à un écart type de 1651 ms) et de 1641 ms pour le groupe 2 (associé à un écart type de 1439 ms).

[70] Groupe 1 : PV-PM ; groupe 2 : PM-PV.

[71] Une donnée est un couple (temps de sélection ; précision de la sélection).

Variable : temps de sélection des cibles (ms)					
Condition	Groupe	Ensemble	Moyenne (ms)	Écart type (ms)	Nombre d'observations
PV	1	G1V	6935	6930	1440
PV	2	G2V	4413	4524	1440
PM	1	G1M	1855	1651	1440
PM	2	G2M	1641	1439	1440

Variable : précision de sélection des cibles					
Condition	Groupe	Ensemble	Nombre d'erreurs	Taux d'erreurs (%)	Nombre d'observations
PV	1	G1V	76	5,2	1440
PV	2	G2V	74	5,1	1440
PM	1	G1M	17	1,2	1440
PM	2	G2M	62	4,3	1440

TAB. 5.4 – Résultats par groupe de sujets.
Le premier tableau présente les résultats concernant les temps de sélection des cibles par condition et par ordre de passation. Le deuxième présente les résultats concernant la précision des sélections par condition et par ordre de passation.

Quelle que soit la condition expérimentale considérée, les sujets du groupe 2 sont plus rapides que les sujets du groupe 1. En effet, dans la condition visuelle, la différence moyenne de 2522 ms observée entre les deux groupes est statistiquement significative ($t=11,56$; $p<0,0001$). Dans la condition multimodale, la différence moyenne observée entre les deux groupes est de 213 ms. Bien que cet écart soit moins important que dans la condition visuelle, il reste statistiquement significatif ($t=3,70$; $p=0,0002$).

Précision de la sélection des cibles

Dans la condition visuelle, on observe 76 erreurs, soit un taux d'erreurs de 5,2%, pour le groupe 1 et 74 erreurs, soit un taux d'erreurs de 5,1% pour le groupe 2. Dans la condition multimodale, on observe 17 erreurs, soit un taux d'erreurs de 1,2%, pour le groupe 1 et 62 erreurs, soit un taux d'erreurs de 4,3% pour le groupe 2. Dans la condition visuelle, la précision des sélections est comparable entre les deux groupes de sujets. En revanche, dans la condition multimodale, le groupe 2 a commis 3 fois plus d'erreurs que le groupe 1.

Toutes conditions expérimentales confondues (PV et PM), les sujets du groupe 1 effectuent les tâches de repérage avec une précision de 98,82% (93 erreurs en tout) tandis que les sujets du groupe 2 les effectuent avec une précision de 95,62% (136 erreurs en tout). Cette différence de plus de 3% est statistiquement significative ($t=5,16$; $p<0,0001$).

Interprétation des résultats

Les résultats de comparaison entre les groupes de sujets 1 et 2 - concernant les temps de sélection d'une part, et la précision des sélections d'autre part - mettent en évidence l'influence de l'ordre de passation sur les performances des sujets. En effet, ces résultats suggèrent que les sujets du groupe 1 auraient adopté une stratégie privilégiant la précision des sélections des cibles, tandis que les sujets du groupe 2 auraient adopté une stratégie privilégiant la rapidité de sélection des cibles. Plusieurs sujets ayant effectué les tâches de repérage dans l'ordre PM-PV ont déclaré que leurs temps de réponse auraient probablement été augmentés dans l'ordre de passation PV-PM. Un sujet, assigné à l'ordre de passation PV-PM, a déclaré avoir recherché la précision plutôt que la rapidité des sélections.

5.7.3 Analyse détaillée des stratégies adoptées par les utilisateurs : rapidité des sélections *versus* précision des sélections

Afin de mettre en évidence les différentes stratégies adoptées par les utilisateurs, nous avons choisi d'analyser l'évolution de leurs performances au cours de l'expérimentation. Cette analyse est basée sur l'évolution des temps moyens de sélection des cibles et sur l'évolution du nombre d'erreurs commises par les sujets dans chaque condition expérimentale, en tenant compte des groupes de sujets (groupe 1 ou groupe 2). Ainsi, pour chaque groupe de sujets, nous avons calculé, dans chacune des deux conditions expérimentales, la moyenne des temps de sélection des cibles ainsi que le nombre total d'erreurs observées toutes les 30 images (cf. annexe B). Les résultats de ces analyses sont représentés graphiquement[72]. À noter que l'ordre d'apparition des couples (scène ; cible) dans les séquences PV et PM étant aléatoire pour chacun des sujets, les données ne sont pas appariées par image.

Rapidité de la sélection des cibles

Nous avons vu dans le paragraphe précédent que le groupe 2 est plus rapide que le groupe 1, dans la condition visuelle, comme dans la condition multimodale. Les différences moyennes entre les deux groupes sont statistiquement significatives [73]. Les graphiques de la figure 5.12 page 106 illustrent ces observations. Les courbes représentant l'évolution des temps moyens de sélection des cibles des deux groupes correspondent à ce résultat. À noter que la différence des temps moyens de sélection des cibles, observée entre les deux groupes est plus importante dans la condition visuelle que dans la condition multimodale.

En effet, dans la condition visuelle, la différence moyenne des temps de sélection des cibles observée entre les deux groupes de sujets est de 2522 ms. La valeur maximale, respectivement minimale, de cette différence est de 3602 ms sur les images 1 à 30, respectivement de 1432 ms sur les images 31 à 60. L'écart sensible des temps moyens de sélection des cibles observé, dans cette condition, entre les deux groupes de sujets, semble provenir de l'ordre de passation. En effet, les sujets du groupe 2 ont effectué les tâches de repérage dans la condition multimodale, avant de les

[72] cf. infra figures 5.12 page5.12 et 5.13 page5.13.

[73] cf. supra 5.7.2 page 103.

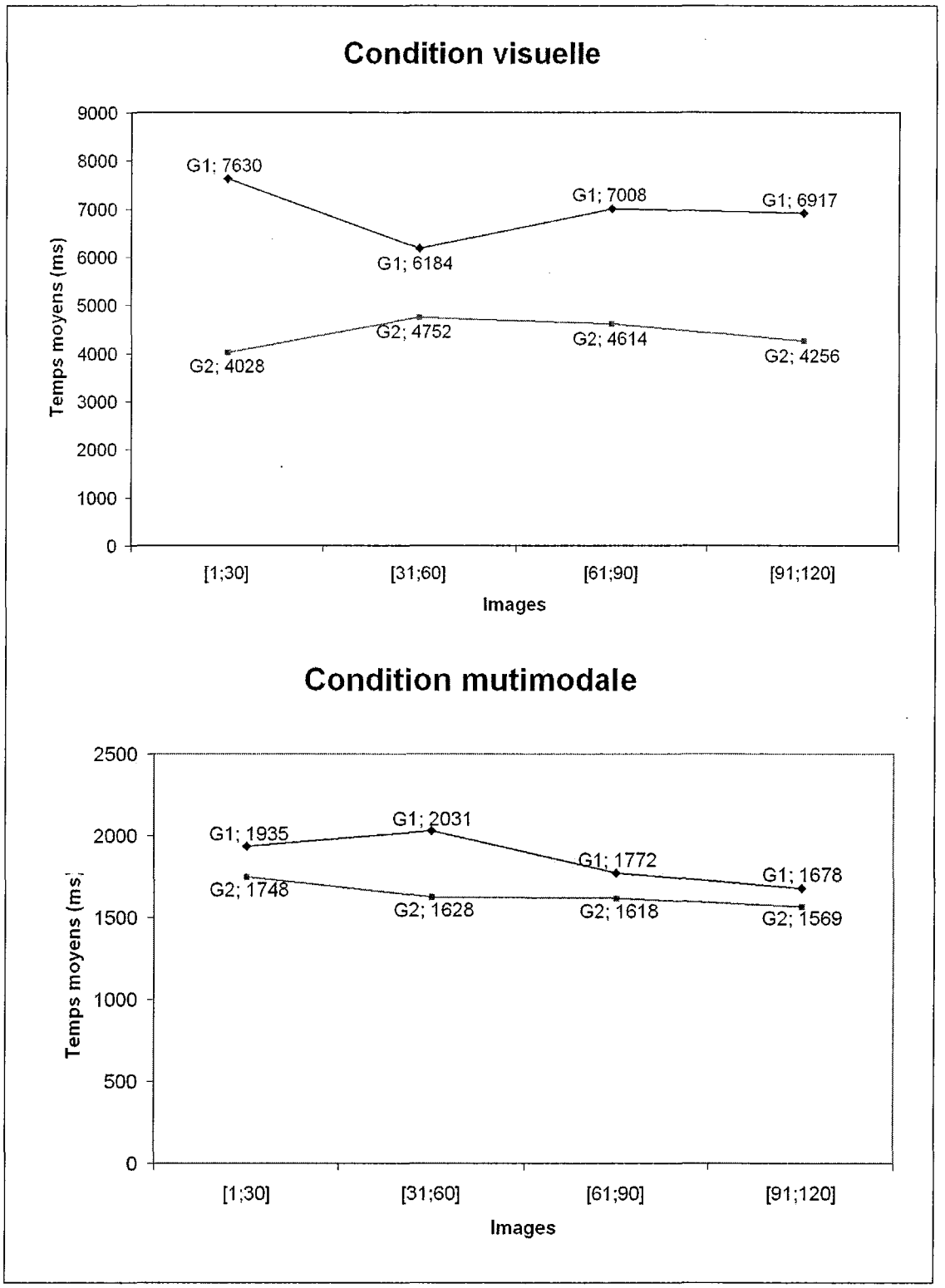


FIG. 5.12 – Évolution des temps moyens de sélection des cibles : Groupe 1 *versus* groupe 2. Le premier graphique représente l'évolution des temps moyens de sélection des cibles dans la condition visuelle (PV), le deuxième dans la condition multimodale (PM).

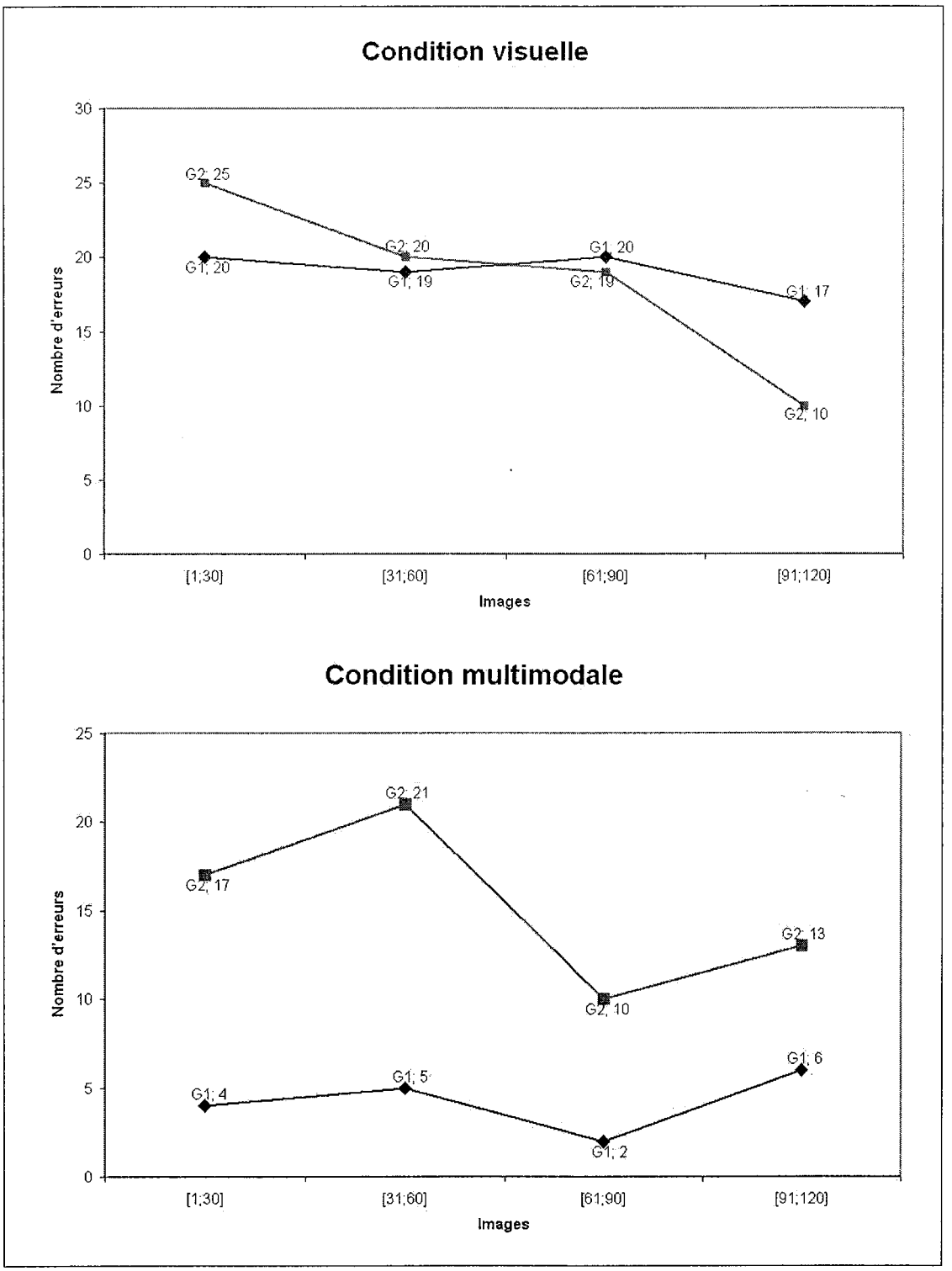


FIG. 5.13 – Évolution de la précision des sélections de cibles : Groupe 1 *versus* groupe 2.
Le premier graphique représente l'évolution du nombre d'erreurs commises par les sujets dans la condition visuelle (PV), le deuxième dans la condition multimodale (PM).

effectuer dans la condition visuelle. Dans la condition visuelle, le groupe 2 connaît déjà l'ensemble des 120 couples (scène ; cible), tandis que le groupe 1 découvre l'ensemble du matériel visuel. Bien que l'ordre de présentation du matériel visuel varie entre les deux conditions expérimentales pour chacun des groupes de sujets, les tâches de repérage dans la condition visuelle sont simplifiées pour les sujets du groupe 2 qui peuvent faire appel à leur mémoire.

Dans la condition multimodale, on observe une différence moyenne des temps de sélection des cibles de 213 ms entre les deux groupes de sujets. La valeur maximale, respectivement minimale, de cette différence est de 403 ms sur les images 31 à 60, respectivement de 109 ms sur les images 91 à 120 (cf. figure 5.12 page 106). Il était légitime de s'attendre à ce que le groupe 1 soit le plus rapide dans cette condition, l'ordre de passation de ce groupe étant PV-PM. Afin d'expliquer ce résultat, nous allons maintenant analyser l'évolution du nombre d'erreurs.

Précision de la sélection des cibles

Les graphiques de la figure 5.13 page 107 illustrent l'évolution du nombre des erreurs commises par les sujets de chaque groupe, par paquet de 30 images. Le faible nombre des erreurs commises par les sujets de chaque groupe dans les deux conditions PV et PM ne permet pas d'effectuer une analyse statistique pertinente des différences.

Dans la condition multimodale, on observe une différence sensible du nombre d'erreurs entre les deux groupes de sujets : 17 erreurs pour le groupe 1 *versus* 61 erreurs pour le groupe 2 ; soit près de 4 fois plus d'erreurs pour le groupe 2 dans cette condition. L'évolution du nombre d'erreurs (cf. figure 5.13 page 107) montre que, dans chacun des paquets de 30 scènes, le groupe 2 commet 4 fois plus d'erreurs que le groupe 1, à l'exception du dernier paquet d'images où on observe 6 erreurs pour le groupe 1 *versus* 13 erreurs pour le groupe 2. Dans la condition multimodale, le groupe 1 est plus précis que le groupe 2.

Dans la condition visuelle, on n'observe pas de différence sensible du nombre d'erreurs entre les deux groupes de sujets : 76 erreurs pour le groupe 1 *versus* 74 erreurs pour le groupe 2, soit une différence de seulement 2 erreurs entre les deux groupes de sujets. Si on observe l'évolution du nombre des erreurs commises par les sujets de chaque groupe, on constate que le groupe 1 commet un nombre constant d'erreurs - entre 20 et 17 erreurs - tandis que le groupe 2 commet de moins en moins d'erreurs - de 25 erreurs sur la première série de 30 images, jusqu'à 10 erreurs sur le dernière série (cf figure 5.13 page 107). Ce résultat peut s'expliquer de la manière suivante. Le groupe 1 effectue les tâches de repérage dans cette condition d'abord ; les sujets ne connaissent pas les couples (scène ; cible) qui leur sont présentés. En revanche, le groupe 2 effectue les tâches de repérage dans la condition visuelle après la condition multimodale ; les sujets connaissent les couples (scène ; cible) qui leur sont présentés et peuvent donc faire appel à leur mémoire pour retrouver les cibles avec plus de précision que les sujets du groupe 1.

Interprétation des résultats

Ces analyses ont mis en évidence l'influence de l'ordre de passation PV-PM *versus* PM-PV sur les stratégies adoptées par les sujets. En effet, le groupe 1 a adopté une stratégie qui privilégie

la précision par rapport à la rapidité de la sélection des cibles, contrairement au groupe 2 qui, lui, a adopté la stratégie inverse, i.e., privilégier la rapidité par rapport à la précision de la sélection des cibles. Pour mémoire, nous avons contrebalancé l'ordre des deux conditions expérimentales afin d'éviter l'effet de séquence[74]. Enfin, on peut faire une interprétation plus fine des courbes présentées sur les figures 5.12 et 5.13.

Concernant l'évolution des temps de sélection des cibles dans la condition visuelle, la courbe du groupe 1 représente l'évolution classique de l'apprentissage d'une nouvelle tâche, en raison du temps limité de la passation. On observe une amélioration avec la pratique sur les 60 premières images, puis une augmentation des temps moyens de sélections des cibles, due, probablement, à la fatigue et/ou au relâchement, des efforts et/ou de la motivation, car les sujets croient "savoir" comment effectuer au mieux la tâche. La courbe d'évolution du groupe 2 dans la condition visuelle représente moins clairement cet apprentissage : les sujets semblent avoir sous-estimé le niveau de difficulté de la tâche sur les 60 premières images (i.e., ils croient "savoir" la réaliser au mieux grâce à la condition multimodale précédente), ils semblent se ressaisir ensuite, mais la fatigue éprouvée interfère alors avec leurs efforts, car il s'agit de la seconde condition expérimentale. Cette "contre-performance" peut également provenir du délai nécessaire pour que les sujets perçoivent l'identité des images.

Concernant l'évolution des temps de sélection des cibles dans la condition multimodale, la courbe du groupe 1 est semblable à celle du groupe 2 dans la condition visuelle. Les hypothèses d'interprétation sont les mêmes (sous-estimation du niveau de complexité de la tâche sur les 60 premières images, puis fatigue). La courbe du groupe 2 dans la condition multimodale représente l'évolution de l'apprentissage d'une nouvelle tâche. L'apprentissage est moins marqué que pour le groupe 1 dans la condition visuelle et il n'y a pas d'augmentation des temps moyens de sélections des cibles sur les 60 dernières images, car la tâche est moins exigeante que dans la condition visuelle.

Concernant la précision des sélections de cibles dans la condition visuelle, il n'y a pas d'évolution des erreurs pour le groupe 1. La courbe du groupe 2 dans la condition visuelle représente l'évolution classique de l'apprentissage d'une nouvelle tâche facile : on n'observe ni fatigue, ni relâchement des erreurs.

Concernant la précision des sélections de cibles dans la condition multimodale, la courbe du groupe 1 met en évidence une augmentation importante du nombre d'erreurs sur 30 dernières images (par rapports aux images 60-90), due à la fatigue : les sujets du groupe 1 sont plus fatigués car "leur" condition visuelle est plus difficile que celle du groupe 2. La courbe d'évolution du nombre d'erreurs du groupe 2 dans la condition multimodale ressemble à celle d'un apprentissage d'une nouvelle tâche. Comment expliquer alors l'augmentation importante du nombre d'erreurs sur les images 30 à 60 ? On peut avancer l'argument du choix de stratégie adopté par ces sujets : privilégier la rapidité des sélections par rapport à leur précision.

Ces résultats sont intéressants car ils fournissent une preuve objective du caractère fatiguant de la tâche de repérage de cibles (cf. l'évolution finale sur les images 60-120 des temps moyens de sélection des cibles dans la condition visuelle et l'évolution finale sur les images 60-120 de

[74] cf. supra section 5.2.4 page 81.

la précision des sélections de cibles dans la condition multimodale). En outre, la tâche est plus exigeante dans l'ordre PV-PM que dans l'ordre PM-PV, ce qui prouve l'influence de l'ordre de passation sur les performances des sujets.

5.8 Influence de la structure des affichages sur les performances des sujets

Il convient de noter que cette partie de l'analyse se découpe en deux volets : un volet statistique qui porte sur la variable temps de sélection des cibles[75] et un volet qualitatif qui porte sur l'analyse des erreurs, d'une part, et sur l'analyse des clics sur le fond noir, d'autre part.

L'objectif que nous nous fixons dans cette partie de l'analyse est de valider les hypothèses de travail B, C, D et E[76]. Il s'agit donc de prouver que :

- dans la condition PM, il existe des différences significatives en termes de temps et de précision de sélection des cibles entre les quatre structures testées (hypothèse B) ;
- dans la condition PM, une structure émerge comme étant la plus efficace en terme de temps et de précision de sélection des cibles (hypothèse C) ;
- dans la condition PV, il existe des différences significatives en terme de temps et de précision de sélection des cibles entre les quatre structures testées (hypothèse D) ;
- dans la condition PV, une structure émerge comme étant la plus efficace en terme de temps et de précision de sélection des cibles (hypothèse E).

Afin de mettre en évidence l'éventuelle influence de l'organisation spatiale des scènes sur les performances des sujets, nous avons comparé la rapidité et la précision de la sélection des cibles entre les quatre structures testées[77], dans chacune des deux conditions expérimentales. Pour ce faire, les données recueillies lors de l'expérimentation ont été regroupées par condition et par structure des affichages, formant ainsi 8 ensembles de 720 données[78] : 1440 observations par structure, soit 720 observations par structure et par condition expérimentale.

5.8.1 Rapidité de la sélection des cibles

Les temps moyens de sélection et écarts types, en millisecondes, par structure et par condition expérimentale, sont présentés dans le tableau 5.5.

Dans la condition PV, le classement des structures par ordre décroissant d'efficacité en terme de temps de sélection des cibles est (cf. supra tableau 5.5 page 111) :

- la structure radiale où le temps moyen de sélection des cibles est de 5081 ms ;
- la structure aléatoire où le temps moyen de sélection des cibles est de 5626 ms ;
- la structure matricielle où le temps moyen de sélection des cibles est de 5738 ms ;

[75] cf. infra paragraphe 5.8.1 page 110.

[76] cf. supra 5.1.2 page72.

[77] Pour mémoire, les quatre structures sont la structure aléatoire, la structure elliptique, la structure matricielle et la structure radiale.

[78] Une donnée est un couple (temps de sélection ; précision de la sélection).

- la structure elliptique où le temps moyen de sélection des cibles est de 6250 ms.

Ces différences en terme des temps moyens de sélection des cibles sont statistiquement significatives entre les structures elliptique et radiale (t=3,64 ; p=0,0003), les structures matricielle et radiale (t=2,18 ; p=0,0296) et les structures elliptique et matricielle (t=1,25 ; p=0,0024). On observe une tendance entre les structures aléatoire et elliptique (t=-1,91 ; p=0,0568) et entre les structures aléatoire et radiale (t=1,82 ; p=0,0696). Il n'existe pas de différence significative entre les structures aléatoire et elliptique (t=-0,30 ; p=0,71).

Dans la condition PM, le classement des structures par ordre décroissant d'efficacité en terme de temps de sélection des cibles est (cf. supra tableau 5.5 page 111) :

- la structure radiale où le temps moyen de sélection des cibles est de 1640 ms ;
- la structure aléatoire où le temps moyen de sélection des cibles est de 1737 ms ;
- la structure matricielle où le temps moyen de sélection des cibles est de 1763 ms ;
- la structure elliptique où le temps moyen de sélection des cibles est de 1851 ms.

La différence, en terme des temps moyens de sélection des cibles, observée entre les structures elliptique et radiale est statistiquement significative (t=2,75 ; p=0,006). On observe une tendance entre les structures matricielle et radiale (t=1,49 ; p=0,1352) et entre les structures aléatoire et radiale (t=1,37 ; p=0,17). Il convient de noter également une légère tendance entre les structures aléatoire et elliptique (t=-1,40 ; p=0,16). Il n'existe de différence significative ni entre les structures aléatoire et matricielle, ni entre les structures elliptique et matricielle.

Variable : temps de sélection des cibles (ms)				
Structure	Condition	Moyenne (ms)	Écart type (ms)	Nombre d'observations
Aléatoire	PV	5626	5819	720
	PM	1737	1437	720
Elliptique	PV	6250	6585	720
	PM	1851	1633	720
Matricielle	PV	5738	5879	720
	PM	1763	1819	720
Radiale	PV	5081	5565	720
	PM	1640	1256	720

TAB. 5.5 – Comparaison de la rapidité des sélections entre les différentes structures.
La variable considérée est le temps de sélection des cibles en millisecondes. Il y a 720 observations par structure et par condition.

Pour résumer, dans chacune des deux conditions PM et PV, on observe des différences en terme de temps de réponse des sujets entre les différentes structures, avec :

- dans la condition PV, des différences statistiquement significatives entre les structures radiale et elliptique, les structures radiale et matricielle, et les structures elliptique et matricielle, mais également une tendance entre les structures radiale et aléatoire et les structures aléatoire et elliptique ;

- dans la condition PM, une différence statistiquement significative entre les structures radiale et elliptique, mais également une tendance statistique entre les structures radiale et matricielle, de même qu'entre les structures radiale et aléatoire.

5.8.2 Précision de la sélection des cibles

Le tableau 5.6 présente le nombre d'erreurs par structure et par condition expérimentale PV et PM, ainsi que la répartition du nombre des erreurs entre les structures en pourcentages.

Variable : précision de sélection des cibles				
Condition	Structure	Nombre d'erreurs	Répartition (%)	Nombre d'erreurs par condition
PV	Aléatoire	20	21	96
	Elliptique	25	26	
	Matricielle	27	28	
	Radiale	24	25	
PM	Aléatoire	13	27	49
	Elliptique	18	37	
	Matricielle	12	24	
	Radiale	6	12	

TAB. 5.6 – Comparaison du nombre d'erreurs observées entre les différentes structures. La variable considérée est la précision de la sélection des cibles. Il y a 720 observations par structure et par condition.

Globalement, le classement des structures, de la plus efficace à la moins efficace en terme de précision, est le suivant : structure radiale (21% des erreurs), structure aléatoire (23% des erreurs), structure matricielle (27% des erreurs) et structure elliptique (30% des erreurs).

Dans la condition visuelle, on observe une répartition équitable des erreurs entre les quatre structures. Le pourcentage d'erreurs observé dans chaque structure est proche de 25. Toutefois, la structure aléatoire se distingue puisque elle entraîne un nombre d'erreurs inférieur à celui observé pour les autres structures.

Le résultat le plus intéressant est celui observé dans la condition multimodale. En effet, entraînant trois fois moins d'erreurs que la structure elliptique et deux fois moins d'erreurs que les structures matricielle ou aléatoire, la structure radiale est dans la condition multimodale, celle qui permet la détection de cibles la plus précise. Autrement dit, les messages sonores d'indication spatiale absolue sont beaucoup plus efficaces dans des scènes à structure radiale que dans les autres organisations spatiales présentées aux sujets.

5.8.3 Clics sur le fond noir

Le tableau 5.7 présente le nombre de clics sur le fond noir, par structure et par condition expérimentale (PV et PM) ainsi que la répartition du nombre de ces clics entre les structures en pourcentages.

Variable : précision de sélection des cibles				
Condition	Structure	Clics sur fond noir	Répartition (%)	Clics sur fond noir par condition
PV	Aléatoire	9	20	45
	Elliptique	17	38	
	Matricielle	9	20	
	Radiale	10	22	
PM	Aléatoire	4	24	17
	Elliptique	6	35	
	Matricielle	4	24	
	Radiale	3	18	

TAB. 5.7 – Comparaison du nombre de clics sur le fond noir entre les différentes structures. La variable considérée est la précision de la sélection des cibles. Il y a 720 observations par structure et par condition.

Globalement, la structure elliptique est l'organisation spatiale la moins efficace en terme de nombre de clics sur le fond noir. En effet, on observe 37% de clics sur le fond noir, soit 23 clics en tout, dans cette structure *versus* seulement 21%, soit une moyenne de 13 clics par structure, dans les trois autres structures testées.

Dans la condition PV, la structure elliptique entraîne près de deux fois le nombre de clics sur le fond noir observés dans les trois autres organisations spatiales (cf. tableau 5.7 page113).

Dans la condition PM, on observe le même ratio par rapport à la structure radiale (6 clics sur le fond noir avec la structure elliptique *versus* 3 avec la structure radiale; cf. tableau 5.7 page113). On observe également une diminution de plus de 30% du nombre de clics sur le fond noir observés sur la structure elliptique par rapport à la structure matricielle et à la structure aléatoire (6 clics sur le fond noir avec la structure elliptique *versus* 4 avec la structure matricielle et la structure aléatoire; cf. tableau 5.7 page 113).

5.8.4 Conclusions

Les analyses quantitatives et qualitatives ont permis de mettre en évidence, dans la condition PM, des différences entre les structures, à la fois en terme de rapidité et de précision de la sélection des cibles. D'une part, concernant la rapidité de sélection des cibles, on observe jusqu'à 211 ms de différence entre la structure radiale et la structure elliptique et ce résultat est statistiquement significatif. En outre, bien que les messages multimodaux nivellent les temps moyens de réponse des sujets au sein des quatre structures, l'organisation spatiale radiale émerge comme permettant

les temps de sélection de cibles les plus rapides par rapport à la structure elliptique qui entraîne les temps moyens de réponse des sujets les plus lents. D'autre part, concernant la précision de sélection des cibles, on observe des différences importantes entre les structures en terme du nombre d'erreurs ; par exemple, le nombre d'erreurs est multiplié par 3 entre la structure radiale (6 erreurs) et la structure elliptique (18 erreurs) et par 2 entre la structure radiale (6 erreurs) et la structure matricielle ou la structure aléatoire (respectivement 12 et 13 erreurs). Ces résultats valident l'hypothèse de travail B.

Au vu des résultats précités, il apparaît également que la structure radiale émerge comme étant la plus efficace dans la condition PM, en termes de temps de réponse des sujets et de précision de la sélection des cibles. Autrement dit, l'apport des messages multimodaux est le plus important pour la structure radiale. Il convient de noter également que cet apport est le moins important pour la structure elliptique. Pour mémoire, l'apport des messages multimodaux a été montré dans la section 5.7 page 99, validant ainsi l'hypothèse A. Ces résultats valident également l'hypothèse de travail C.

Les différences constatées entre les structures dans la condition PM sont plus sensibles encore dans la condition PV. En effet, on observe jusqu'à plus d'une seconde d'écart en moyenne (1169 ms) entre la structure elliptique et la structure radiale, la plus petite différence étant observée entre la structure matricielle et la structure aléatoire (112 ms). Seule la différence de 112 ms entre structure aléatoire et structure matricielle n'est pas statistiquement significative, au contraire des autres différences entre les structures qui expriment, dans le pire des cas, une tendance. Quant au nombre d'erreurs observé entre les différentes structures, il avoisine les 25% pour chacune d'entre elles. Seule la structure aléatoire se démarque avec seulement 20 erreurs représentant 20% des erreurs commises dans la condition PV. En revanche, en ce qui concerne le nombre de clics sur le fond noir, on observe à nouveau des différences importantes entre la structure elliptique (17 sur le fond noir) et les structures aléatoire, matricielle et radiale (respectivement 9, 9 et 10 clics sur le fond noir). Le nombre de clics sur le fond noir est pratiquement multiplié par 2 dans la structure elliptique, par rapport aux autres structures. Ces résultats valident l'hypothèse D.

Au vu des résultats précités, il apparaît également que la structure radiale émerge comme étant la plus efficace, en termes de temps de réponse des sujets et de précision de la sélection des cibles dans la condition PV. Autrement dit, même en l'absence de messages multimodaux, cette structure reste l'organisation spatiale au sein de laquelle les sujets sont le plus rapide et le plus précis. Ce résultat valide l'hypothèse E.

5.8.5 Interprétation des résultats et discussion

La structure radiale permet aux sujets le repérage de cibles le plus efficace, à la fois en terme de leurs temps de réponse et en terme de précision des sélections, avec et sans assistance multimodale. Par opposition, la structure elliptique est la moins efficace pour les tâches de repérage visuel proposées. En ce qui concerne les préférences des sujets entre les quatre structures, l'analyse des questionnaires et des debriefings a révélé d'importantes contradictions :

- globalement, dans les questionnaires, les sujets préfèrent la structure elliptique à la structure radiale ; le classement des structures par ordre décroissant de préférence est structure elliptique, structure radiale, structure matricielle, et enfin, structure aléatoire ;
- cependant, au cours des debriefings, excepté trois sujets, tous ont déclaré que la structure radiale est leur préférée dans la condition multimodale ; les sujets se sont plaints de la structure elliptique qui, selon eux, était mal adaptée au découpage spatial associé aux indications orales ; ils ont déclaré qu'elle leur faisait perdre du temps car "il y avait plus d'images à regarder".

L'efficacité relative de la structure matricielle, mise en évidence par les performances moyennes des sujets dans cette structure, trouve peut-être une justification dans l'analyse détaillée des debriefings. Cette structure figure parmi les structures les plus utilisées pour l'organisation d'icones au sein d'environnement graphiques (cf. les systèmes d'exploitations Windows et UNIX), donc elle est familière aux sujets sollicités, pour la recherche d'items notamment. Néanmoins, son parcours reste lent, car les sujets semblent adopter une stratégie de parcours linéaire, d'après les debriefings. En outre, au moment où la structure apparaît, certains sujets n'ont pas su par où commencer leur exploration de la scène. Ces commentaires sur la structure matricielle semblent correspondre à ceux observés pour la navigation au sein de services en ligne [Perkins, 1995].

Enfin, l'avantage des structures aléatoires par rapport aux structures elliptique ou matricielle semble provenir du fait qu'elles induisent une exploration exhaustive des scènes pour trouver la cible. C'est, du moins, ce qui a été exprimé assez fréquemment au cours des debriefings.

5.9 Influence du niveau de difficulté des scènes sur les performances des sujets

Il convient de noter que cette partie de l'analyse se découpe en deux volets : un volet statistique qui porte sur la variable temps de sélection des cibles[79] et un volet qualitatif qui porte, à la fois, sur l'analyse des erreurs et sur l'analyse des clics sur le fond noir observés entre chacune des quatre structures testées[80]. Afin de mettre en évidence l'éventuelle influence du niveau de difficulté des scènes sur les performances des sujets, nous avons comparé la rapidité et la précision de la sélection des cibles entre les trois niveaux de difficulté, dans chacune des deux conditions expérimentales. Pour ce faire, les données recueillies lors de l'expérimentation ont été regroupées par condition et par niveau de difficulté, formant ainsi 8 ensembles de 720 données[81] : soit 1920 observations par niveau de difficulté, et 960 observations par niveau de difficulté et par condition expérimentale PV ou PM.

5.9.1 Rapidité de la sélection des cibles

Les temps de réponse des sujets, en fonction du niveau de difficulté des scènes et de la condition expérimentale, sont présentées dans le tableau 5.8 page 116.

[79] cf. infra paragraphe 5.9.1 page 115.
[80] cf. infra paragraphe 5.9.2 page 116 et 5.9.3 page 117.
[81] Une donnée est un couple (temps de sélection ; précision de la sélection).

On observe dans chacune des deux conditions des différences entre les niveaux de difficulté. Dans la condition visuelle, les différences observées entre les 3 niveaux de difficulté sont statistiquement significatives :

- la différence de 520 ms entre les niveaux 1 et 2 est statistiquement significative (t=-2,07 ; p=0,0369) ;
- la différence de 1744 ms entre les niveaux 1 et 3 est statistiquement hautement significative (t=-6,40 ; p<0,0001) ;
- la différence de 1224 ms entre les niveaux 2 et 3 est statistiquement hautement significative (t=-4,22 ; p<0,0001).

Variable : temps de sélection des cibles (ms)				
Niveau de difficulté	Condition	Moyenne (ms)	Écart type (ms)	Nombre d'observations
Facile	PV	4919	5011	960
	PM	1611	1387	960
Moyen	PV	5439	5879	960
	PM	1620	1272	960
Difficile	PV	6663	6801	960
	PM	2012	1893	960

TAB. 5.8 – Comparaison de la rapidité des sélections entre les différents niveaux de difficulté. La variable considérée est le temps de sélection des cibles en millisecondes. Il y a 960 observations par niveau de difficulté et par condition.

Dans la condition multimodale, la différence de 9 ms entre les niveaux de difficulté 1 et 2 n'est pas statistiquement significative. En revanche, la différence de 401 ms, respectivement 392 ms, observée entre les niveaux de difficulté 1 et 3, respectivement 2 et 3, est statistiquement hautement significative (t=-5,29 ; p<0,0001 respectivement t=-5,33 ; p<0,0001).

5.9.2 Précision de la sélection des cibles

Le nombre d'erreurs commises par les sujets, en fonction du niveau de difficulté des scènes et de la condition expérimentale, est présenté dans le tableau 5.9 page 117.

On observe dans chacune des deux conditions des différences entre les niveaux de difficulté. Dans la condition visuelle, le nombre d'erreurs est multiplié par 2 entre les niveaux 1 et 2. Il est presque multiplié par 3 entre les niveaux 1 et 3. Ces différences entre les niveaux sont réduites dans la condition multimodale. Il convient de noter toutefois, que dans chacune des deux conditions, le niveau de difficulté 3 entraîne près de 50% des erreurs : 48% des erreurs commises sur les scènes de niveau 3 dans la condition visuelle, 49% dans la condition multimodale.

Il convient de noter également que pour chaque niveau de difficulté, le nombre des erreurs est inférieur dans la condition PM par rapport à celui observé dans la condition PV. Par exemple, concernant le niveau 2 de difficulté, le nombre d'erreurs est divisé par plus de 2 dans la condition

Variable : précision de sélection des cibles				
Condition	Niveau de difficulté	Nombre d'erreurs	Répartition (%)	Nombre d'erreurs par condition
PV	Facile	17	18	96
	Moyen	33	34	
	Difficile	46	48	
PM	Facile	10	20	49
	Moyen	15	31	
	Difficile	24	49	

TAB. 5.9 – Comparaison de la précision des sélections entre les différents niveaux de difficulté. La variable considérée est le nombre d'erreurs. Il y a 960 observations par niveau de difficulté et par condition.

multimodale : 33 erreurs dans PV *versus* 15 erreurs dans PM. On approche ce résultat dans les deux autres niveaux de difficulté (cf. tableau 5.9 page 117).

5.9.3 Clics sur le fond noir

Le nombre de clics sur le fond noir, en fonction du niveau de difficulté des scènes et de la condition expérimentale, est présenté dans le tableau 5.10 page 118.

Contre toute attente, les résultats observés concernant les temps moyens de sélection des cibles et le nombre d'erreurs effectives, ne se répètent pas dans l'analyse des clics sur le fond noir. C'est le niveau de difficulté 2 qui entraîne le plus grand nombre de clics sur le fond noir (26 clics ; PV et PM confondues), suivi du niveau 3 (21 clics ; PV et PM confondues), puis du niveau 1 (11 clics ; PV et PM confondues). Ce résultat global se répète dans chacune des conditions PV et PM.

Le résultat observé entre les niveaux 2 et 3 de complexité des scènes peut s'expliquer de la manière suivante, en comparant les résultats présentés dans les tableaux 5.9 et 5.10 : les sujets adoptent une stratégie en fonction du niveau de difficulté de la scène. Sur les scènes de difficulté 2, lorsque les sujets ne parviennent pas à identifier rapidement la cible, ils cliquent plus volontiers dans le noir car aucune photographie ne "ressemble assez" à la cible. Sur les scènes de difficulté 3, il "tentent" plus souvent leur chance, car le nombre de non cibles visuellement proches de la cible sont plus nombreux.

5.9.4 Conclusions

Les résultats relatifs aux temps de réponse et à la précision, ou nombre d'erreurs effectives, des sujets confirment les hypothèses formulées sur les caractéristiques des collections d'images, mais aussi sur les niveaux de difficulté des scènes[82]. Les collections de photographies hétérogènes

[82] cf. supra "Caractéristiques des collections d'images", paragraphe 5.1.2 page 72).

Variable : précision de sélection des cibles				
Condition	Niveau de difficulté	Clics sur fond noir	Répartition (%)	Nombre total de clics dans le noir
PV	Facile	11	24	45
	Moyen	19	42	
	Difficile	15	33	
PM	Facile	4	24	17
	Moyen	7	41	
	Matricielle	6	35	

TAB. 5.10 – Comparaison du nombre de clics sur le fond noir observés entre les différents niveaux de difficulté.
La variable considérée est la précision de la sélection des cibles en terme du nombre de clics sur le fond noir. Il y a 960 observations par niveau de difficulté et par condition.

permettent aux sujets de repérer plus facilement les cibles par rapport aux collections homogènes qui augmentent les temps de recherche et accroissent le nombre des erreurs. De la même façon, la tâche de repérage est plus simple pour les sujets lorsque les collections de photographies ne présentent pas, en plus d'être homogènes, une complexité de détail au sein de chaque pastille.

Autrement dit, les items non cibles peuvent être rejetés plus facilement s'ils ne partagent pas ou peu de propriétés avec la cible.

Par ailleurs, ce résultat renforce davantage notre conclusion concernant l'apport des messages sonores de localisation spatiale des cibles pour leur repérage[83]. En effet, ce résultat démontre que plus la tâche de repérage de cibles est complexe en terme du niveau de difficulté des scènes, plus les présentations multimodales des cibles s'avèrent efficaces. Statistiquement, en terme du temps moyen de réponse, ce résultat s'exprime de la manière suivante :

- entre les niveaux 1 et 2, la différence observée entre les conditions PV et PM est de 512 ms ; cette différence est statistiquement significative (t=-2,07 ; p=0,0382) ;
- entre les niveaux 1 et 3, la différence observée entre les conditions PV et PM est de 1344 ms ; cette différence est statistiquement significative (t=-4,92 ; p<0,0001) ;
- entre les niveaux 2 et 3, la différence observée entre les conditions PV et PM est de 831 ms ; cette différence est statistiquement significative (t=-2,90 ; p=0,0037).

Autrement dit, l'utilisation ou non de présentations multimodales pour faciliter la tâche de repérage visuel de cibles doit dépendre du niveau de complexité des scènes. Sur des scènes de complexité peu élevée (nombre d'éléments constitutifs peu élevé, hétérogénéité, faible complexité de détail), les messages sonores d'indication à caractère spatial constituent un apport moindre par rapport à leur puissance au sein de scènes complexes (nombre d'éléments constitutifs très élevé, homogénéité, complexité de détail élevée). Ce résultat correspond à celui exprimé dans [Althoff *et al.*, 2001].

[83] cf. supra 5.7 page 99.

5.10 Analyses complémentaires

5.10.1 Paysages *versus* objets complexes

Nous avons évalué également l'influence du contenu des photographies sur les performances des sujets. Nous avons regroupé les performances des sujets par type de photographies contenues dans la scène (paysages *versus* objets complexes) formant ainsi deux groupes de 2880 observations. Les résultats sont présentés dans le tableau 5.11.

Type d'affichage	Niveau d'observations	Temps moyen (ms)	Écart type (ms)	Taux d'erreurs brutes (%)
Objets	2880	3235	4175	3,84
Paysages	2880	4186	5297	5,53

TAB. 5.11 – Comparaison objets complexes *versus* paysages.
Les variables considérées sont le temps moyen et la précision de la sélection des cibles. Il y a 2880 observations par type de photographies.

On observe des différences en termes de temps moyen et de précision de la sélection des cibles. Ces différences sont statistiquement hautement significatives, à la fois, concernant les temps moyens de sélection des cibles ($t=-7.57$; $p<0,0001$), mais aussi concernant la précision de sélection des cibles (-3,04 ; $p<0,0001$).

Ces résultats montrent que le repérage visuel de cibles, quel que soit le mode de présentation de celle-ci, est plus facile pour les sujets lorsque la cible est un objet complexe plutôt qu'un paysage : les sujets repèrent plus vite la cible et commettent moins d'erreurs dans ce cas. Nous avons établi que la plupart des sujets adoptent la stratégie suivante : description mentale d'un détail, jugé pertinent *a priori* pour la discrimination de la cible. Le manque de tels indices au sein des photographies de paysages les empêche d'appliquer cette stratégie. Des paysages neutres, tels que ceux choisis pour élaborer les collections de niveau de difficulté 2 et 3, redent difficile leur caractérisation verbale. En outre, la recherche d'un paysage peu familier ou inconnu, s'avère plus difficile que la recherche d'un objet visuellement connu, familier, voire déjà utilisé. Ce résultat semble compatible avec le modèle temporel "coarse to fine" de perception des fréquences spatiales [Huges *et al.*, 1996].

5.10.2 Répartition des erreurs en fonction de la position de la cible

La répartition des erreurs en fonction de la position de la cible, selon les 9 zones définies par la figure 5.3 page 76, est présentée dans le tableau 5.12 page 120.

Les zones au sein desquelles on observe le plus d'erreurs sont : la zone "en bas" avec 26% des erreurs, la zone "en haut" avec 17% des erreurs et la zone "à gauche" avec 16% des erreurs. On observe moins d'erreurs dans les zones "en haut à gauche", "en bas à droite", "à droite" et "en bas à gauche", avec respectivement 10%, 8%, 7% et 7%. Les deux zones au sein desquelles on observe

HG (3976 ms ; 10%) 576 observations	H (3366 ms ; 17%) 864 observations	HD (3730 ms ; 4%) 480 observations
G (3965 ms ; 16%) 672 observations	C (2625 ms ; 5%) 624 observations	D (4061 ms ; 7%) 480 observations
BG (3464 ms ; 7%) 528 observations	B (4053 ms ; 26%) 864 observations	BD (4168 ms ; 8%) 672 observations

TAB. 5.12 – Analyse des erreurs en fonction de la position des cibles (conditions PV et PM confondues).
Les variables considérées sont le temps moyen et la précision de la sélection des cibles. Le nombre d'observations par zone est variable, compte tenu des contraintes spatiales imposées par les structures.

le moins d'erreurs sont la zone "au centre" avec 5% des erreurs et la zone "en haut à droite" avec 4% des erreurs.

Ce résultat peut s'expliquer de la manière suivante : l'exploration des scènes commence au centre en raison du repositionnement de la souris au centre de l'écran imposé par le bouton "OK", donc les sujets atteignent plus facilement la cible que dans les autres zones de l'écran. Ce résultat aurait peut-être été différent en l'absence du bouton "OK", ou si le bouton "OK" avait été placé ailleurs dans la scène ? En outre, le parcours oculaire lors de l'exploration d'une scène commence-t'il par le centre de la scène, pour se poursuivre vers les coins et s'achever au bord haut, bas, gauche ou droite ?

5.11 Conclusions générales

Chacune des hypothèses A, B, C, D et E a été validée. Dans un premier temps, nous avons mis en évidence l'apport des messages multimodaux pour le repérage visuel de cibles, en termes de temps moyen de sélection des cibles et de précision de la sélection des cibles. Nous avons démontré, en outre, que cet apport varie en fonction du niveau de difficulté de la scène. En d'autres termes, plus la tâche est complexe, par la nature même de la scène (homogénéité de la collection d'items) ou par le manque de familiarité des sujets avec la cible, plus les messages multimodaux s'avèrent efficaces. Tous ces résultats valident l'hypothèse A.

En outre, par l'analyse détaillée des commentaires recueillis lors des debriefings, il apparaît que les sujets suivent les indications spatiales orales incluses dans les présentations multimodales pour explorer visuellement les scènes. Ces indications leur permettent de réduire efficacement la zone de recherche. Par suite, en n'imposant pas un parcours systématique de l'organisation ou structure spatiale de la scène, et en réduisant le nombre de confusions possibles entre la cible et les non-cibles, les présentations multimodales permettent aux sujets de réduire leurs temps de réponse et d'être plus précis.

La stratégie de recherche est différente si le mode de présentation de la cible est uniquement visuel. Dans la condition visuelle, les sujets se construisent une représentation mentale de la cible, d'un détail qu'ils espèrent être discriminant. Au bout de 10 scènes, ils ont observé la "cohérence thématique" de chaque scène : trouver le détail pertinent devient plus aisé. Au moment où la cible apparaît, il parcourent la scène en suivant sa structure visuelle. Dans le cas des structures aléatoires, ils se dirigent vers les zones écran les plus denses.

En outre, les sujets ne semblent pas gênés par les messages sonores et les acceptent comme une "assistance à la tâche". Les messages sonores ne semblent pas augmenter leur charge cognitive : au contraire, les sujets ressentent moins de fatigue lors de la condition multimodale que lors de la condition visuelle. La tâche, dans cette condition, est décrite comme plus facile et plus rapide. Nous avons observé également que l'ordre PV-PM est plus fatiguant que l'ordre PM-PV. La fusion des deux types d'informations ne semble pas surcharger la mémoire de travail. Enfin, ces commentaires des sujets recueillis à l'issue de l'expérimentation correspondent à ceux présentés dans [Pelz *et al.*, 2001] sur la création d'une structure mentale de la tâche.

Dans un deuxième temps, nous avons mis en évidence des différences entre les structures pour chaque condition expérimentale. Entre autres, il apparaît que la structure radiale émerge comme étant la plus efficace, en termes de temps de réponse des sujets et de précision de la sélection des cibles dans les deux conditions expérimentales. Les résultats présentés au paragraphe 5.8.4 valident les hypothèses de travail B, C, D et E.

Plus précisément, la structure radiale permet aux sujets le repérage de cibles le plus efficace, à la fois en terme de leurs temps de réponse et en terme de précision des sélections, avec et sans assistance multimodale. Par opposition, la structure elliptique est la moins efficace pour les tâches de repérage visuel proposées. L'efficacité relative de la structure matricielle, mise en évidence par les performances moyennes des sujets dans cette structure, peut se justifier de la manière suivante : la structure matricielle est celle à laquelle les sujets sont le plus habitués (cf. la recherche d'items au sein des environnement Windows et UNIX). Le parcours, vraisemblablement linéaire de la structure matricielle, permet un repérage précis des cibles, mais accompagné d'un temps de recherche très long. Pour être efficace au sein de la structure matricielle, la recherche doit porter sur un item dont la position dans la matrice est connue, ce qui n'est pas le cas dans notre expérience. Enfin, l'avantage des structures aléatoires par rapport aux structures elliptique ou matricielle semble provenir du fait qu'elle permettent une exploration exhaustive rapide des scènes pour trouver la cible, la recherche progressant d'une zone de forte densité spatiale à l'autre, en fonction des regroupements d'items créés par leur répartition aléatoire dans l'affichage. C'est, du moins, ce qui a été exprimé assez fréquemment au cours des debriefings.

Dans ce chapitre, nous avons établi l'influence des messages oraux d'assistance à la recherche au sein de visualisations 2D interactives. Nous avons également mis en évidence l'influence de l'organisation visuelle des scènes pour la recherche visuelle de cibles. Des stratégies d'exploration visuelle différentes semblent être adoptées par les sujets en fonction de la structure spatiale même de la scène, selon leurs propres commentaires qui demandent à être confirmés par des mesures objectives avant d'être considérés comme valides. D'après leurs commentaires, il semblerait que :

- le parcours oculaire des structures matricielles soit linéaire : la recherche est donc lente, mais précise ;
- le parcours oculaire des structures radiales parte du centre vers les bords : la recherche est par conséquent très rapide et très précise ;
- le parcours oculaire des structures elliptiques soit circulaire : son efficacité en termes du temps et de la précision des sélections des cibles peut-être discutable ;
- le parcours oculaire des structures aléatoire soit dirigé par les zones denses de l'affichage : une recherche rapide et précise est donc possible.

Peut-on vérifier ces hypothèses ? Est-il possible de mettre en évidence des comportements différents des sujets, en terme de trajectoires oculaires, au sein des différentes structures ? Peut-on parler de tâche différente selon l'affichage proposé au sujet ? C'est l'objet du prochain chapitre.

Chapitre 6

Troisième étude

La troisième étude, comme les précédentes, est conçue dans le cadre d'une approche expérimentale. Elle a pour objectif d'analyser les parcours oculaires adoptés par les sujets pour la recherche visuelle d'items au sein d'affichages 2D interactifs, et de montrer si, oui ou non, ces parcours dépendent de l'organisation spatiale des affichages. En d'autres termes, quelle est l'influence de l'organisation spatiale des affichages sur les parcours oculaires lors de la recherche visuelle de cibles ? L'organisation spatiale peut-elle être perçue comme une forme de guidage visuel à elle seule ? Certaines structures des affichages obligent-elles à des parcours oculaires plus lents, moins précis ?

Telles sont les interrogations suggérées par les résultats quantitatifs et qualitatifs observés lors de la deuxième étude, particulièrement en l'absence de tels messages. En effet, en l'absence de messages sonores d'indication spatiale, on observe des différences significatives entre les quatre structures 2D testées, en terme de temps de sélection des cibles. C'est pour apporter des éléments de réponse à ces questions que nous avons mené une troisième étude expérimentale, en nous appuyant sur des données recueillies à l'aide d'un oculomètre (*eye-tracker*). Non seulement les performances des sujets sont mesurées en termes de temps et de précision de la sélection à la souris des cibles, mais en plus, leurs fixations oculaires sont recueillies grâce à l'oculomètre pour chaque scène de l'expérimentation.

Ce chapitre suit le même plan que celui adopté dans le chapitre précédent. Nous décrivons tout d'abord la méthodologie utilisée dans la conception du plan expérimental. Puis, nous décrivons le protocole. Enfin, nous présentons l'analyse des données en deux volets : le premier portant sur les performances des sujets, le deuxième sur les stratégies de recherche visuelle.

6.1 Méthodologie

Les conclusions relatives à l'organisation spatiale des affichages présentées dans le chapitre consacré à la deuxième étude expérimentale sont les suivantes [84] :

[84] cf. supra 5.11 page 120.

- en l'absence de messages sonores d'indication spatiale, on observe des différences significatives entre les quatre structures 2D testées, en terme de temps de sélection des cibles ;
- la structure radiale émerge par rapport aux structures en ellipse, en matrice ou aléatoire, comme étant la plus efficace en terme de temps de sélection des cibles.

Le parcours oculaire d'une scène graphique semble donc être dépendant de l'organisation spatiale des éléments la composant. Compte tenu de ces résultats, nous avons choisi de centrer cette troisième étude expérimentale sur l'analyse des parcours oculaires effectués par les sujets lors de la recherche visuelle de cibles en nous basant sur l'analyse des fixations oculaires sur chaque scène, les fixations oculaires étant recueillies à l'aide d'un eye-tracker.

Dans la suite de cette section, nous présentons dans un premier temps, la méthodologie adoptée. Nous énonçons, dans un deuxième temps, les objectifs détaillés du travail.

6.1.1 Présentation générale

Dans cette étude, nous ne conservons que le mode de présentation visuel des cibles. Il est inutile de maintenir le mode de présentation multimodal des cibles, car l'étude est centrée sur l'analyse des stratégies d'exploration visuelle adoptées par les sujets pour le repérage visuel de cibles, en fonction de l'organisation spatiale des affichages. L'étude précédente a montré, non seulement que les messages oraux contenus dans les présentations multimodales constituent une assistance à la tâche, mais aussi que, dans la condition multimodale (PM), les sujets suivent les indications orales à caractère spatial sans tenir compte de la structure ; c'est ce qui a été le plus fréquemment exprimé spontanément lors des debriefings[85].

Comme dans l'étude précédente, les scènes présentées aux sujets peuvent prendre la structure matricielle, la structure radiale, la structure elliptique, ou encore la structure aléatoire. En outre, nous avons conservé la même caractérisation du matériel visuel[86] :

- chaque scène est composée de N photographies de même forme et de même taille, portant toutes sur le même thème et formant ainsi une collection, soit d'objets, soit de paysages ;
- chaque collection est caractérisée par un niveau de difficulté établi en fonction de l'homogénéité/hétérogénéité et du niveau de détail des photographies la composant ;
- chaque cible est caractérisée par sa position et sa saillance dans la scène.

Par ailleurs, comme dans le cadre du projet pluridisciplinaire Micromégas [Micromégas, 2003] qui porte sur la conception et l'évaluation d'approches multiéchelles pour la navigation dans les masses des données familières[87], nous distinguons le repérage visuel de cibles familières, du repérage visuel de cibles non familières. Dans la suite du travail, nous parlerons de cibles familières, respectivement non familières, pour désigner des cibles déjà connues visuellement, respectivement inconnues visuellement, par les sujets, définissant ainsi deux conditions expérimentales :

- la condition notée F où les sujets effectuent les tâches de repérage visuel au sein de couples (scène ; cible) qui leur sont familiers, car ils les ont vus deux fois déjà lors de la deuxième étude ;

[85] cf. supra 5.11 page 120.
[86] cf. supra 5.1.3 page 73.
[87] cf. supra 2.4.2 page 14.

- la condition notée NF où les sujets effectuent les tâches de repérage visuel au sein de couples (scène ; cible) qui ne leur sont pas familiers, car ils les découvrent au cours de cette troisième expérimentation.

L'objectif est de déterminer si, oui ou non, il existe une différence entre la recherche d'un item au sein d'une collection d'items, cette recherche ayant déjà été effectuée antérieurement, et la recherche dans une scène d'un item présenté visuellement à l'écran pendant quelques secondes.

6.1.2 Objectifs de l'étude

L'expérience réalisée porte sur l'influence des structures spatiales des affichages sur les stratégies d'exploration visuelle dans un contexte de repérage de cibles. Elle a pour objectif de fournir des éléments de réponse aux interrogations suivantes :

- la structure des affichages agit-elle comme une forme de guidage visuel ? Les structures matricielles sont-elles parcourues par balayage horizontal et/ou vertical, les structures elliptiques de façon circulaire, les structures radiales suivant leurs rayons ?
- dans le cas particulier des structures aléatoires, le regard est-il guidé vers les zones de l'affichage denses en informations ? Ces zones plus "informatives" peuvent être définies en terme de la position ou arrangement des items les uns par rapport aux autres (cf. l'ouvrage sur la photographie [Aumont, 2001]) ou encore, en terme de la densité relative des éléments qui composent l'affichage (cf. les travaux sur la recherche d'informations au sein de visualisations arborescentes [Pirolli *et al.*, 2000]) ;
- pour un même sujet, les stratégies d'exploration visuelle des affichages varient-elles en fonction de l'organisation spatiale des scènes ou, à l'inverse, ceux-ci adoptent-ils la même stratégie d'exploration visuelle pour toutes les structures spatiales ?
- pour une même structure, peut-on comparer les éventuelles différentes stratégies adoptées par les sujets, et par suite, établir pour chacune des structures, une typologie des parcours individuels ?

Ce sont autant de questions auxquelles nous allons tenter de répondre, grâce à l'analyse des parcours oculaires lors de l'exploration visuelle des scènes pour la tâche de repérage de cibles, à l'aide d'un eye-tracker dont s'est doté l'équipe MErLIn.

6.2 Protocole expérimental

6.2.1 Généralités

Le protocole expérimental est semblable à celui adopté pour les deux études précédentes. Le scénario d'interaction est le même [88] avec :

- la présentation visuelle de la cible au centre de l'écran pendant 3 secondes ;
- le repositionnement de la souris au centre de l'écran grâce à la sélection obligatoire du bouton OK ;

[88] cf. supra paragraphes 4.2.1 page 33 et 5.2.1 page 79.

- l'affichage de la scène ;
- la sélection à la souris d'un item dans la scène.

Comme dans les deux études précédentes, la taille de la cible ne varie pas entre sa présentation visuelle et son affichage au sein de la scène. Une fois la scène affichée, les sujets ne disposent que d'un seul clic pour sélectionner la cible. Au clic, on passe à la présentation visuelle de la prochaine cible.

Les variables libres sont, comme dans la deuxième étude, la structure spatiale de la scène, le type de contenu des photographies, ainsi que le niveau de difficulté des collections de photographies et la position de la cible dans la scène[89]. Les variables liées sont le temps et la précision de la sélection de la cible.

Par ailleurs, un logiciel de rejeu des parcours oculaires des affichages graphiques a été développé par Jérôme Simonin, doctorant dans l'équipe MErLIn. Cet outil nous permet de recueillir, pour chacun des sujets et pour chaque scène, le temps et le nombre de fixations oculaires d'un point A à un point B de la scène, la durée des fixations oculaires, la durée des saccades, la distance parcourue sur la scène. Le logiciel fournit, en outre, pour chacun des sujets, la représentation graphique du parcours oculaire effectué sur chaque scène lors de l'expérimentation.

Pour atteindre les objectifs d'analyse que nous avons fixés au paragraphe 6.1.2 page 125, l'analyse des données comprend deux phases :

- l'analyse quantitative statistique des données brutes telles que temps et précision des sélections, temps de parcours entre deux fixations particulières et, sur de tels intervalles, nombre de fixations, durée des fixations et des saccades, distance parcourue par le regard ;
- l'analyse qualitative détaillée des parcours oculaires sur les scènes, associée à une analyse quantitative inter-sujets.

6.2.2 Conditions expérimentales

Les conditions expérimentales sont au nombre de deux : la condition de repérage de cibles familières (F) et la condition de repérage de cibles non familières (NF). Pour réaliser ces deux conditions, nous avons choisi les participants parmi les sujets de la deuxième étude. Dans la condition F, les sujets effectuent les tâches de repérage sur 60 scènes familières, i.e., les couples (scène + cible) dans cette condition sont les mêmes que dans l'étude précédente. Dans la condition NF, les sujets effectuent les tâches de repérage sur 60 scènes non familières, i.e., les couples (scène + cible) dans cette condition ne sont pas les mêmes que dans l'étude précédente : nous avons repris les mêmes collections de photographies que pour l'étude précédente, en modifiant les positions des photographies au sein de la scène et en choisissant une cible n'appartenant pas à la collection initiale, i.e., celle présentée lors de la deuxième expérimentation. Chaque scène contient 30 photographies (N=30).

Pour maintenir la validité interne de cette étude, relativement à la précédente, nous avons contrebalancé l'ordre des conditions expérimentales, i.e., les sujets effectuent les tâches de repérage visuel de cibles, soit dans l'ordre F puis NF, soit dans l'ordre NF puis F. Comme dans

[89] cf. supra tableau 5.1 page 80.

l'étude précédente, les sujets ont été affectés de façon aléatoire à l'ordre de passation. Enfin, tous les sujets effectuent les tâches de repérage visuel de cibles sur les mêmes images, mais l'ordre d'apparition de chaque image varie aléatoirement d'un sujet à l'autre.

6.2.3 Choix des sujets

Nous avons sélectionné 10 sujets ayant participé à la deuxième expérimentation. Pour ce faire, nous avons, dans un premier temps, établi une classification hiérarchique[90] (*clustering*) des 24 sujets en fonction de leurs performances en termes du temps et de la précision des sélection de cibles. Cette classification a été réalisée par la méthode des centres mobiles. À noter que la précision n'influe pas sur la classification hiérarchique, en raison du faible nombre d'erreurs observé lors de la passation.

Nous avons, dans un deuxième temps, formé plusieurs groupes de sujets :

- le groupe 1 comptant 13 sujets[91], assez rapides dans les deux conditions PV (temps moyens de réponse compris entre 4 secondes et 6 secondes) et PM (temps moyens de réponse compris entre 1 et 2,25 secondes) ;
- le groupe 2 comptant 2 sujets, très rapides dans les deux conditions PV (temps moyens de réponse inférieurs à 4 secondes) et PM (temps moyens de réponse inférieurs à 1,5 secondes) ;
- le groupe 3 comptant 3 sujets, lents dans la condition PV (temps moyens de réponse compris entre 6 et 8 secondes) et rapides dans la condition PM (temps moyens de réponse compris entre 1,5 et 1,75 secondes) ;
- le groupe 4 comptant 2 sujets, moyens dans la condition PV (temps moyens de réponse légèrement inférieurs à 6 secondes) et lents dans la condition PM (temps moyens de réponse compris entre 2,25 et 2,75 secondes) ;
- le groupe 5 comptant 3 sujets, très lents dans la condition PV (temps moyens de réponse compris entre 8 et 10 secondes) et lents dans la condition PM (temps moyens de réponse compris entre 1,9 et 2,4 secondes) ;
- le groupe 6 comptant un seul sujet (atypique), très lent dans les deux conditions PV (temps moyen de réponse égal à 11,4 secondes) et PM (temps moyen de réponse égal à 2,9 secondes).

Nous avons exclu de la troisième étude le sujet du groupe 6, en raison du caractère atypique de ses performances. Puis, nous avons sélectionné un sujet dans chacun des groupes 2, 3, 4 et 5, et 6 sujets du groupe 1. Pour mémoire, l'objectif de cette étude est d'analyser les stratégies adoptées par les sujets pour la tâche de repérage visuel de cibles. Ainsi, sélectionner 10 sujets aux profils très différents devrait nous permettre de couvrir un maximum de stratégies d'exploration visuelle. En effet, la grande variabilité interindividuelle existant entre ces 10 sujets devrait pouvoir s'expliquer en fonction des stratégies mises en place pour le repérage de cibles, en termes de distance parcourue dans la scène, de nombre de fixations, ou encore, de parcours oculaire.

[90] cf. figure 6.1 page 128. Cette classification a été réalisée sous SAS avec la collaboration de François-Xavier Jollois, Maître de Conférence à l'Université René Descartes (Paris V).

[91] Il convient de noter que ce groupe peut encore être subdivisé en plusieurs sous-groupes (cf. figure 6.1 page 128).

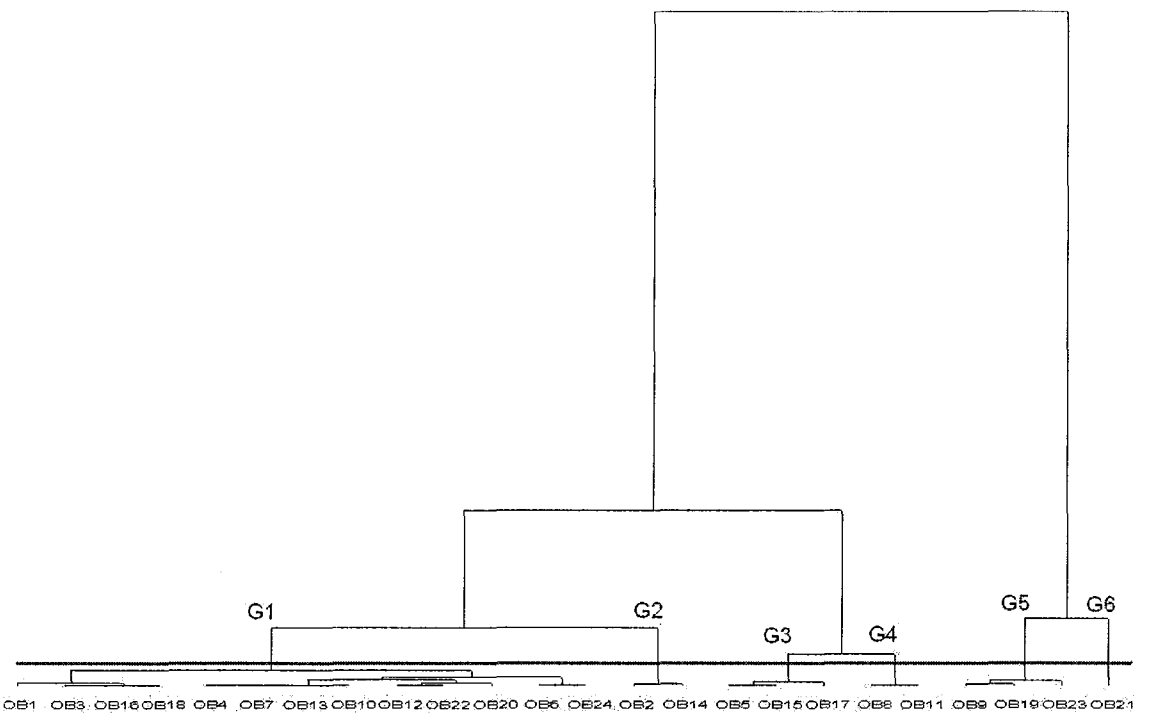


FIG. 6.1 – Classification hiérarchique ; 24 sujets ; temps et précision des sélections de cibles. Cette classification hiérarchique a été réalisée sous SAS par la méthodes des centres mobiles en tenant compte des deux conditions expérimentales PV et PM, des temps de sélection des cibles ainsi que de la précision des sélections. Les notations OB1 à OB24 désignent les 24 sujets dans l'ordre croissant. Le découpage est matérialisé par la droite horizontale rouge. Le groupe 1 est formé de l'ensemble des sujets {1,3,16,18,4,7,13,10,12,22,20,6,24}, le groupe 2 de l'ensemble {2,14}, le groupe 3 de l'ensemble {5,15,17}, le groupe 4 de l'ensemble {8,11}, le groupe 5 de l'ensemble {9,19,23} et le groupe 6 du singleton {21}.

6.2.4 Déroulement

Le déroulement d'une passation est le même pour tous les sujets, à l'exception de l'ordre de passation des conditions expérimentales F et NF. Chaque sujet lit d'abord la consigne, puis l'expérimentateur procède au calibrage de l'eye-tracker. Le sujet effectue alors les tâches de repérage sur 10 scènes d'entraînement. Enfin, il réalise les tâches de repérage dans chacune des deux conditions[92]. La durée globale d'une passation est d'une heure environ.

6.3 Analyse quantitative des données

L'analyse quantitative des données a été réalisée selon le plan suivant :
- le premier volet correspond à la validation du protocole expérimental par le biais d'analyses statistiques des performances des sujets en termes de temps et, lorsque le nombre d'observations le permet, de précision de la sélection des cibles (cf. infra 6.3.1) ;
- le deuxième volet porte sur l'analyse statistique des données recueillies à l'aide de l'oculomètre, notamment les temps de parcours des scènes, le nombre de fixations par scène, leur durée, ou encore la distance parcourue (cf. infra 6.3.2 page 132).

Les statistiques ont été réalisées sous SAS en collaboration avec François-Xavier Jollois, Maître de Conférence à l'Université René Descartes (Paris V).

6.3.1 Validation du protocole expérimental

Nous avons, dans un premier temps, vérifié la validité du protocole expérimental. Plus précisément, nous avons vérifié s'il existe des différences de performance significatives entre les deux conditions expérimentales (F et NF), entre les différentes organisations spatiales des affichages (structures aléatoire, matricielle, elliptique et radiale), entre les différents niveaux de difficulté des scènes (niveaux 1, 2 ou 3) et enfin, entre les types d'affichages (objets ou paysages). Les résultats de ces analyses sont présentés ci-dessous.

Conditions expérimentales F et NF

Pour mémoire, la condition F définit le repérage de cibles familières, la condition NF le repérage de cibles non familières[93]. Il y a 600 observations par condition (10 sujets, 60 scènes par condition). La moyenne des temps observés dans la condition F est de 4696 ms. La moyenne des temps de sélection observés dans la condition NF est de 4204 ms. Cette différence est significative (t=2,00 ; p=0,0455). Globalement, les sujets sont meilleurs dans la condition NF. Nous avons été surpris par ce résultat qui semble pouvoir s'expliquer de la façon suivante : dans la condition F, les sujets savaient qu'il s'agissait des mêmes tâches de repérage que lors de la deuxième expérimentation (mêmes scènes, mêmes cibles). Ils semblent avoir été déroutés par cette information, car bien qu'ils aient reconnu certaines cibles, ils ne se souvenaient pas de leur position dans la scène. C'est ce que la plupart ont déclaré spontanément à la fin de la passation.

[92] 60 scènes par condition, soit 120 scènes en tout.
[93] cf. supra 6.2.2 page 126.

Bien que surprenant, ce résultat semble pouvoir s'expliquer de la manière suivante : d'après [Ware, 2004] page 304, pour être retenues, les images doivent être porteuses de sens et susceptibles d'être incorporées au sein d'un cadre cognitif. Ce qui signifie qu'une image ne peut pas être reconnue si l'information qu'elle véhicule est nouvelle ou représentée de manière abstraite ou hors-contexte. Certes, les thèmes utilisés au sein de chaque collection sont courants (animaux, véhicules, montagnes, ...). En revanche, le contenu de chaque scène, et notamment la cible, est représenté de manière inhabituelle (structure spatiale des affichages) et hors-contexte (sur un fond d'écran neutre). C'est pourquoi dans la condition dite familière (F), i.e., où chaque couple (scène + cible) a déjà été présenté deux fois à chaque sujet (cf. expérimentation précédente : conditions PV et PM), même si les cibles ont quelquefois été reconnues, les sujets n'ont pas pu faire appel à leur mémoire pour retrouver la position de la cible dans la scène. La différence des temps moyens de sélection de la cible entre les deux conditions F (4696 ms) et NF (4204 ms) semble pouvoir s'expliquer par la limite de la mémorisation à long terme des tâches visuelles, les deux études ayant été réalisées à trois mois d'intervalle. Cependant, des études spécifiques sont nécessaires pour pouvoir expliquer pourquoi les cibles ont été bien mémorisées tandis que leurs positions dans les scènes ne l'ont pas été.

Organisation spatiale des affichages

L'analyse globale des performances des sujets en fonction de la structure spatiale des affichages a révélé qu'il n'existe aucune différence significative entre les structures, i.e., toutes conditions et tous sujets confondus.

L'absence de différence statistiquement significative observée entre les structures de manière globale semble provenir de la grande variabilité interindividuelle entre les sujets. En effet, nous avons choisi des sujets issus de groupes différents (cf. la section portant sur le clustering 6.2.3 page 127). Cette variabilité interindividuelle peut provenir de différents facteurs : l'entraînement des sujets à la tâche proposée (joueurs *versus* non joueurs, par exemple : on peut supposer les joueurs plus entraînés pour ce type de tâche), leur dextérité, ou encore l'efficacité des stratégies visuelles adoptées. Par exemple, on peut supposer que le balayage circulaire se révélera efficace sur la structure elliptique, mais totalement inefficace sur la structure matricielle ou radiale. C'est ce que nous allons tenter de vérifier avec l'analyse qualitative, par sujet, des parcours oculaires sur les scènes visuelles (cf. infra paragraphe 6.4 page 142).

Le tableau 6.1 ci-dessous présente les temps moyens de sélection des cibles par structure en tenant compte de la condition expérimentale F ou NF. Pour chaque structure, il y a 150 observations par condition. Pour la structure elliptique, on observe une différence statistiquement significative entre les condition F et NF (t=2,75 ; p=0,0063). Ce résultat observé pour la structure elliptique est similaire à celui observé de façon globale, i.e., toutes structures spatiales confondues (cf. supra 129). Bien qu'il n'existe pas d'autre différence statistiquement significative par structure entre les deux conditions, il n'en reste pas moins que, pour chaque structure spatiale, les temps moyens de sélection observés dans la condition NF sont plus courts que ceux observés dans la condition F. Ce qui confirme l'interprétation proposée dans le premier paragraphe de cette section.

Comparaison des temps de sélection par condition			
Structure	Condition F	Condition NF	Test t
Aléatoire	4565	4357	t=0,42 ; p=0,6728
Elliptique	4897	3668	t=2,75 ; p=0,0063
Matricielle	4624	4368	t=0,47 ; p=0,6374
Radiale	4697	4422	t=0,57 ; p=0,5677

TAB. 6.1 – Résultats par structure et par condition F/NF.
Ce tableau présente les temps de sélection moyens des cibles par structure et par condition expérimentale F ou NF. Les temps moyens de sélection des cibles sont fournis en millisecondes. La colonne 2 contient les moyennes dans la condition F, la colonne 3 les moyennes dans la condition NF.

Niveaux de difficulté des scènes

Le tableau 6.2 ci-dessous présente les temps moyens de sélection des cibles par niveau de difficulté ainsi que le taux d'erreurs, exprimé en pourcentages, pour chaque niveau de difficulté. Pour chaque niveau de difficulté, il y a 400 observations par variable.

Comparaison des niveaux de difficulté des scènes		
Niveau	Temps moyen (ms)	Taux d'erreurs (%)
Facile	3774	1,75
Moyen	4101	5,5
Difficile	5474	8,5

TAB. 6.2 – Résultats par niveau de difficulté et par condition F/NF.
Ce tableau présente les temps de sélection moyens des cibles par structure et par condition expérimentale F ou NF. Les temps moyens sont exprimés en millisecondes, les taux d'erreurs, exprimé en pourcentages, par rapport au nombre de scènes (400) par niveau de difficulté.

Concernant la rapidité de sélection des cibles, la différence de 327 ms observée entre le niveau facile et le niveau moyen n'est pas statistiquement significative (t=-1,29 et p=0,1981). En revanche, la différence de 1700 ms, respectivement 1373 ms, observée entre les niveaux facile et difficile, respectivement moyen et difficile, est statistiquement hautement significative (t=-5,60 et p<0,0001, respectivement t=-4,17 et p<0,0001).

On observe un taux d'erreurs de 1,75% pour les scènes faciles, de 5,5% pour les scènes de difficulté moyenne et de 8,5% pour les scènes difficiles. La différence observée entre les niveaux facile et moyen, respectivement facile et difficile, est statistiquement significative (t=-2,85 et p=0,0045, respectivement t=-4,38 et p<0,0001). En revanche, la différence observée entre les niveaux moyen et difficile ne l'est pas (t=-1,66 et p=0,0966).

Ces résultats sont conformes à ceux exprimés dans le chapitre précédent[94] et valident le protocole expérimental.

Objets *versus* paysages

Il y a 600 observations par type de contenu des photographies (par sujet, 60 scènes contenant des objets, 60 scènes contenant des paysages). On observe un temps moyen de sélection des cibles de 4165 ms sur les scènes contenant des objets et de 4735 ms sur les scènes contenant des paysages. Cette différence de 570 ms, observée sur les temps moyens de sélection de la cible entre "objets" et "paysages", est statistiquement significative (t=-2,32 ; p=0,0204).

On observe un taux d'erreurs de 4,83%, soit 29 erreurs, sur les scènes contenant des objets et de 5,67%, soit 34 erreurs, sur les scènes contenant des paysages. Cette différence entre les deux types d'affichages n'est pas statistiquement significative (t=0,65 ; p=0,5179).

Conclusions

Les résultats concernant les niveaux de difficulté et le contenu des photographies (paysages *versus* objets) sont conformes à ceux présentés dans la chapitre précédent. Ils valident en outre le protocole expérimental établi pour cette troisième étude concernant les trois niveaux de difficulté pour les scènes, et la distinction entre les scènes représentant des objets et celles représentant des paysages. En effet, d'après [Price et Humphreys, 1989], il semblerait que les sujets perçoivent d'abord la forme et la structure globale d'un objet, puis analysent les détails.

De la même façon, le résultat concernant les structures met en évidence l'influence de la variabilité interindividuelle et valide le choix des sujets par clustering. Ce choix va nous permettre de couvrir le maximum de stratégies d'exploration visuelle des scènes lors de l'analyse qualitative des données fournies par l'eye-tracker.

6.3.2 Analyse des stratégies de recherche visuelle

Cette partie de l'analyse porte sur les données recueillies à l'aide de l'eye-tracker. Pour mémoire, ces données sont, par sujet et par scène, et pour différents intervalles de temps, le nombre de fixations oculaires, la durée des fixations et des saccades oculaires, ainsi que la distance parcourue par le regard. En outre, on dispose, pour chaque sujet et pour chaque scène, de la représentation graphique du parcours oculaire effectué lors de la recherche visuelle. Les parcours oculaires sont définis en assimilant la trajectoire du regard entre deux fixations consécutives à un segment de droite. Les distances parcourues sont donc les schématisations à l'aide de segments de droites des parcours réels.

Ces données n'ont pu être exploitées que pour cinq sujets sur dix, en raison d'incidents techniques. Il a été impossible de calibrer correctement un sujet. Pour les quatre autres sujets, les fichiers des parcours oculaires fournis par le logiciel de rejeu des séquences se sont avérés

[94] cf. supra paragraphe 5.9 page 115.

inexploitables, en raison d'une défaillance du système de fixation de la caméra sur le casque qui a entraîné une imprécision croissante des mesures en cours de passation.

Méthodologie

Nous avons, dans un premier temps, analysé manuellement et pour chaque scène, les parcours oculaires et les positions des fixations, à l'aide du logiciel de rejeu (cf. supra paragraphe 6.2.1 page 125). Cette analyse nous a permis d'identifier deux étapes successives dans la recherche visuelle, impliquant ainsi deux formes distinctes d'activité visuelle :

- la première est l'exploration visuelle de la scène jusqu'à atteindre la cible ;
- la deuxième est la validation de cette détection visuelle par comparaison avec les autres éléments ressemblant à la cible.

En tenant compte de ces observations, nous avons choisi d'utiliser pour les analyses quantitatives le temps de parcours en millisecondes (T), le nombre de fixations oculaires (N), la durée des fixations oculaires (TF), la durée des saccades (TS) ainsi que la distance parcourue (D), en distinguant, pour chacune de ces variables :

- sa valeur entre le début de la première fixation sur la scène et la fin de première fixation sur la cible ;

de :

- sa valeur entre la fin de la première fixation sur la cible et la fin de la dernière fixation sur la scène.

Nous avons, dans un deuxième temps, analysé ces données de façon à mettre en évidence une différence statistique entre la phase d'exploration visuelle de la scène jusqu'à la détection de la cible et la phase de validation de cette détection. Puis, pour chaque phase, nous avons effectué des tests statistiques entre les deux types de contenu des photographies (objet *versus* paysage), les trois niveaux de difficulté des scènes (facile, moyen et difficile) et les quatre structures spatiales des affichages testées (aléatoire, elliptique, matricielle et radiale).

Enfin, l'analyse manuelle des parcours oculaires a également mis en évidence des différences en fonction de la position centrée *versus* excentrée de la cible. Plus la cible est centrée, plus le repérage visuel semble facile en terme du nombre de fixations notamment. Inversement, plus la cible est excentrée, plus le repérage visuel semble difficile en termes du temps de parcours de la scène jusqu'à la première fixation sur la cible, de la distance parcourue et du nombre de fixations au cours de cette phase. Nous présentons dans la suite du chapitre la validation de ces intuitions par l'analyse statistique des données recueillies sur les mouvements oculaires des cinq sujets.

La présentation de ces résultats suit un plan en trois parties. La première partie contient les analyses statistiques portant sur les deux phases du repérage visuel. La deuxième partie porte sur la mise en évidence de différences statistiquement significatives entre les deux types de contenu des photographies, les niveaux de difficulté et les structures spatiales, en distinguant les deux phases impliquées dans la recherche visuelle. La troisième et dernière partie porte sur l'analyse statistique des différences en fonction de la position centrée *versus* excentrée de la cible.

Exploration *versus* validation : deux phases distinctes du repérage de cibles

Dans toute la suite, les termes phase 1, respectivement phase 2, désignent la phase d'exploration de la scène jusqu'à la première fixation sur la cible, respectivement de validation de cette détection depuis la première fixation sur la cible jusqu'à la dernière fixation sur la scène. Il y a 567 observations par phase, car pour 33 scènes la cible n'a fait l'objet d'aucune fixation. Les résultats de l'analyse statistique des différences entre ces deux phases sont présentés dans la tableau 6.3 ci-dessous.

Variable	Phase 1	Phase 2	(1-2)	Test t
T (ms)	2583	1571	1012	t=7,06 ; p<0,0001
N	9,67	5,05	4,62	t=9,19 ; p<0,0001
TF (ms)	155	281	-126	t=-17 ; p<0,0001
TS (ms)	134	74	60	t=17,3 ; p<0,0001
D (pixels)	1794	606	1188	t=12,14 ; p<0,0001

TAB. 6.3 – Exploration *versus* validation.
La colonne (1-2) contient la différence, pour chacune des variables T, N, TF, TS et D, entre la phase 1 et la phase 2. Les valeurs contenues dans les colonnes phase 1, phase 2 et (1-2) sont des moyennes. La colonne test t contient les résultats des tests de Student effectués sur les données.

Pour chacune des variables T, N, TF, TS et D la différence entre les deux phases est statistiquement hautement significative. Autrement dit, le temps moyen de parcours d'exploration de la scène (2583 ms) est plus long que celui de la validation du choix de la cible (1571 ms). Ce temps plus long lors de la phase 1 est accompagné d'un nombre de fixations moyen de 10 environ (9,67), le double par rapport à la phase 2 (4,62 fixations en moyenne). En outre, lors de l'exploration de la scène, la durée moyenne des saccades est plus longue, accompagnée d'une distance parcourue moyenne quatre fois plus importante, que lors de la validation du choix de la cible (respectivement 134 ms *versus* 74 ms pour la durée moyenne des saccades et 1794 pixels *versus* 606 pixels pour la distance parcourue moyenne). Enfin, le temps de fixation moyen est plus court lors de la phase 1 que lors de la phase 2.

Ces résultats valident l'existence de deux phases distinctes lors de la recherche de cibles. La première est plus longue que la deuxième. Il s'agit de l'exploration très rapide d'un grand nombre de photographies. En effet, on observe des temps de fixation deux fois plus courts accompagnés d'un nombre de fixations oculaires deux fois plus élevé, par rapport à la seconde phase. Les sujets semblent percevoir rapidement la forme générale ainsi que les principales caractéristiques (couleurs, taille) des photographies explorées, conformément à [Duncan et Humphreys, 1989] et [Price et Humphreys, 1989]. En outre, les sujets atteignent pour la première fois la cible à l'issue d'environ 10 fixations oculaires, soit après avoir exploré seulement le tiers des photographies contenues dans la scène. L'exploration semble guidée par la saillance visuelle des items contenus dans la scène. L'importante distance parcourue ainsi que les saccades très longues, associées à des temps de fixations très courts, semblent corroborer cette interprétation.

En effet, on peut supposer que le regard se fixe tout d'abord sur les photographies les plus ressemblantes à la cible, en terme de saillance visuelle (contraste, couleur, direction, forme) grâce à la vision périphérique. Ce qui explique les longues saccades et longues distances parcourues. Pour résumer, lors de la première phase de recherche visuelle, les sujets semblent adopter une stratégie de recherche "par la saillance visuelle d'abord".

La deuxième phase de recherche est plus courte. Il s'agit de la validation de la détection de la cible "candidate" par comparaison avec d'autres items contenus dans la scène. Nombre de fixations moyen, durée moyenne des saccades et distances parcourues moyennes sont réduits, par rapport à la phase précédente. En revanche, la durée moyenne des fixations est plus importante (près de deux fois plus longue que lors de la phase précédente). Ce qui signifie que lors de cette phase les sujets exploitent les détails des photographies similaires à la cible pour valider ou non le résultat de leur précédente exploration, conformément à [Duncan et Humphreys, 1989] et [Price et Humphreys, 1989].

Analyse détaillée de la première phase du repérage visuel de cibles : l'exploration de la scène

Dans ce paragraphe, nous présentons l'analyse détaillée des performances des sujets, en termes des variables T, N, TF, TS et D, en fonction des structures spatiales, du type de contenu des photographies et des niveaux de difficulté des scènes. Les résultats sont présentés dans les tableaux 6.4, 6.6 et 6.7 suivants.

Structure des affichages				
Variable	Aléatoire	Radiale	Elliptique	Matricielle
T (ms)	2485,2	2685,2	2274,2	2887,5
N	9,383	9,7589	8,7483	10,782
TF (ms)	152,8	157,45	153,05	156,86
TS (ms)	134,44	143,12	129,72	128,94
D (pixels)	1818,4	1972,6	1439,9	1951

TAB. 6.4 – Analyse détaillée de la phase de "exploration de la scène" : (1).
On dénombre 141 observations pour chacune des structures aléatoire et radiale, 142 observations pour la structure matricielle et 143 observations pour la structure elliptique.

Variable : distance parcourue (pixels)		
Structures	Différence (pixels)	Test t
Aléatoire-Elliptique	378,5	t=2,08 ; p=0,0380
Radiale-Elliptique	532,7	t=2,44 ; p=0,0462
Matricielle-Elliptique	511,1	t=2,53 ; p=0,0119

TAB. 6.5 – Différences statistiques observées pour la variable D.

On constate que le classement en termes de temps de parcours moyen croissant des structures et de nombre de fixations par structure est le suivant : ellipse (2274,4 ms et 8,7 fixations en moyenne), aléatoire (2485,2 ms et 9,4 fixations en moyenne), radiale (2685,2 ms et 9,8 fixations en moyenne) et matricielle (2887,5 ms et 10,8 fixations en moyenne). Seule, la différence de 613,1 ms observée entre les structures elliptique (2274,4 ms) et matricielle (2887,5 ms) est statistiquement significative (t=-2,00 ; t=0,0462). Autrement dit, l'écart de plus d'une demi-seconde observé entre le parcours des ellipses et le parcours des matrices est statistiquement significatif. On n'observe de différence statistiquement significative, ni concernant les durées de fixation moyennes, ni concernant les durées moyennes des saccades. Les durées de fixation moyennes sont comprises entre 152,8 ms pour la structure aléatoire et 157,45 ms pour la structure radiale. Les durées moyennes des saccades sont comprises entre 129,72 ms pour la structure elliptique et 143,12 ms pour la structure radiale. En revanche, concernant la distance parcourue dans la scène, exprimée en pixels, la structure elliptique se distingue par rapport aux autres. En effet, on observe une différence statistiquement significative entre l'ellipse et les trois autres structures, comme le montre le tableau 6.5.

Types d'affichages		
Variable	Objets	Paysages
T (ms)	2365,3	2800,7
N	8,9472	10,389
TF (ms)	156,25	153,82
TS (ms)	131,89	136,49
D (pixels)	1686,7	1902,8

TAB. 6.6 – Analyse détaillée de la phase "exploration de la scène" : (2).
On dénombre 284 observations pour les affichages contenant des objets et 283 observations pour les affichages contenant des paysages.

Niveaux de difficulté			
Variable	Facile	Moyen	Difficile
T (ms)	2386	2372,4	3008,7
N	8,9286	8,8032	11,344
TF (ms)	150,16	162,33	152,77
TS (ms)	138,99	139,66	123,41
D (pixels)	1790,7	1600,5	1998,2

TAB. 6.7 – Analyse détaillée de la phase "exploration de la scène" : (3).
On dénombre 196 observations pour le niveau facile, 188 pour le niveau moyen et 183 pour le niveau difficile.

Concernant le type de contenu des photographies, on constate d'après le tableau 6.6 que les temps moyens de parcours des scènes sont plus courtes lorsque celles-ci contiennent des objets plutôt que des paysages : 2365 ms pour les objets *versus* 2801 ms pour les paysages, soit

une différence de 435,4 ms statistiquement significative (t=-1,98 ; p=0,0479). Excepté pour la variable T, il n'existe aucune autre différence statistiquement significative concernant le type de contenu des photographies. Les nombres moyens de fixations par scène sont proches avec, en moyenne, 8,9 fixations sur les objets *versus* 10,4 sur les paysages. De même, seulement 2,43 ms séparent les temps de fixations moyens observés sur les objets (156,25 ms) de ceux observés sur les paysages (153,82 ms). Enfin, ni la différence de 4,6 ms entre les durées moyennes des saccades, ni la différence de 216,1 pixels entre les distances moyennes parcourues, ne sont statistiquement significatives. Néanmoins, le repérage visuel semble être plus rapide, donc plus facile, lorsque les cibles sont des objets, car les sujets perçoivent plus rapidement (voir les résultats obtenus pour les variables T, N, TS et D) les détails discriminant les items les uns des autres. Ce qui n'est pas le cas dans le repérage visuel de paysages, car ceux-ci ne véhiculent aux sujets que des indications globales, comme la forme (montagne, par exemple) ou la couleur (couchers de soleil, par exemple). Ce résultat est compatible avec le modèle "coarse to fine" de recherche visuelle [Huges *et al.*, 1996].

Concernant le niveau de difficulté des scènes, on constate d'après le tableau 6.7 que les temps de parcours moyens des scènes sont proches entre les niveaux 1 et 2. Contre toute attente, les sujets se sont même globalement révélés plus rapides sur les scènes de niveau moyen que sur celles de niveau facile, sans que cette légère différence soit significative. En revanche, entre les niveaux 1 et 3, respectivement 2 et 3, les différences d'environ 630 ms observées pour la variable T sont statistiquement significatives : (t=-2,23 ; p=0,0264), respectivement (t=-2,21 ; p=0,0280). Excepté pour les durées moyennes des fixations, il n'existe aucune différence statistiquement significative entre les niveaux 1 et 2 (t=-2,00 ; p=0,0463). Il convient de noter que la durée moyenne des fixations est plus longue pour les scènes moyennes que pour les scènes faciles ou difficiles. Ce résultat semble s'expliquer par le fait que les scènes associées au niveau 2 sont plus homogènes visuellement que celles du niveau 1. La saillance visuelle des items contenus dans les scènes de niveau 2 étant moindre, il faut examiner les photographies plus longtemps avant de les éliminer en tant que cible candidate.

Les différences statistiquement significatives entre les niveaux 1 et 3, respectivement 2 et 3, sont présentées dans le tableau 6.8, respectivement le tableau 6.9. Le nombre important de différences statistiquement significatives observées, excepté entre les niveaux 1 et 2 où seule la variable TF fait l'objet d'une différence significative, suggère que les stratégies d'exploration visuelle adoptées par les sujets diffèrent entre les niveaux 1 et 3 d'une part, et entre les niveau 2 et 3 d'autre part. Les stratégies de réalisation des tâches de repérage visuel sur des scènes faciles ou de niveau de difficulté moyenne semblent, en revanche, identiques.

Facile *versus* difficile		
Variable	1-3	Test t
T (ms)	622,7	t=-2,23 ; p=0,0264
N	2,4154	t=-2,44 ; p=0,0151
TS (ms)	15,58	t=2,44 ; p=0,0153

TAB. 6.8 – Différences observées entre les scènes faciles et difficile.

Moyen *versus* difficile		
Variable	2-3	Test t
T (ms)	636,3	t=-2,21 ; p=0,0280
N	2,5408	t=-2,49 ; p=0,0131
TS (ms)	16,25	t=2,20 ; p=0,0283
D (pixels)	397,7	t=-2,03 ; p=0,0429

TAB. 6.9 – Différences observées entre les scènes de difficulté moyenne et difficile.

Pour résumer, les stratégies d'exploration visuelle adoptées par les sujets ne semblent pas être affectées par le type de contenu des photographies, objet ou paysage. En revanche, des niveaux de difficulté des scènes différents semblent impliquer des stratégies différentes, notamment en termes de nombre de fixations et de durée des saccades. Les scènes difficiles nécessitent un plus grand nombre de fixations pour découvrir la cible, en raison probablement du manque de saillance visuelle de celle-ci. Par ailleurs, il apparaît que la structure des affichages a une influence directe sur les temps d'exploration des scènes, sur le nombre de fixations, et sur la distance parcourue au sein d'une scène. En particulier, les structures elliptiques semblent être les plus propices au repérage efficace de cibles, en termes surtout de temps d'exploration moyens, de nombre moyen de fixations et de distance moyenne parcourue, par rapport aux structures matricielles. Ces dernières semblent les moins efficaces en terme de temps d'exploration. En outre, les saccades plus courtes traduisent un parcours oculaire de la matrice probablement par balayage linéaire. Enfin, il convient de noter que les structures aléatoires permettent une exploration efficace des scènes, en termes surtout de temps d'exploration moyen, de nombre moyen de fixations et de distance parcourue en moyenne. L'exploration visuelle des scènes dont la structure est aléatoire semble guidée par la saillance visuelle. Il peut s'agir de la saillance visuelle d'un item en particulier, mais aussi de la saillance de certaines zones de l'affichage, probablement les zones où la densité des photographies est la plus forte. Pour infirmer ou valider ces interprétations, nous avons effectué l'analyse détaillée des parcours oculaires individuels[95].

Analyse détaillée de la deuxième phase du repérage visuel de cibles : la validation du choix de la cible candidate

Dans ce paragraphe, nous présentons l'analyse détaillée des performances des sujets au cours de la phase 2, en termes des variables T, N, TF, TS et D, en fonction de la structure spatiale des affichages. Les résultats sont fournis dans le tableau 6.10.

Nous n'avons observé, pour aucune des variables utilisées, de différence statistiquement significative. En revanche, l'analyse de ces résultats quantitatifs montre que, globalement, c'est pour les structures matricielles que cette phase de validation, ou vérification, semble la plus efficace. En effet, non seulement le temps moyen de parcours de la scène et la durée moyenne des fixations oculaires sont les plus courts (respectivement, 1389 et 266,05 ms) mais en plus le nombre moyen de fixations oculaires est le plus petit (environ 4,5). En outre, comme pour les

[95] cf. infra 6.4 page 142.

Structure des affichages				
Variable	Aléatoire	Radiale	Ellipse	Matrice
T (ms)	1746,8	1522,7	1624,1	1389
N	5,617	4,8936	5,1608	4,5282
TF (ms)	277,68	279,88	303,45	266,05
TS (ms)	73,101	77,294	71,267	73,486
D (pixels)	722,88	677,62	510,85	516,03

TAB. 6.10 – Analyse détaillée de la phase de "validation" par structure.
On dénombre 141 observations pour chacune des structures aléatoire et radiale, 142 observations pour la structure matricielle et 143 observations pour la structure elliptique.

structures elliptiques, la distance moyenne parcourue est faible (516,03 pixels), par rapport aux structures radiales (677,62 pixels) et aux structures aléatoires (722,88 pixels). La durée moyenne des saccades, intermédiaire relativement aux autres structures (71,267 ms pour les structures elliptiques, 73,101 ms pour les structures aléatoires, 73,486 ms pour les structures matricielles et enfin, 77,294 ms pour les structures radiales), semble indiquer un parcours par balayage linéaire de la matrice plutôt qu'un parcours guidé par la saillance visuelle des items. Ce qui laisse à penser que la phase d'exploration, plus lente dans cette structure et assortie d'un nombre plus élevé de fixations s'avère somme toute efficace. Une fois la cible atteinte, la phase de validation est plus courte, peut-être parce que le repérage initial paraît plus sûr aux sujets (moins de fixations, fixations plus courtes).

Inversement, les structures aléatoires semblent être les moins efficaces concernant la phase de validation du choix de la cible candidate. En effet, le temps moyen de parcours y est important, accompagné du nombre de fixations moyen le plus élevé et de la distance moyenne parcourue la plus longue. Dans ce type de structure, la phase de validation semble guidée davantage par la saillance que par la structure globale de la scène (cf. la durée moyenne des saccades qui est intermédiaire).

Enfin, les structures radiales semblent plus efficaces que les structures elliptiques, en termes de temps moyen de parcours, de durée moyenne des fixations et en terme de nombre moyen de fixations. L'analyse détaillée des parcours oculaires, section 6.4 page 150, devrait nous fournir plus d'informations concernant la phase de validation. La comparaison de ces résultats à ceux observés pour la première phase permet d'expliquer pourquoi lorsqu'on examine les temps de sélection de la cible qui regroupent les deux phases, aucune structure ne se distingue nettement des autres (cf. tableau 6.1 page 131).

Position centrée *versus* position excentrée : analyse détaillée

C'est l'analyse manuelle des parcours oculaires qui nous a amené à considérer la position centrée *versus* excentrée de la cible comme facteur pouvant influencer les performances des sujets, donc les temps moyens de parcours, le nombre et la durée moyens des fixations, la durée moyenne des saccades et les distances moyennes parcourues. Pour les structures elliptiques notamment,

nous avons constaté que les performances des sujets étaient meilleures si la cible était située sur l'ellipse intérieure, plutôt que sur l'ellipse extérieure.

Pour les structures elliptique, matricielle et radiale[96], nous avons découpé la scène en deux zones : la zone des cibles centrées et la zone des cibles excentrées. En ce qui concerne la structure elliptique, appartiennent à la zone centrée les photographies (8) situées sur l'ellipse intérieure, les photographies (22) situées sur l'ellipse extérieure appartenant à la zone excentrée. En ce qui concerne la structure radiale, appartiennent à la zone centrée les photographies de chaque rayon (8), les autres photographies (22) appartenant à la zone excentrée, comme le montre la figure 6.2. En ce qui concerne la structure matricielle, appartiennent à la zone centrée les photographies (6) situées sur une matrice intérieure 2×3, les autres photographies (24) appartenant à la zone excentrée, comme le montre la figure 6.3. Ces trois zones centrales ont été définies en fonction de la position des photographies par rapport au centre de l'écran en utilisant un angle visuel de 12°, qui correspond à l'estimation courante de la taille du champ visuel humain, vision périphérique incluse.

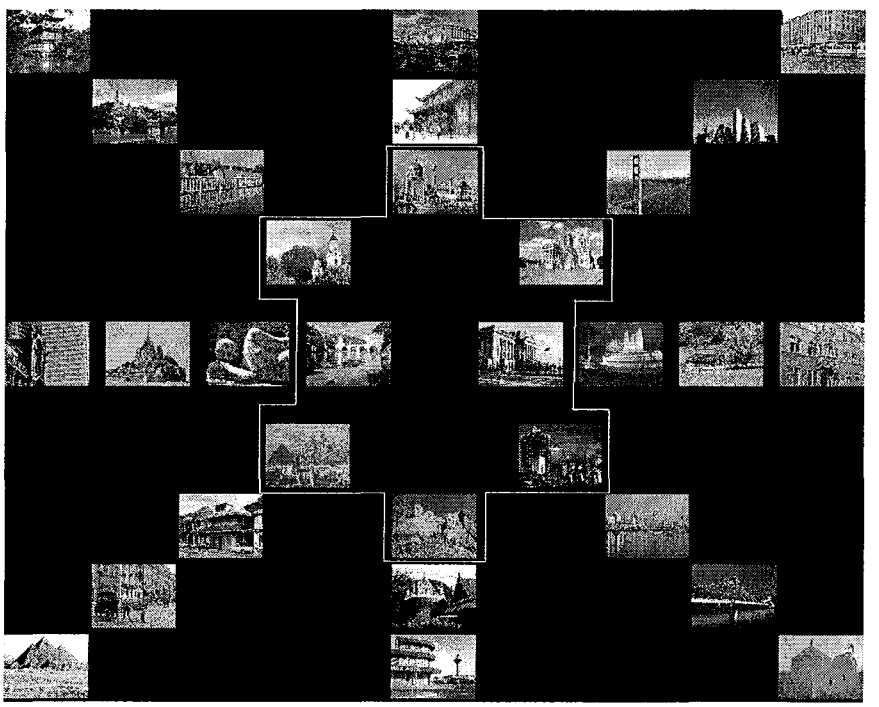


FIG. 6.2 – Zone centrée : structure radiale.
La zone centrée est délimitée par le polygone blanc.

Les tableaux 6.11 et 6.12 présentent les données des variables T, N, TF, TS et D pour la position centrée *versus* excentrée de la cible, respectivement pour les phases 1 et 2 impliquées dans le repérage visuel.

[96]Nous avons écarté les structures aléatoires de cette étude en raison même de leur caractère aléatoire, les structures spatiales des 30 scènes aléatoires étant toutes distinctes les unes des autres.

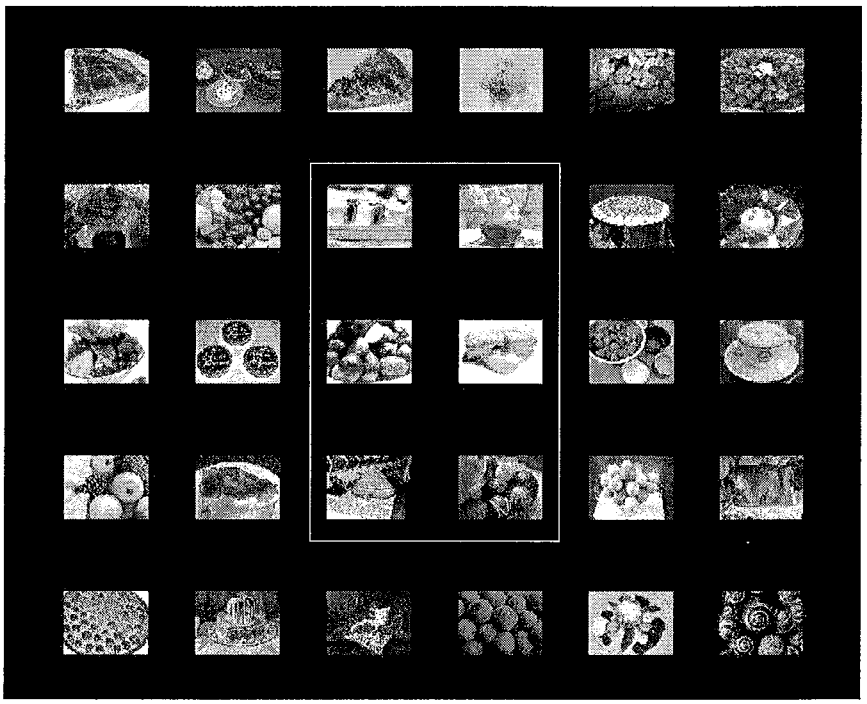


FIG. 6.3 – Zone centrée : structure matricielle.
La zone centrée est délimitée par le rectangle blanc.

Variable	Excentré	Centré	Différence	Test t
T (ms)	2848,2	1854	994,25	t=320 ; p=0,0015
N	10,632	6,92	3,7119	t=3,39 ; p=0,0008
TF (ms)	148,76	178,66	-29,9	t=-3,13 ; p=0,0023
TS (ms)	142,99	105,1	37,898	t=-4,03 ; p<0,0001
D (pixels)	1985,9	1137,2	848,75	t=4,00 ; p<0,0001

TAB. 6.11 – Position centrée *versus* position excentrée de la cible : phase 1.
Il y a 100 observations pour les positions centrées, 326 pour les positions excentrées.

Variable	Excentré	Centré
T (ms)	1626,1	1140,8
N	5,2178	3,7
TF (ms)	285,13	276,83
TS (ms)	75,606	68,77
D (pixels)	610,64	428,01

TAB. 6.12 – Position centrée *versus* position excentrée de la cible : phase 2.
Il y a 100 observations pour les positions centrées, 326 pour les positions excentrées.

Toutes les différences observées dans le tableau 6.11 sont statistiquement significatives. Autrement dit, l'efficacité de l'exploration visuelle, exprimée par les variables T, N, TF, TS et D, dépend de la position centrée *versus* excentrée de la cible dans la scène. Parmi les différences présentées dans le tableau 6.12, seules les différences de 485,36 ms (T) et de 1,5178 (N) sont statistiquement significatives : respectivement (t=-2,32 ; p=0,0213) et (t=-2,07 ; p=0,0398). Autrement dit, temps de parcours et nombre de fixations sont influencés par la position de la cible pendant la phase de validation de sa détection. Seule une analyse détaillée des parcours oculaires individuels permettrait d'expliquer ces observations et d'identifier les stratégies qui les sous-tendent. L'analyse manuelle des quelques parcours oculaires pendant cette phase suggère deux types de stratégies :

- soit les sujets parcourent la scène pour s'assurer que la cible choisie est la seule candidate possible ;
- soit les sujets hésitent entre plusieurs photographies candidates qu'ils comparent, d'où des mouvements de va-et-vient entre les différents candidats.

Nous n'avons pas été surpris par ces résultats concernant la première phase. En effet, le bouton "OK" de repositionnement de la souris entre la présentation de la cible et l'affichage de la scène est positionné au centre de l'écran. Ainsi, les premières fixations partent du centre et, grâce à la vision périphérique, les cibles situées près du centre sont plus faciles à détecter. Ce résultat est très intéressant pour faciliter des activités visuelles comme la navigation sur Internet ou la recherche au sein de grands ensembles d'informations. Pouvoir anticiper la position des premières fixations, permet d'attirer le regard de l'utilisateur vers les composants de l'affichage considérés comme importants.

6.4 Analyse détaillée des performances et des parcours oculaires individuels

L'analyse des parcours oculaires des scènes se présente en deux parties. La première partie porte sur l'analyse quantitative des performances, en termes de temps de sélection des cibles et de précision, pour les 10 sujets ayant participé à l'expérimentation. L'objectif est double :

- étudier l'évolution de leurs performances entre les études expérimentales 2 et 3 ;
- établir une classification hiérarchique des sujets en terme de temps de sélection des cibles et en tenant compte des deux conditions expérimentales.

La deuxième partie porte sur l'analyse qualitative des parcours oculaires des cinq sujets pour lesquels les données recueillies à l'aide de l'eye-tracker étaient exploitables[97]. L'analyse, sujet par sujet, des parcours oculaires devrait permettre d'identifier des groupes de sujets au sein desquels les stratégies d'exploration visuelle adoptées sont similaires, conformément à la classification hiérarchique établie dans la première partie de l'analyse. En outre, elle devrait permettre de montrer si ces stratégies diffèrent d'une structure spatiale à l'autre. Autrement dit, les sujets adaptent-ils leur stratégie de recherche visuelle en fonction de l'organisation spatiale des affichages ? Auquel cas, les structures spatiales agiraient comme une forme de guidage visuel. Enfin, cette analyse de-

[97] cf. supra section 6.3.2 page 132.

vrait permettre de déterminer, pour chacune des structures spatiales testées, la ou les stratégies qu'il convient d'adopter pour une recherche visuelle efficace.

6.4.1 Analyse des performances individuelles des sujets

Repérage visuel de cibles et apprentissage

Le tableau 6.13 présente, pour chacun des 10 sujets, le temps moyen de sélection de la cible dans l'ordre de passation. Par exemple, si le sujet a effectué les tâches de repérage dans l'ordre F/NF, alors, dans la colonne "1ère condition", apparaît le temps moyen des sélections dans la condition F et, dans la colonne "2ème condition", le temps moyen dans la condition NF.

Temps moyens de sélections des cibles (ms)			
Sujet	Ordre	1ère condition	2ème condition
1 (23)	NF/F	5794	5420
2 (1)	F/NF	4646	3514
3 (14)	F/NF	3449	3172
4 (15)	NF/F	4968	5410
5 (13)	F/NF	3606	3077
6 (20)	F/NF	5610	5231
7 (22)	NF/F	3728	5337
8 (3)	NF/F	3651	4357
9 (8)	F/NF	5358	4125
10 (6)	NF/F	4778	3769

TAB. 6.13 – Résultats des sujets selon l'ordre de passation.
Pour chacun des dix sujets, on dénombre 60 observations par condition expérimentale. Le numéro des sujets figure dans la première colonne avec, entre parenthèses, leur numéro lors de la deuxième expérimentation. L'ordre de passation F/NF ou NF/F figure dans la colonne 2.

On constate que seulement trois sujets ont été plus lents dans la deuxième partie de la passation. Pour ces trois sujets, l'ordre de passation était NF/F. Comme nous l'avons mentionné page 129, certains sujets ayant effectué les tâches de repérage dans l'ordre NF puis F semblent avoir été perturbés dans la condition F, ne se souvenant pas de la position des cibles dites familières. Néanmoins, pour les sept sujets qui ont été plus rapides dans la deuxième partie de l'expérimentation, il y a apprentissage de la tâche. Par ailleurs, globalement, les sujets ont été meilleurs lors de cette expérimentation, relativement à la précédente (cf. chapitre précédent). La moyenne des temps de sélection de cibles observée lors de la deuxième expérimentation dans la condition visuelle (PV) est de 5316 ms *versus* 4450 ms lors de la troisième expérimentation, conditions expérimentales F et NF confondues. Cette différence est statistiquement hautement significative ($t=4{,}39$; $p<0.0001$). Ces résultats traduisent le phénomène d'apprentissage de la tâche pour le repérage visuel de cibles.

Le tableau 6.14 présente l'évolution des performances des sujets entre les deux expérimentations. Pour chaque sujet, ce tableau fournit dans les colonnes 3, 4 et 5 les temps de sélection moyens observés lors de la 2ème expérimentation (condition PV), puis ceux observés lors de la 3ème expérimentation et enfin le test statistique associé.

Variable : temps moyen de sélection des cibles (ms)				
Sujet	Ordre	Moyenne 2 (ms)	Moyenne 3 (ms)	Test t
1 (23)	NF/F	9113	5607	t=4,41 ; p<0,0001
2 (1)	F/NF	5388	4080	t=2,63 ; p=0,0091
3 (14)	F/NF	3167	3311	t=-0,39 ; p=0,6959
4 (15)	NF/F	7178	5189	t=2,43 ; p=0,0158
5 (13)	F/NF	4403	3341	t=2,32 ; p=0,0209
6 (20)	F/NF	4262	5421	t=-1,72 ; p=0,0873
7 (22)	NF/F	4011	4532	t=-1,22 ; p=0,2239
8 (3)	NF/F	5009	4004	t=2,12 ; p=0,0351
9 (8)	F/NF	5720	4741	t=1,52 ; p=0,1292
10 (6)	NF/F	4905	4273	t=0,92 ; p=0,3584

TAB. 6.14 – Évolution des performances des sujets entre les deux expérimentations.
Le numéro des sujets figure dans la première colonne avec, entre parenthèses, leur numéro lors de la deuxième expérimentation. L'ordre de passation F/NF ou NF/F figure dans la colonne 2. Il y a 120 observations par étude.

Les différences observées sont statistiquement significatives, excepté pour les sujets 3, 6, 7, 9, 10. Ce résultat peut s'expliquer comme suit : il s'agit des cinq sujets qui avaient effectué les tâches de repérage, lors de l'étude précédente, dans l'ordre PM-PV. Ils avaient réalisé de très bonnes performances dans la condition PV par rapport à l'autre groupe de sujets, faisant appel à leur mémoire pour réaliser les tâches de repérage dans cette condition. Trois d'entre eux ont été plus lents lors de la 3ème étude, les deux autres ont été plus rapides. Ce résultat illustre l'importante variabilité interindividuelle qui se traduit également, dans le tableau 6.14, par les écarts importants entre les sujets, quelle que soit l'étude considérée. Pour les cinq autres sujets, les différences observées entre les deux études sont statistiquement significatives. Tous les cinq ont été meilleurs lors de la 3ème étude.

En résumé, que ce soit à court terme, ou bien à long terme, il y a apprentissage de la tâche de repérage visuel de cibles pour un sujet sur deux.

Classification hiérarchique (10 sujets)

À partir des temps moyens de sélection et de la précision observés dans les conditions F et NF, nous avons réalisé sous SAS une classification hiérarchique par la méthode des centres mobiles des dix sujets ayant participé à la troisième expérimentation. La figure 6.4 illustre les résultats obtenus.

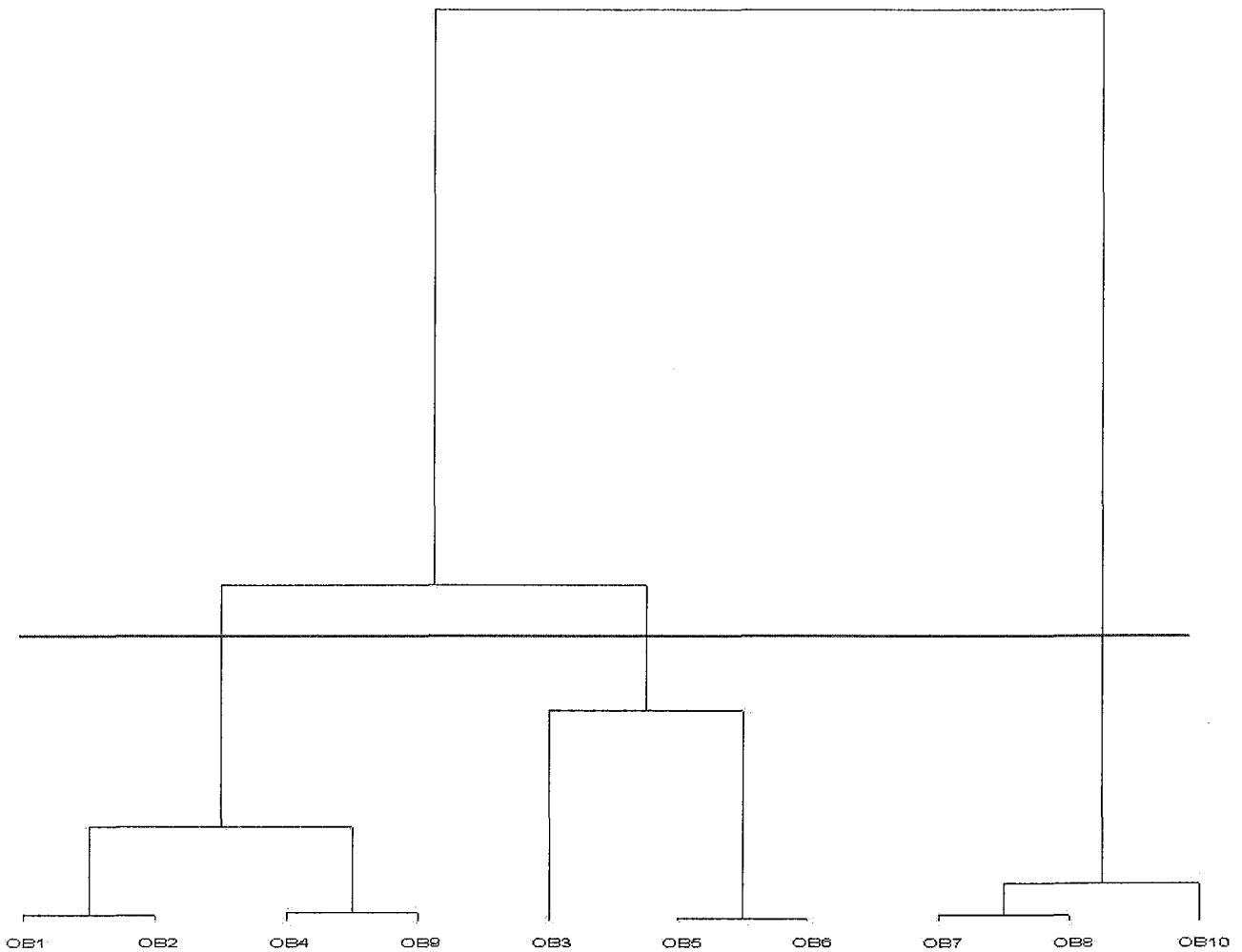


FIG. 6.4 – Classification hiérarchique ; 10 sujets ; temps de sélection des cibles.
Cette classification hiérarchique a été réalisée sous SAS par la méthodes des centres mobiles en tenant compte des deux conditions expérimentales F et NF, des temps de sélection des cibles ainsi que de la précision des sélections. Les notations OB1 à OB10 désignent les 10 sujets ayant participé à la troisième expérimentation, soit OB1 pour le sujet n°1, et ainsi de suite. Le groupe 1 est formé de l'ensemble des sujets {1,2,4,9}, le groupe 2 de l'ensemble {3,5,6}, le groupe 3 de l'ensemble {7,8,10}.

Au sein de cette classification hiérarchique, nous avons distingué trois groupes de sujets dont le profil semble différent : le groupe de sujets {1,2,4,9}, le groupe de sujets {3,5,6} et le groupe de sujets {7,8,10}. Nous avons choisi ce découpage de façon à disposer des parcours oculaires d'au moins un sujet par groupe. Pour mémoire, seuls les parcours oculaires des sujets 5, 6, 7, 8 et 9 étaient exploitables. Nous allons vérifier, dans la suite de l'analyse, si à ces groupes correspondent des stratégies différentes de recherche visuelle : les parcours oculaires du sujet 9, des sujets 5 et 6, des sujets 7 et 8 diffèrent-ils entre-eux ? C'est ce que nous allons analyser dans la première partie de la section suivante.

6.4.2 Les structures spatiales : une forme de guidage pour le repérage visuel de cibles

Analyse intra-sujet (5 sujets) : phase d'exploration (1)

L'analyse manuelle des parcours oculaires pour chacun des sujets 5, 6, 7, 8 et 9 a permis de montrer que les parcours oculaires sont effectivement de nature différente entre les trois groupes de sujets (sujet 9 *versus* sujets 5 et 6 *versus* sujets 7 et 8). Pour cette analyse, nous nous plaçons dans le cas de scènes de niveau de difficulté moyenne ou difficiles. En effet, sur les scènes faciles, le parcours oculaire peut n'être composé que d'une unique fixation sur la cible. Globalement, sauf cas extrême, nous n'avons observé sur les scènes faciles que des parcours oculaires dont le nombre de fixations est inférieur à six.

Le sujet 9 ne suit pas les structures spatiales, à l'exception des structures elliptiques (dans les cas où la scène est d'un niveau de difficulté facile ou moyen). Les parcours oculaires adoptés par le sujet 9 présentent les caractéristiques suivantes :

- les saccades sont très longues, quelle que soit la structure spatiale ;
- les distances parcourues sont très importantes, surtout lors de la recherche au sein de structures matricielles ;
- les parcours oculaires des scènes contiennent, pour la plupart, un nombre très important d'aller-retour sur les mêmes photographies (cf. figure 6.5).

Nous avons comparé le parcours oculaire du sujet 9 illustré par la figure 6.5 à celui observé pour le sujet 5, le plus rapide pour toute la passation parmi les sujets 5, 6, 7, 8 et 9 (cf. figure 6.6). On observe sur la même scène, pour le sujet 9, un temps de parcours global de la scène de 2,6 secondes *versus* 20,9 secondes pour le sujet 9. Le sujet 5 a atteint la cible à l'issue de 10 fixations en 2,1 secondes *versus* respectivement 19,9 secondes et 68 fixations pour le sujet 9. Enfin, pour les scènes à structure radiale, le sujet 9 effectue un parcours systématique de toutes les photographies. Pour les scènes à structure elliptique, il navigue entre les deux ellipses. Enfin, d'après le debriefing de la deuxième étude, c'est au sein des structures aléatoires que ce sujet trouve le repérage de cibles le plus confortable.

Les sujets 5 et 6 adoptent des parcours oculaires plus efficaces que le sujet 9 en terme d'aller-retour. En outre, ils semblent guidés d'abord par la saillance visuelle de certains items ressemblant à la cible ; ils ne suivent les structures spatiales qu'après quelques fixations guidées par la saillance. Les parcours oculaires adoptés par les sujets 5 et 6 présentent les caractéristiques suivantes :

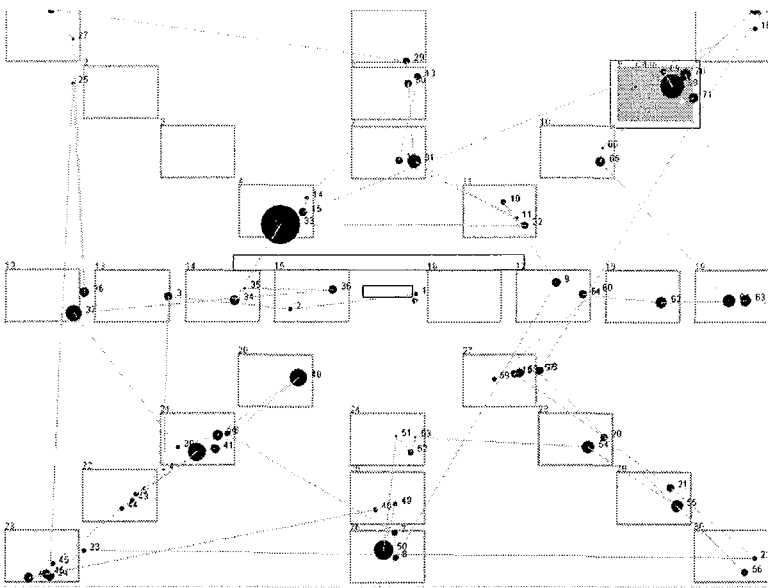


FIG. 6.5 – Sujet 9 : exemple de nombreux aller-retour.
Condition NF ; structure radiale ; difficulté moyenne. Le point de départ des fixations est la fixation numérotée 1 située au centre de l'écran. La dernière fixation est en jaune. Le sujet a atteint la cible, en vert, à l'issue de 68 fixations et 19,9 secondes. 72 fixations au total lui ont été nécessaires pour sélectionner la cible. Les disques noirs représentent les fixations. Leur diamètre est proportionnel à la durée de la fixation.

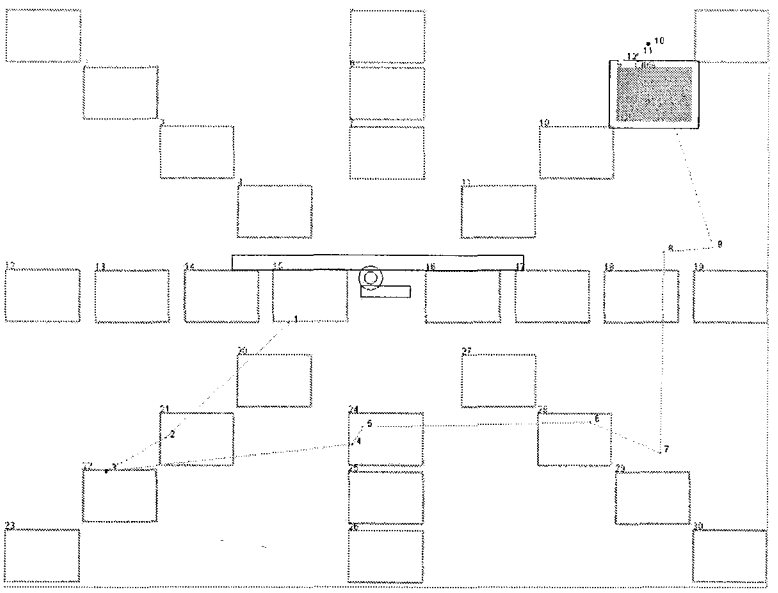


FIG. 6.6 – Sujet 5 : exemple de parcours oculaire de la structure radiale.
Même scène que celle présentée figure 6.5 : condition NF ; difficulté moyenne. Le point de départ des fixations est la fixation numérotée 1 située au centre de l'écran. La dernière fixation est en jaune. Le sujet a atteint la cible, en vert, à l'issue de 10 fixations en 2,1 secondes. À noter la brièveté des fixations qui dénote un examen sommaire des photographies sur le parcours oculaire, vraisemblablement fondé principalement sur leurs caractéristiques visuelles.

- les parcours oculaires au sein des structures aléatoires sont basés sur la saillance visuelle des items : d'abord vers les items dont la couleur est proche de celle de la cible, puis vers les zones les plus informatives (denses) de la scène ;
- les parcours oculaires au sein des structures elliptiques sont basés sur la saillance visuelle des items d'abord, puis, balayage circulaire de l'ellipse extérieure, si la première stratégie échoue ; l'ellipse intérieure n'est que très peu visitée (on suppose que c'est la vision périphérique qui permet ce type de stratégie d'exploration visuelle) ;
- les parcours oculaires au sein des structures radiales ne suivent pas les rayons, mais forment un cercle passant au milieu de chacun des rayons ;
- les parcours oculaires au sein des structures matricielles sont guidés par la saillance visuelle des items ressemblant à la cible ; les sujets n'adoptent un parcours par balayage horizontal ou vertical que si la recherche par la saillance s'avère être un échec, ou si aucun item n'est saillant.

On peut faire l'hypothèse que ces deux sujets utilisent leur vision périphérique pour explorer en priorité les photographies qui possèdent les mêmes propriétés visuelles que la cible (i.e., contraste, direction, forme, à l'exclusion de la couleur).

Les sujets 7 et 8 suivent les structures spatiales :

- les parcours oculaires au sein des structures aléatoires sont guidés par les agglomérations ou alignements de photographies au sein des affichages ;
- les parcours oculaires au sein des structures elliptiques sont réalisés par balayage circulaire de l'ellipse extérieure ; on dénombre très peu de fixations oculaires sur l'ellipse intérieure ;
- les parcours oculaires au sein des structures radiales suivent les rayons ;
- les parcours oculaires au sein des structures matricielles sont réalisés par balayages horizontaux et verticaux.

Il convient de noter également la présence de nombreux aller-retour quelle que soit la structure spatiale pour le sujet 8, contrairement au sujet 7.

L'analyse intra-sujet a permis d'identifier des groupes de sujets en fonction des stratégies d'exploration visuelle mise en œuvre : le groupe constitué du sujet 9, le groupe constitué des sujets 5 et 6 et le groupe constitué des sujets 7 et 8. Le sujet 9 (temps moyen de sélection : 4711 ms) utilise une stratégie de recherche qui consiste à "analyser" tous les éléments de la scène en effectuant de nombreux aller-retour. Les sujets 5 et 6 (temps moyens de sélection : respectivement 3341 ms et 5421 ms) adoptent une stratégie de recherche "par la saillance visuelle d'abord". Cette stratégie utilise la structure de l'affichage et la vision périphérique pour guider le regard. Enfin, les sujets 7 et 8 (temps moyens de sélection : respectivement 4532 ms et 4004 ms) suivent les structures.

Ces résultats sont intéressants car, non seulement ils correspondent à ceux établis par classification hiérarchique (cf. paragraphe page 6.4), mais en plus ils prouvent la grande variabilité interindividuelle entre les sujets. Il est d'autant plus regrettable de n'avoir pu utiliser le logiciel de rejeu pour les cinq autres sujets. Néanmoins, cela montre que, même au sein d'une population homogène de sujets (doctorants en informatique ou informaticiens), il existe une importante variabilité interindividuelle en terme de stratégie de recherche, i.e. en terme de parcours oculaire adopté pour effectuer la même tâche.

Enfin, les structures semblent agir comme une forme de guidage visuel pour certains sujets. Mais nous avons pu constater que ce n'est pas systématique : les sujets 5 et 6 privilégient la saillance visuelle des items ressemblant à la cible, tandis que le sujet 9 ne suit pas les structures spatiales. À l'exception du sujet 9, tous adaptent leurs stratégies de recherche, en terme de parcours oculaire, à la structure spatiale de l'affichage.

Analyse intra-sujet (5 sujets) : phase de validation (2)

L'analyse manuelle des parcours oculaires n'a mis en évidence aucune différence entre les sujets 5, 6, 7, 8 et 9 en ce qui concerne la phase de validation du choix de la cible candidate impliquée lors du repérage visuel de cibles. Pour ces cinq sujets, les parcours observés lors de cette phase sont de trois types :

- l'analyse détaillée de la cible qui se traduit par plusieurs fixations sur celle-ci (environ 75% des cas) ;
- l'analyse d'autres photographies de la scène qui n'ont pas été visitées lors de la phase d'exploration (environ 15% des cas) ;
- l'analyse de photographies déjà visitées lors de la phase d'exploration (environ 10% des cas).

En revanche, nous avons observé des différences entre les structures aléatoire et elliptique et les structures radiale et matricielle dans le cas de l'analyse de photographies déjà visitées lors de la phase d'exploration. En effet, dans ce cas, les sujets parcourent intégralement les ellipses et les zones les plus denses des structures aléatoires, tandis qu'ils n'analysent que quelques photographies des structures radiales et matricielles. Ce résultat semble expliquer le temps moyen de parcours plus long ainsi que le nombre de fixations moyen plus élevé pour les structures elliptique et aléatoire, par rapport aux structures radiale et matricielle pendant cette seconde phase de la recherche (cf. tableau 6.10 page 139). Dans le prochain paragraphe, en nous basant sur les parcours oculaires les plus rapides, i.e., les parcours oculaires du sujet 5 (le plus rapide parmi les sujets 5, 6, 7, 8 et 9 et le second sur les 10 en terme de temps de sélection, derrière le sujet 3 ; cf. tableau 6.13 page 143), nous présentons les types de parcours oculaires les plus efficaces en fonction de la structure spatiale des affichages.

Les parcours oculaires les plus rapides

Pour les quatre structures (aléatoire, elliptique, radiale et matricielle), le sujet 5 a adopté une stratégie privilégiant l'analyse des items ressemblant à la cible (couleur, forme, par exemple) : c'est ce que nous avons qualifié de "stratégie par la saillance visuelle d'abord". En cas d'échec de cette stratégie, le sujet 5 a opté pour un parcours oculaire suivant les structures.

La figure 6.7 page 151 illustre cette stratégie pour la structure aléatoire. On constate en effet que le regard est guidé vers les zones les plus denses de la scène. La figure 6.8 page 152 illustre la stratégie adoptée par le sujet 5 sur les structures elliptiques, où la cible est saillante visuellement. On constate que le regard ne suit pas l'ellipse, mais est guidé vers les photographies ressemblant à la cible. La figure 6.9 page 152 illustre la stratégie adoptée par le sujet 5 sur les structures elliptiques où la cible n'est pas saillante visuellement. Le regard suit l'ellipse extérieure.

La figure 6.10 page 153 illustre la stratégie adoptée par le sujet 5 sur les structures matricielles. La scène étant difficile, les photographies contenues dans la scènes appartiennent à une collection homogène. La stratégie par la saillance d'abord est donc inefficace. On constate que le regard suit la matrice par balayages horizontaux et verticaux. La figure 6.11 page 153 illustre la stratégie adoptée par le sujet 5 sur les structures radiales. Le regard forme un cercle autour du centre de la scène. Ainsi, la vision périphérique permet au sujet de "voir" en une seule fixation un rayon entier.

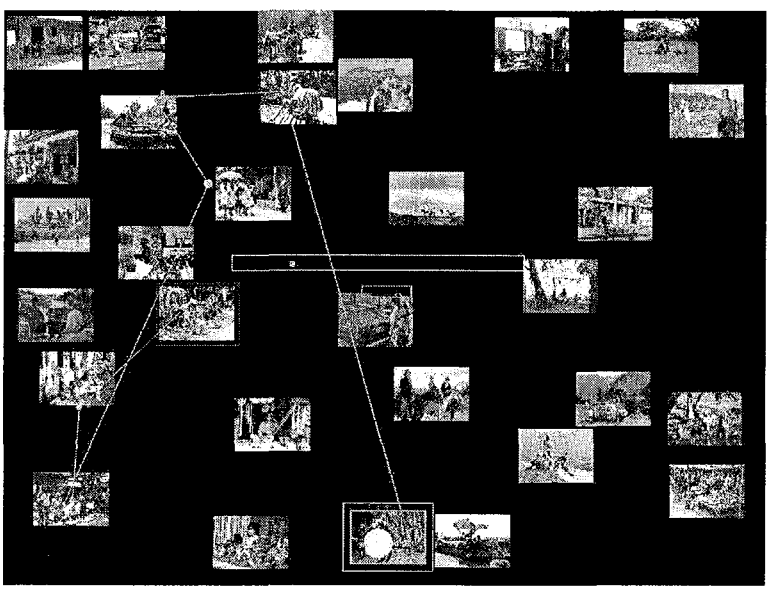


FIG. 6.7 – Sujet 5 : parcours oculaire type des structures aléatoires.
Les fixations se portent d'abord sur les zones les plus informatives de l'affichage. Le point de départ est matérialisé par le rectangle rouge. La dernière fixation est le gros point jaune sur la cible en vert (9 fixations au total). À noter que le logiciel de rejeu ne fournit pas, pour les scènes aléatoires, la représentation graphique épurée dont on dispose pour les trois autres structures, i.e., sur fond blanc, avec la cible en vert, la première fixation en bleu et la dernière en jaune.

6.5 Conclusions

Dans ce chapitre, consacré à l'analyse des parcours oculaires effectués lors de la tâche de repérage visuel de cibles, nous avons mené une troisième étude expérimentale visant à mettre en évidence l'influence de la structure des affichages sur les stratégies d'exploration visuelle adoptées par les utilisateurs. Dix sujets, ayant participé à la deuxième étude expérimentale, ont été sélectionnés en fonction de leur profil. Ils avaient pour consigne d'effectuer les tâches de repérage visuel dans deux conditions expérimentales : la condition F, où chaque couple (scène + cible) leur avait été présenté lors de la deuxième étude, et la condition NF, où aucun couple (scène + cible) ne leur avait été présenté antérieurement.

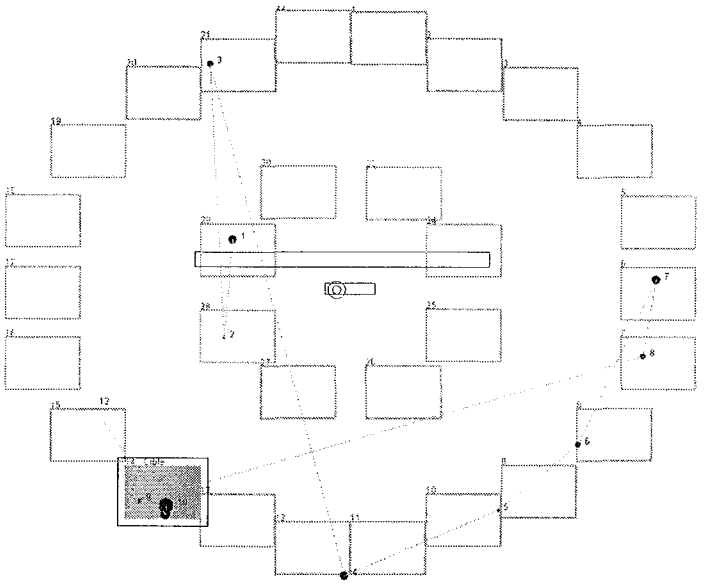


FIG. 6.8 – Sujet 5 : parcours oculaire type des structures elliptiques (1).
La scène est facile. Les fixations sont guidées par la saillance visuelle des items ressemblant à la cible (12 fixations au total). La cible est en vert. La première fixation sur la scène est en bleu, la dernière est en jaune.

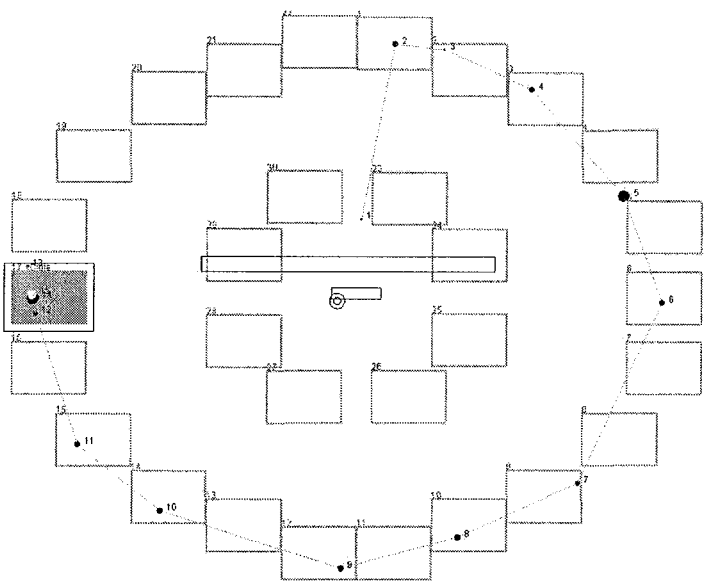


FIG. 6.9 – Sujet 5 : parcours oculaire type des structures elliptiques (2).
La scène est difficile. Les fixations sont guidées par la structure spatiale de la scène (15 fixations au total). La cible est en vert. La première fixation sur la scène est en bleu, la dernière est en jaune.

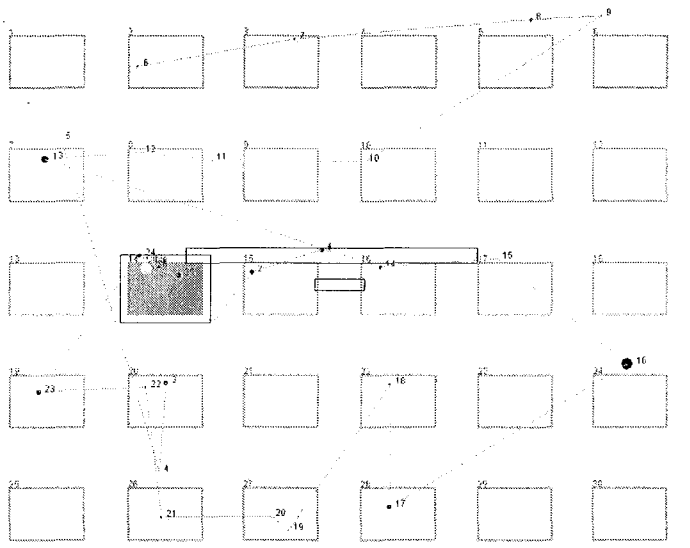


FIG. 6.10 – Sujet 5 : parcours oculaire type des structures matricielles.
La scène est difficile. Il y a 25 fixations au total. La cible est en vert. La première fixation sur la scène est en bleu, la dernière est en jaune.

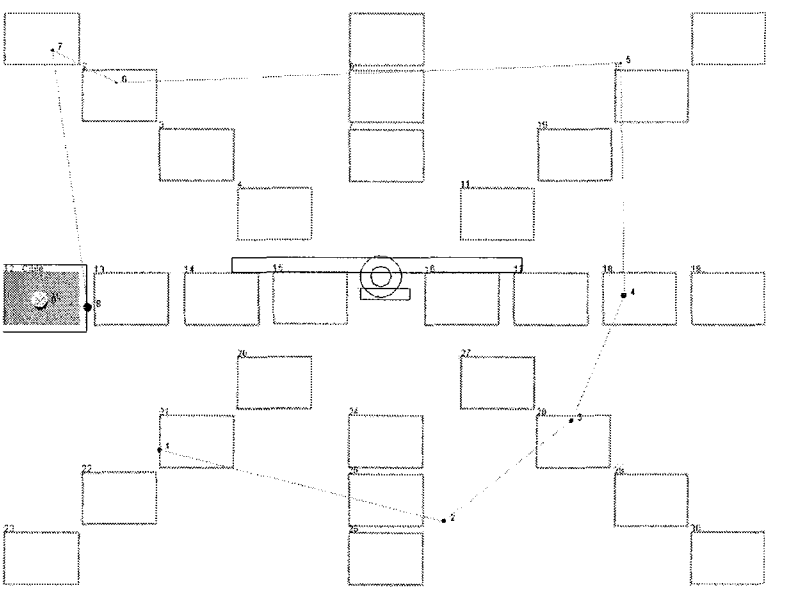


FIG. 6.11 – Sujet 5 : parcours oculaire type des structures radiales.
La scène est classée difficile. La première fixation sur la scène est en bleu, la dernière est en jaune (10 fixations au total).

Temps et précision de la sélection des cibles L'analyse des données, à savoir les temps et la précision de sélection des cibles, pour les dix sujets ayant participé à l'expérimentation a montré les limites de la mémorisation à long terme des tâches visuelles : les sujets étaient meilleurs dans la condition NF que dans la condition F, bien que la plupart des sujets ont reconnu les cibles dans la condition F. Des études spécifiques sur ce point sont nécessaires pour expliquer pourquoi les cibles ont été mémorisées tandis que leur position dans la scène ne l'a pas été. Cette analyse quantitative des performances des sujets a permis en outre de valider les choix de conception des protocoles expérimentaux concernant le niveau de difficulté de la tâche (niveau de difficulté des scènes présentées aux sujets) et le type de contenu des photographies de la scène (objets *versus* paysages).

Exploration *versus* validation : deux étapes successives dans le repérage visuel de cibles L'analyse quantitative des données recueillies à l'aide de l'eye-tracker pour cinq des dix sujets ayant participé à la troisième expérimentation a permis de mettre en évidence l'existence de deux phases distinctes dans le repérage visuel de cibles : une première phase d'exploration de la scène jusqu'à atteindre la cible, suivie d'une deuxième phase de validation/vérification du choix de la cible candidate. La première est plus longue que la deuxième. Il s'agit de l'exploration rapide d'un grand nombre de photographies, où les sujets semblent percevoir forme générale (grâce aux contrastes lumineux) ainsi que principales caractéristiques (couleurs, directions) des photographies explorées. La deuxième phase de recherche est plus courte. Il s'agit de la validation de la détection de la cible candidate par comparaison avec d'autres items contenus dans la scène. Lors de cette phase, les sujets semblent exploiter les détails des photographies pour valider le résultat de la phase précédente : les fixations sont beaucoup plus longues que dans la première phase et une même photographie peut faire l'objet de plusieurs fixations.

Différentes stratégies pour le repérage visuel de cibles L'analyse manuelle détaillée des parcours oculaires individuels effectuée, jointe à l'analyse quantitative des données oculaires par sujet, pour cinq des dix sujets ayant participé à la troisième expérimentation a permis de mettre en évidence des différence entre les sujets, en terme de stratégie d'exploration visuelle. Un premier groupe de sujets suit les structures spatiales pour explorer la scène. Un autre groupe de sujets adopte une stratégie d'exploration "par la saillance visuelle d'abord", i.e., par exploration des items ressemblant à la cible d'abord, pour s'appuyer ensuite sur les structures spatiales si la première stratégie mise en œuvre a échoué. Enfin, un dernier groupe fait une exploration minutieuse de la *quasi* totalité des items avant d'en sélectionner un. Ce résultat important de notre étude prouve que les stratégies d'exploration visuelle d'affichages complexes dépendent du profil des utilisateurs. Autrement dit, même au sein d'une population homogène de sujets (étudiants, ingénieurs et chercheurs en informatique, tous entre 24 et 29 ans), il existe une grande variabilité interindividuelle en terme de comportement et, par suite, en terme de stratégie adoptée de réalisation de tâches visuelles comme la détection de cibles.

Il convient de noter que les stratégies d'exploration des affichages par la "saillance visuelle des items d'abord" semblent les plus efficaces en termes de temps de sélection des cibles, de confort et de fatigue.

Enfin, si les stratégies adoptées pour le repérage visuel de cibles diffèrent d'un groupe de sujets à l'autre, en revanche, on retrouve, pour un groupe de sujets donné, le même type de stratégie d'exploration visuelle quelle que soit la structure des affichages. Ainsi, les sujets qui adoptent une stratégie par la saillance visuelle des items d'abord au sein des structures elliptiques, semblent adopter la même stratégie pour les structures aléatoires, radiales et matricielles. De même, d'autres sujets "suivent" la structure qu'elle soit elliptique, matricielle ou radiale. Ceci pourrait expliquer, en partie, pourquoi les différences observées entre les structures, concernant la durée de la phase d'exploration, ne sont pas significatives dans cette troisième étude (sauf pour les structures elliptique et matricielle). En effet, l'échantillon de sujets réduit (10) par rapport à la deuxième étude (24) implique une plus grande influence des variabilités interindividuelles sur les performances globales des sujets, et donc, sur les performances des sujets par structure. Il faut en outre tenir compte de l'influence de l'apprentissage de la tâche, probablement plus marqué pour la structure elliptique, plus insolite pour certains sujets, que pour la structure matricielle, très couramment utilisée au sein d'affichages de type "browser", par exemple. Enfin, le fait que la phase de validation du choix de la cible est nettement plus longue pour la structure elliptique que pour les structures matricielle et radiale, alors que la durée de la première phase est sensiblement plus courte pour la structure elliptique que pour les autres structures (différence statistiquement significative avec la structure matricielle) peut expliquer pourquoi les sujets n'ont pas obtenus les meilleurs temps de sélection pour cette structure lors de la seconde expérimentation (cf. section 5.11 page 120).

Chapitre 7

Conclusions et perspectives

L'objectif initial de notre recherche était d'évaluer l'utilité de la multimodalité parole + graphique en tant qu'expression complémentaire du système en situation d'interaction avec des visualisations 2D. Le constat du volume et de la densité croissants des informations affichées simultanément, d'une part, et l'intégration quasi-systématique de la parole en sortie aux interfaces des applications graphiques grand public, d'autre part, expliquent pourquoi nous avons choisi d'explorer l'interaction Homme-Machine multimodale qui associe la parole au graphique dans une même intervention du système.

Nous avons adopté une démarche expérimentale afin de déterminer l'influence d'indications spatiales formulées oralement sur l'efficacité et la satisfaction d'utilisateurs potentiels dans des activités d'exploration visuelle d'affichages complexes. La situation d'interaction multimodale que nous avons retenue est le repérage visuel de cibles car il intervient dans de nombreuses activités interactives, comme la navigation dans les grands ensembles d'informations, la navigation dans les réalités virtuelles, la navigation sur Internet ou l'inspection visuelle.

Le programme expérimental envisagé comportait initialement deux étapes. D'abord, en l'absence de résultats scientifiques suffisants sur la recherche visuelle en situation d'interaction Homme-Machine, nous avons mené une étude exploratoire fondée sur des hypothèses *a priori* avec, pour objectif, la comparaison de trois modes de présentation de la cible : visuel, oral et multimodal. Cette étude a montré, principalement, que les présentations multimodales, i.e., présentations visuelles de la cible accompagnées d'une indication spatiale absolue de sa localisation dans la scène, sont les plus efficaces en termes de temps de recherche et de précision de la sélection de la cible. Une deuxième étude s'est avérée nécessaire pour consolider ces premiers résultats. Par ailleurs, nous avons modifié partiellement le protocole expérimental de cette deuxième étude par rapport à celui élaboré pour la précédente, de façon à pouvoir évaluer l'influence de la structure spatiale des affichages sur les performances des sujets. Cette étude a non seulement permis d'asseoir les résultats préliminaires concernant la multimodalité parole + présentation visuelle, mais elle a permis, en outre, de vérifier l'hypothèse selon laquelle l'organisation spatiale des items contenus dans les affichages a une influence sur les performances des sujets. Pour cette étude, nous avons limité les mesures, comme pour l'étude préliminaire, au temps et à la précision de la sélection des cibles.

C'est pour comprendre et interpréter les résultats prometteurs de cette deuxième étude que nous avons élaboré une troisième et dernière étude expérimentale centrée sur les stratégies d'exploration visuelle adoptée par les utilisateurs pour les tâches de repérage visuel de cibles. Dans cette étude, en plus de la précision et des temps de sélection des cibles réalisés par les sujets, nous avons recueilli leurs fixations oculaires afin d'analyser leurs stratégies d'exploration visuelle des affichages.

Présentations visuelles *versus* présentations multimodales Le principal objectif des deux premières études expérimentales résidait dans la démonstration de l'apport spécifique des indications spatiales contenues dans les présentations multimodales pour le repérage visuel de cible, par rapport aux présentations visuelles des cibles isolées. Les différentes analyses statistiques portant sur la comparaison des performances des sujets en fonction du mode de présentation de la cible, d'une part, et l'analyse des debriefings post-expérimentation, d'autre part, ont mis en évidence les bénéfices retirés par les utilisateurs, non seulement en termes de précision et de rapidité, mais aussi en terme de confort, dans la réalisation de tâches visuelles comme le repérage de cibles.

Les présentations multimodales permettent aux sujets de tirer avantage de l'information spécifique véhiculée par chacune des modalités :

- la présentation visuelle contenue dans la présentation multimodale "montre" la cible, ses caractéristiques, comme sa forme, sa nature, ou sa couleur, par exemple ;
- le message sonore d'indication spatiale permet d'indiquer précisément et de réduire la zone de recherche.

Ces informations permettent aux sujets de repérer la cible sans explorer la scène dans son intégralité et sans avoir recours à la description linguistique d'un détail, processus cognitif qui entraîne une charge de travail plus importante (cf. les résultats obtenus lors de l'étude préliminaire concernant les présentations exclusivement orales). Les présentations multimodales semblent permettre aux sujets de moins se focaliser, lors de la présentation de la cible, sur les détails la caractérisant. Les messages sonores d'indication spatiale contenus dans les présentations multimodales réduisent de façon importante la zone de recherche : elle est divisée par neuf, approximativement, en raison même de la nature des messages sonores. Au sein d'une zone de recherche de cette taille, moins de détails caractéristiques de la cible s'avèrent nécessaires pour la différencier des non cibles.

Par ailleurs, nous avons démontré que l'apport des messages sonores d'indications à caractère spatial varie en fonction du niveau de difficulté de la scène, et donc, en fonction du niveau de difficulté de la tâche. L'apport est plus important sur des scènes où les items sont proches visuellement et où le niveau de détail du contenu de chaque photographie est plus fin (cf. l 'analyse statistique concernant les différences entre les niveaux de difficulté des scènes). Si l'apport des messages multimodaux se révèle moindre sur des scènes dont les constituants sont hétérogènes ou sans complexité de détail, il se révèle, en revanche, très important sur des scènes dont les composants sont homogènes et avec une grande complexité de détail.

Enfin, les choix de conception du protocole expérimental se sont avérés judicieux, en termes de type de contenu des photographies (objets *versus* paysages), niveau de difficulté des scènes (facile,

moyen, difficile) et structure spatiale des affichages (aléatoire, elliptique, radiale et matricielle), dans la mesure où nous avons observé, pour chacune de ces variables, des différences statistiquement significatives. Ces résultats impliquent la nécessité de tenir compte de ces différents critères (contenu des affichages, difficulté, organisation spatiale) lors d'analyses expérimentales de tâches visuelles.

Exploration *versus* vérification : deux phases successives du repérage visuel de cibles Lors de la troisième étude expérimentale, nous avons pu mettre en évidence l'existence de deux phases successives lors du repérage visuel d'une cible. Il s'agit de l'exploration visuelle de la scène jusqu'à atteindre une première fois la cible du regard *versus* la vérification du choix de la cible candidate. La phase d'exploration de la scène est plus courte que la phase de vérification. En outre, la phase d'exploration visuelle présente les caractéristiques suivantes : plus de photographies sont "visitées" et plus rapidement que lors de la phase de vérification (fixations oculaires plus longues).

L'analyse quantitative des données recueillies à l'aide l'eye-tracker lors de la troisième étude n'a révélé aucune différence statistiquement significative entre les structures, sauf pour la première phase d'exploration visuelle. C'est pour la structure elliptique que la distance moyenne parcourue (en nombre de pixels) est la plus courte pendant cette phase. Cette structure s'avère plus efficace en terme de temps moyen de sélection des cibles que la structure matricielle, et ce résultat est également statistiquement significatif. Néanmoins, nous avons pu constaté que les structures matricielles et radiales sont les moins efficaces lors de la phase d'exploration visuelle ; probablement en raison de leurs discontinuités. En effet, plus l'organisation spatiale des affichages est continue et régulière en terme d'allure globale, plus elle semble efficace. Or, les structures radiales étaient composées de 8 rayons, les structures matricielles étaient constituées d'une matrice à 5 lignes et 6 colonnes, tandis que les structures elliptiques ne comptaient que 2 ellipses. Donc, on peut supposer que la régularité *versus* irrégularité de l'organisation spatiale d'un affichage en affecte l'exploration visuelle, en terme de rapidité, efficacité et confort de l'utilisateur. Davantage d'études semblent nécessaires pour conclure sur ce point précis. Il convient de noter également que les structures aléatoires entraînent des performances intermédiaires de la part des sujet, i.e., entre celles observées pour les structures radiales ou matricielles et celles observées pour les structures elliptiques.

Cette tendance observée entre les structures lors de la phase d'exploration de la scène s'inverse lors de la phase de vérification : les structures matricielles et radiales sont les plus efficaces par rapport aux structures elliptiques pour lesquelles les performances des sujets sont intermédiaires, et les structures aléatoires. Le repérage au sein des structures radiales et matricielles est certes plus long, mais il semble plus sûr.

Les différentes stratégies de recherche visuelle : analyse et comparaison L'analyse manuelle détaillée des parcours oculaires individuels recueillis lors de la troisième étude expérimentale a montré des différences importantes entre les stratégies adoptées par les utilisateurs, conduisant, par suite, à la définition des profils utilisateur différents. Trois stratégies d'exploration visuelle distinctes émergent. La première, la plus efficace en termes de temps et de précision

de la sélection des cibles, est basée sur la saillance visuelle des items ressemblant à la cible : les utilisateurs explorent d'abord les items semblables, visuellement, à la cible, pour ensuite, en cas d'échec de la recherche, utiliser les structures spatiales pour explorer le reste de la scène. La deuxième, la moins efficace en termes de temps et de précision de la sélection des cibles, consiste à explorer minutieusement un nombre important d'items sans s'appuyer sur la structure spatiale de l'affichage : les utilisateurs effectuent, dans ce cas, de nombreux aller-retour sur des items déjà explorés ; ce qui rend la recherche plus lente et plus fatigante par rapport à la première stratégie. La dernière, intermédiaire en termes de performances par rapport aux deux autres, consiste à suivre la structure spatiale des affichages jusqu'à atteindre la cible.

Ce résultat est important puisqu'il impose de distinguer au moins trois catégories d'individus pour la réalisation de tâches visuelles comme le repérage de cibles :

- ceux qui se laissent guider par la structure spatiale des affichages pour explorer une scène ;
- ceux qui se concentrent, en priorité, sur les propriétés graphiques de l'élément recherché pour n'explorer que certains items ressemblant à la cible ;
- ceux qui adoptent une stratégie de recherche "aléatoire", n'étant guidés ni par l'organisation spatiale des affichages, ni par la saillance visuelle des items ressemblant à la cible.

Même au sein d'une population homogène de sujets, nous avons pu identifier d'importantes différences entre les individus, à la fois en terme de rapidité d'exécution de la tâche, mais aussi en terme de stratégie d'exploration visuelle. En l'absence d'un modèle de l'utilisateur, la conception d'une interface 2D interactive pour la recherche visuelle d'items, efficace et confortable pour une vaste population d'individus, semble donc impossible.

Néanmoins, il est fort probable que d'autres facteurs, comme la dextérité, l'acuité visuelle, ou l'utilisation avancée de la vision périphérique, interviennent dans le repérage visuel de cibles. Par ailleurs, nous avons pu mettre en évidence l'influence importante de l'apprentissage de la tâche pour de telles activités. L'apprentissage à court terme a permis à tous les sujets d'améliorer leurs performances au cours d'une même passation. L'apprentissage à long terme a permis à certains d'entre-eux d'améliorer celles-ci entre deux expérimentations. Ce résultat est compatible avec ceux présentés dans [Drury, 1992; Gramopadhye et Madhani, 2001] : les performances des sujets en terme d'efficacité de recherche augmentent avec l'entraînement. L'utilisateur régulier ou averti d'une telle application sera plus efficace que l'utilisateur occasionnel quelle que soit la priorité qu'il accorde entre saillance visuelle et structure spatiale des éléments.

Perspectives Les perspectives de ce travail sont nombreuses. D'une part, davantage de travaux sont nécessaires sur un groupe de sujets plus proches de la moyenne. En effet, en voulant "couvrir" le maximum de stratégies de recherche visuelle, nous nous sommes heurtés aux problèmes créés par une variabilité interindividuelle importante. Il serait intéressant de reproduire partiellement la troisième étude avec une population homogène en termes de performances et en excluant les cas extrêmes. Une telle étude serait susceptible de mettre en évidence des différences plus nettes entre les structures aléatoire, elliptique, matricielle et radiale, entre autres.

D'autre part, dans ce travail, nous avons considéré des structures spatiales simples des affichages. Les scènes présentées aux sujets comprenaient 30 items, photographies d'objets ou de paysages, organisés selon une structure soit aléatoire, soit radiale, soit matricielle, soit elliptique.

Les résultats obtenus, pourvoyeurs d'informations précieuses quant aux stratégies d'exploration visuelle adoptées par les utilisateurs au sein de structures 2D simples, doivent être considérés comme le premier pas de l'analyse plus ambitieuse des stratégies adoptées par les utilisateurs au sein de visualisations 2D plus complexes. En effet, nous envisageons de poursuivre nos recherches vers l'étude des tâches de navigation au sein de grands ensembles d'informations 2D interactifs en nous appuyant sur les résultats obtenus concernant le repérage visuel de cibles.

L'étude envisagée consisterait à définir, puis évaluer d'un point de vue ergonomique, l'efficacité des techniques de visualisation interactive pour la navigation vers un élément en utilisant une, voire plusieurs, organisations spatiales simples comme la structure elliptique, par exemple. Les tâches de repérage visuel de cibles semblent facilitées au sein des structures elliptiques que nous avons utilisées, car celles-ci présentent le double avantage d'être régulières, i.e., les éléments sont collés les uns aux autres, et aérées, i.e., elles laissent beaucoup d'espace autour des deux ellipses, ce qui est de nature à renforcer leur influence sur les mouvements oculaires de l'utilisateur et, par conséquent, les effets du guidage visuel qui en résultent. Cette intuition demande à être vérifiée. La tâche expérimentale envisagée pourrait consister à naviguer vers une photographie connue visuellement au sein d'une banque de photographies.

Annexe A

Étude préliminaire

L'annexe A contient les images présentées aux sujets, la liste des messages sonores de l'étude préliminaire (cf. infra sections A.1 et A.2), ainsi que les questionnaires post-expérimentation présentés au sujets (cf. infra figures A.3 et A.4).

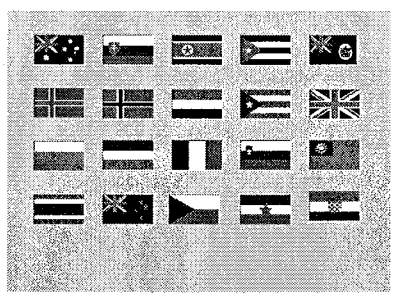


Image 1(1)

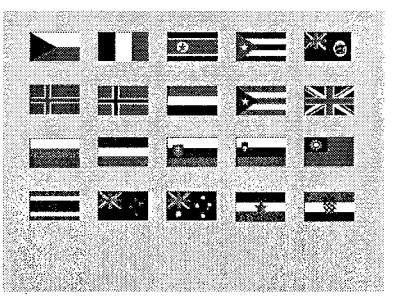


Image 2(1)

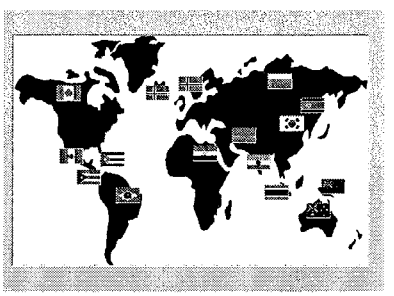


Image 3(1)

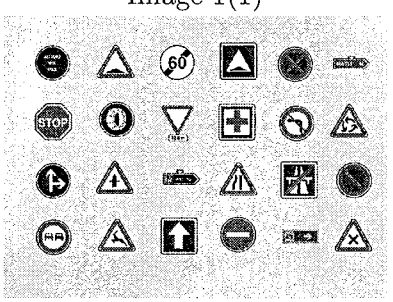


Image 4(1)

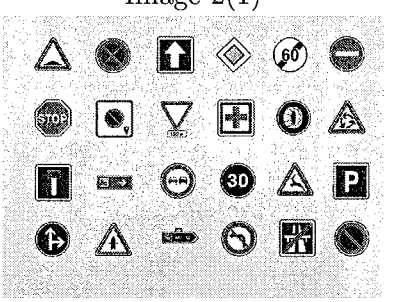


Image 5(1)

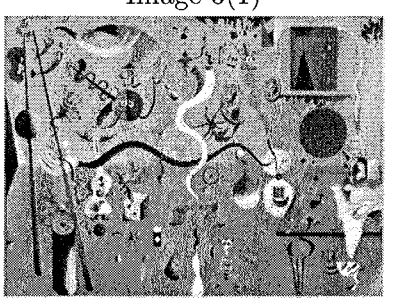


Image 6(1)

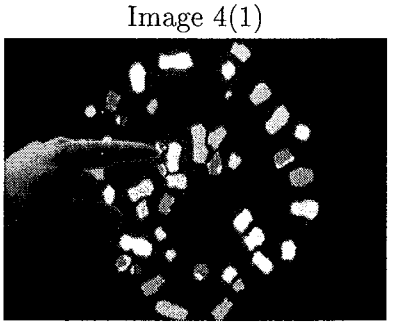


Image 7(1)

Image 8(1)

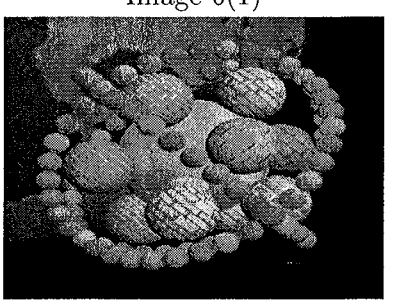


Image 9(1)

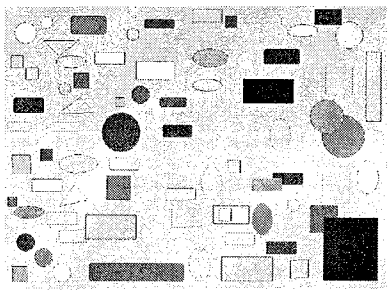


Image 10(1)

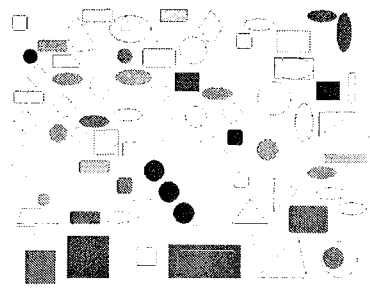


Image 11(1)

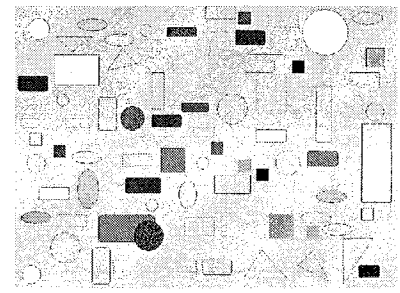


Image 12(1)

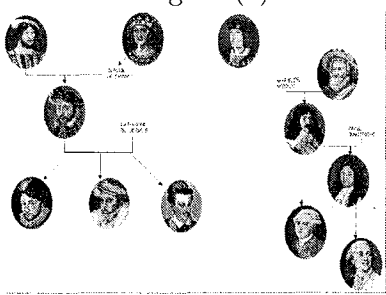


Image 13(1)

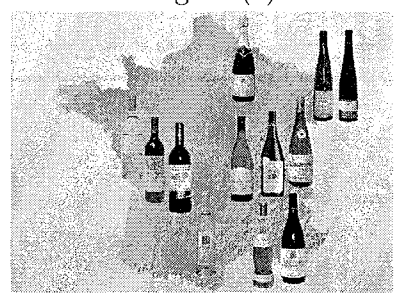


Image 14(1)

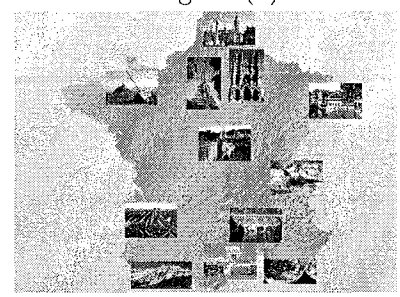


Image 15(1)

Image 16(1)

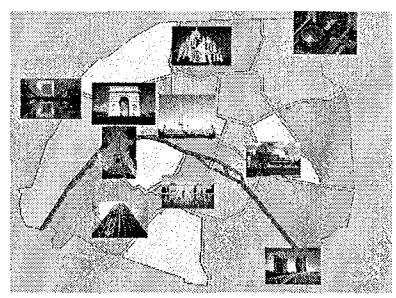


Image 17(1)

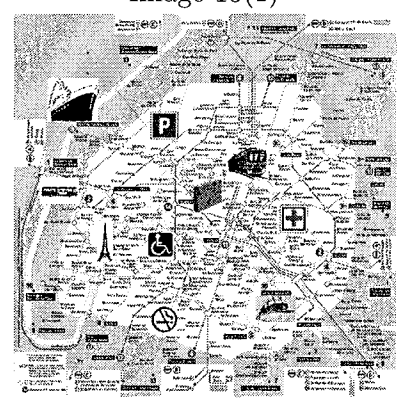


Image 18(1)

Image 1(2)

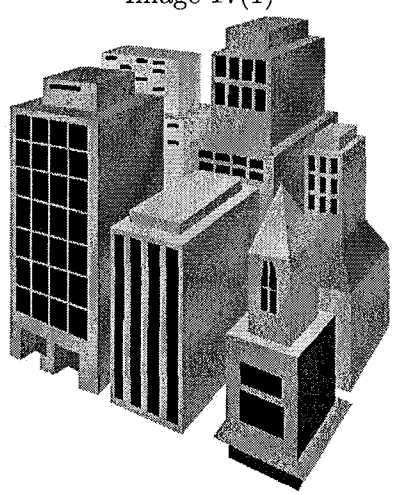


Image 2(2)

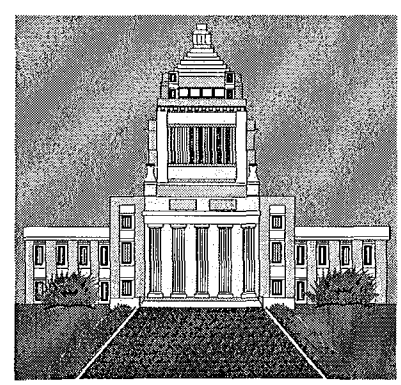


Image 3(2)

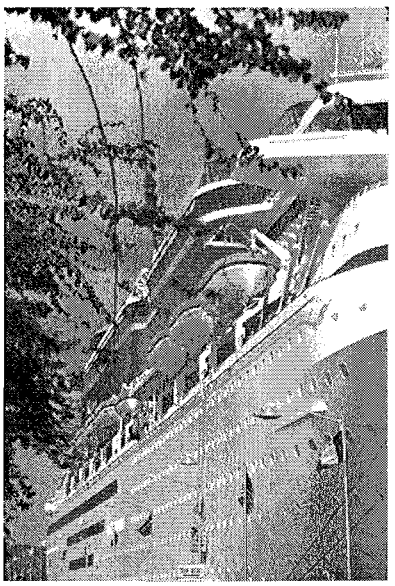


Image 4(2)

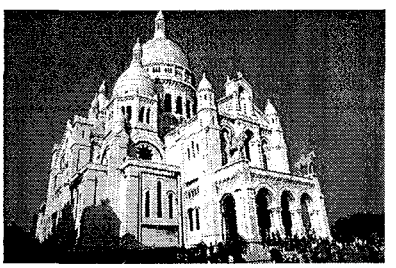


Image 5(2)

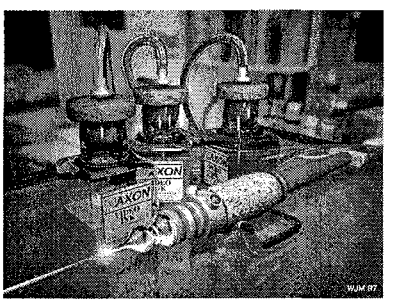


Image 6(2)

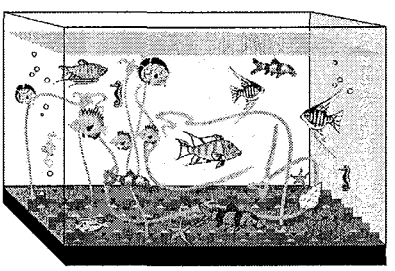


Image 7(2)

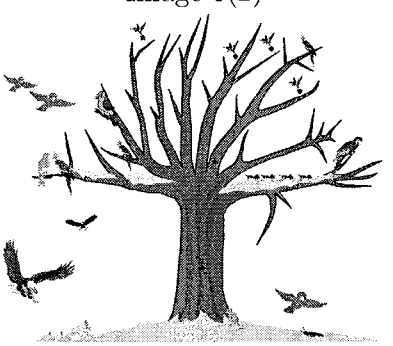


Image 8(2)

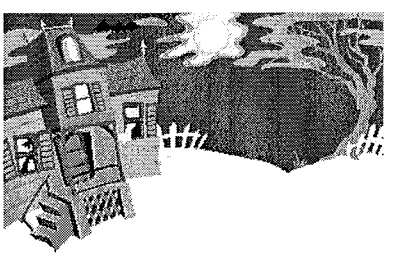


Image 9(2)

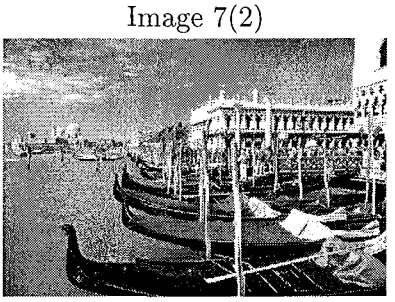


Image 10(2)

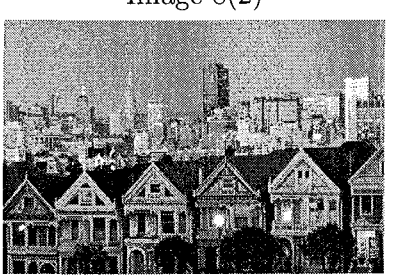


Image 11(2)

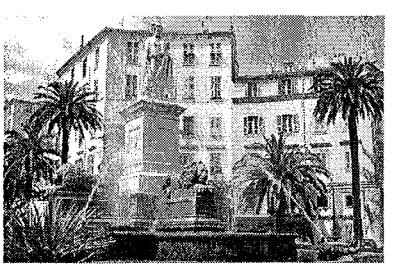


Image 12(2)

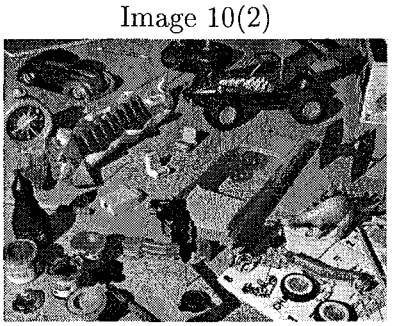


Image 13(2)

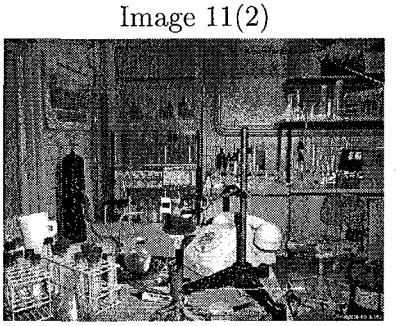


Image 14(2)

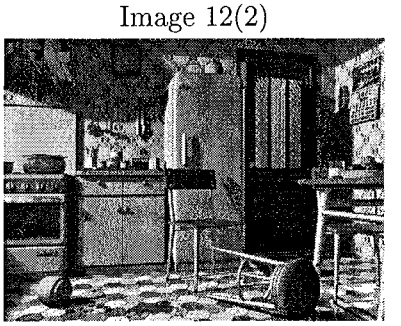


Image 15(2)

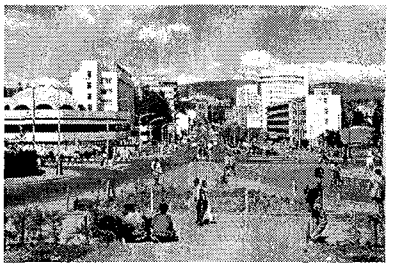


Image 16(2)

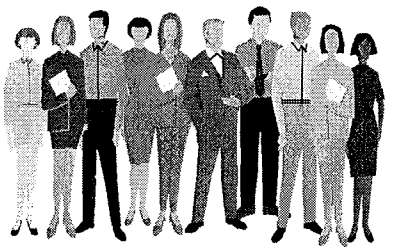


Image 17(2)

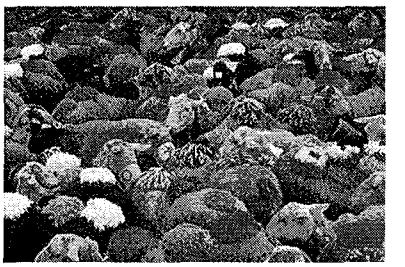


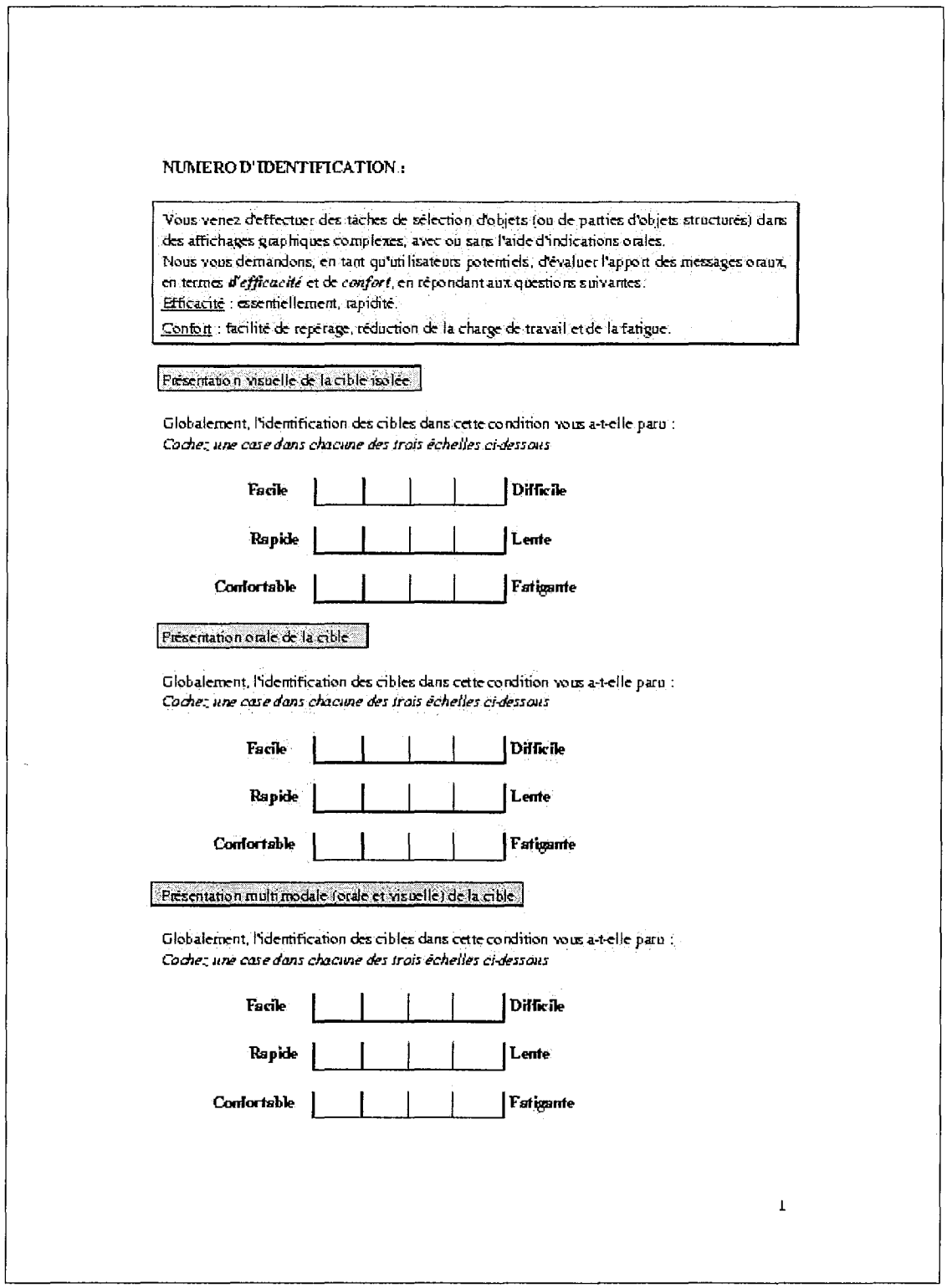
Image 18(2)

Image 1	"Dans la dernière ligne, le second drapeau."
Image 2	"Dans la première ligne, le drapeau à droite du drapeau français."
Image 3	"Le drapeau du Mexique."
Image 4	"Dans la dernière colonne, le panneau "Sens giratoire"."
Image 5	"Dans la troisième ligne, le panneau "Gibier"."
Image 6	"Le dé à jouer."
Image 7	"En bas, la tache brune à droite de la tache bleu clair."
Image 8	"En haut à droite, la forme en bois jaune."
Image 9	"En bas, la boule verte au premier plan."
Image 10	"En bas à droite, le carré mauve."
Image 11	"Le cercle à droite du disque orange."
Image 12	"En haut à gauche, le triangle au-dessus du rectangle vert clair."
Image 13	"À droite, le Roi Soleil fils d'Anne dAutriche."
Image 14	"Le Côtes de Provence."
Image 15	"Les Alpes."
Image 16	"Le Roi des Animaux."
Image 17	"La tour Montparnasse."
Image 18	"À l'Ouest de Paris, le bus."

FIG. A.1 – Messages pour la classe d'images abstraites (1)

Image 1	"À gauche de la pomme, la poire."
Image 2	"À droite, le toit du bâtiment au premier plan."
Image 3	"Au premier étage, la fenêtre la plus à gauche."
Image 4	"Le canot de sauvetage au premier plan."
Image 5	"Le clocheton du petit dôme."
Image 6	"Le bouton rouge du pistolet à peinture."
Image 7	"Sur le fond de l'aquarium, l'étoile de mer en arrière-plan."
Image 8	"Au pied de l'arbre, le mulot."
Image 9	"La fenêtre au-dessus de l'escalier."
Image 10	"À l'horizon, le plus gros des dômes."
Image 11	"En bas à droite, le porche éclairé."
Image 12	"La cheminée à gauche de la mansarde."
Image 13	"À gauche de la voiture rose, la gomme."
Image 14	"Au milieu en bas, le crayon rouge et bleu."
Image 15	"En haut à droite, le calendrier."
Image 16	"À droite, les deux personnes en train de courir."
Image 17	"Sur la droite, la cravate de l'homme en pantalon noir."
Image 18	"Au fond, la tête du mouton."

FIG. A.2 – Messages pour la classe d'images réalistes (2)



NUMERO D'IDENTIFICATION :

Vous venez d'effectuer des tâches de sélection d'objets (ou de parties d'objets structurés) dans des affichages graphiques complexes, avec ou sans l'aide d'indications orales.
Nous vous demandons, en tant qu'utilisateurs potentiels, d'évaluer l'apport des messages oraux, en termes ***d'efficacité*** et de ***confort***, en répondant aux questions suivantes.
Efficacité : essentiellement, rapidité.
Confort : facilité de repérage, réduction de la charge de travail et de la fatigue.

Présentation visuelle de la cible isolée

Globalement, l'identification des cibles dans cette condition vous a-t-elle paru :
Cochez une case dans chacune des trois échelles ci-dessous

Facile		**Difficile**
Rapide		**Lente**
Confortable		**Fatigante**

Présentation orale de la cible

Globalement, l'identification des cibles dans cette condition vous a-t-elle paru :
Cochez une case dans chacune des trois échelles ci-dessous

Facile		**Difficile**
Rapide		**Lente**
Confortable		**Fatigante**

Présentation multimodale (orale et visuelle) de la cible

Globalement, l'identification des cibles dans cette condition vous a-t-elle paru :
Cochez une case dans chacune des trois échelles ci-dessous

Facile		**Difficile**
Rapide		**Lente**
Confortable		**Fatigante**

1

FIG. A.3 – Deuxième questionnaire (page 1).

- Les messages oraux vous ont-ils facilité le repérage des cibles ? oui non

  Si non, pourquoi ?

- Les messages oraux vous ont-ils gêné ? oui non

  Si oui, pourquoi ?

- Classez les trois modes de présentation des cibles (visuel, oral, multimodal) par ordre décroissant de préférence :

  1. .....

  2. .....

  3. .....

- Classez les trois modes de présentation des cibles (visuel, oral, multimodal) par ordre décroissant d'efficacité :

  1. .....

  2. .....

  3. .....

Merci du temps que vous nous avez accordé !

2

FIG. A.4 – Deuxième questionnaire (page 2).

Annexe B

Deuxième étude

L'annexe B contient les questionnaires post-expérimentation présentées au sujets (cf. infra figures B.1 à B.6), de même que l'évolution des performances des sujets en fonction de l'ordre dans lequel ils ont effectué les tâches de repérage, i.e., PV-PM ou PM-PV (cf. infra figures B.7 et B.8).

PREMIER QUESTIONNAIRE
SUJET NUMÉRO :

Cette expérimentation préservera votre anonymat.

Les informations sollicitées dans ce questionnaire nous serviront à préciser votre profil en tant qu'utilisateur de logiciels grand public (connaissances générales et expérience informatique).

M. ☐ Mme ☐ Mlle ☐

NOM :
Prénom : Âge :
Tél. : Mail :

CONNAISSANCES GÉNÉRALES

Dernier diplôme obtenu : Année :
Diplôme en cours de préparation : Année :
Profession :

COMPÉTENCES INFORMATIQUES

Quelles sont vos activités informatiques ? Précisez le nombre d'heures par semaine.

☐ Internet ..... heures par semaine
☐ Saisie/consultation de données ..... heures par semaine
☐ Logiciels grand public ..... heures par semaine
☐ Logiciels graphiques (Dessin, CAO, etc.) ..... heures par semaine
☐ Conception et développement d'applications ..... heures par semaine
☐ Jeux ..... heures par semaine
☐ Autres - Préciser .... heures par semaine

FIG. B.1 – Premier questionnaire : Statut du participant.

DEUXIÈME QUESTIONNAIRE
SUJET NUMÉRO :

Vous venez d'effectuer des tâches de sélection dans des affichages graphiques complexes dans la condition visuelle.

Nous vous demandons, en tant qu'utilisateurs potentiels, d'évaluer l'efficacité de la tâche de repérage dans cette condition, en répondant au questionnaire suivant.

PRÉSENTATION VISUELLE DE LA CIBLE

Le repérage des cibles vous a semblé :

Difficile	_____	Facile
Lent	_____	Rapide
Fatiguant	_____	Confortable
Ennuyeux	_____	Amusant

Il vous est arrivé d'hésiter entre plusieurs cibles :

- ☐ Jamais
- ☐ Moins de cinq fois
- ☐ Entre cinq et 15 fois
- ☐ Plus de 15 fois

Il vous est arrivé de choisir une cible au hasard :

- ☐ Jamais
- ☐ Une ou deux fois
- ☐ Entre trois et cinq fois
- ☐ Plus de cinq fois

FIG. B.2 – Deuxième questionnaire : Évaluation de la condition visuelle (page 1).

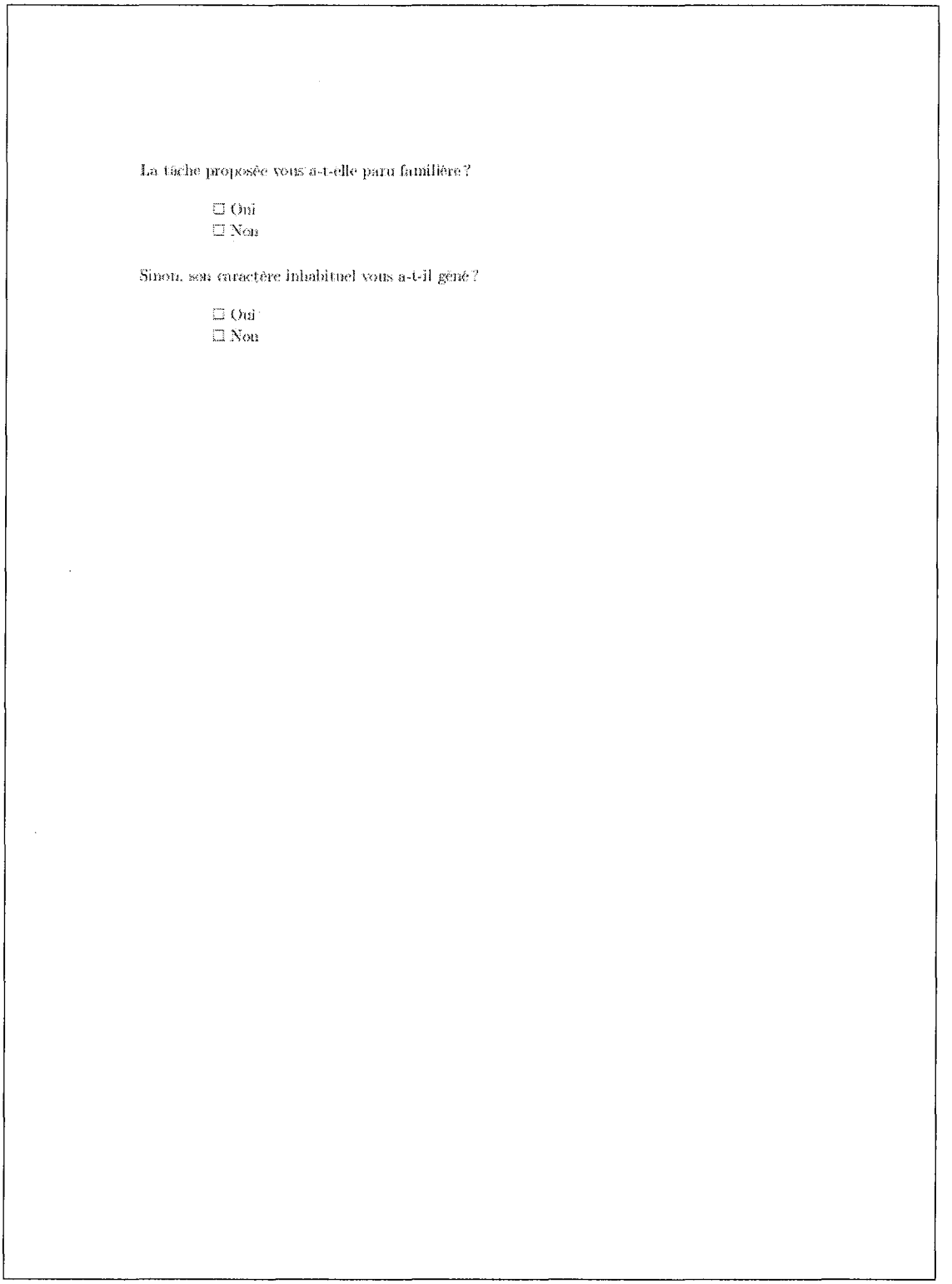


FIG. B.3 – Deuxième questionnaire : Évaluation de la condition visuelle (page 2).

TROISIÈME QUESTIONNAIRE
SUJET NUMÉRO :

Vous venez d'effectuer des tâches de sélection dans des affichages graphiques complexes dans la condition multimodale.

Nous vous demandons, en tant qu'utilisateurs potentiels, d'évaluer la contribution éventuelle des messages sonores pour les tâches de repérage, en répondant au questionnaire suivant.

PRÉSENTATION MULTIMODALE DE LA CIBLE

Le repérage des cibles vous a semblé :

Difficile	_____|_____|_____|_____|_____	Facile
Lent	_____|_____|_____|_____|_____	Rapide
Fatigant	_____|_____|_____|_____|_____	Confortable
Ennuyeux	_____|_____|_____|_____|_____	Amusant

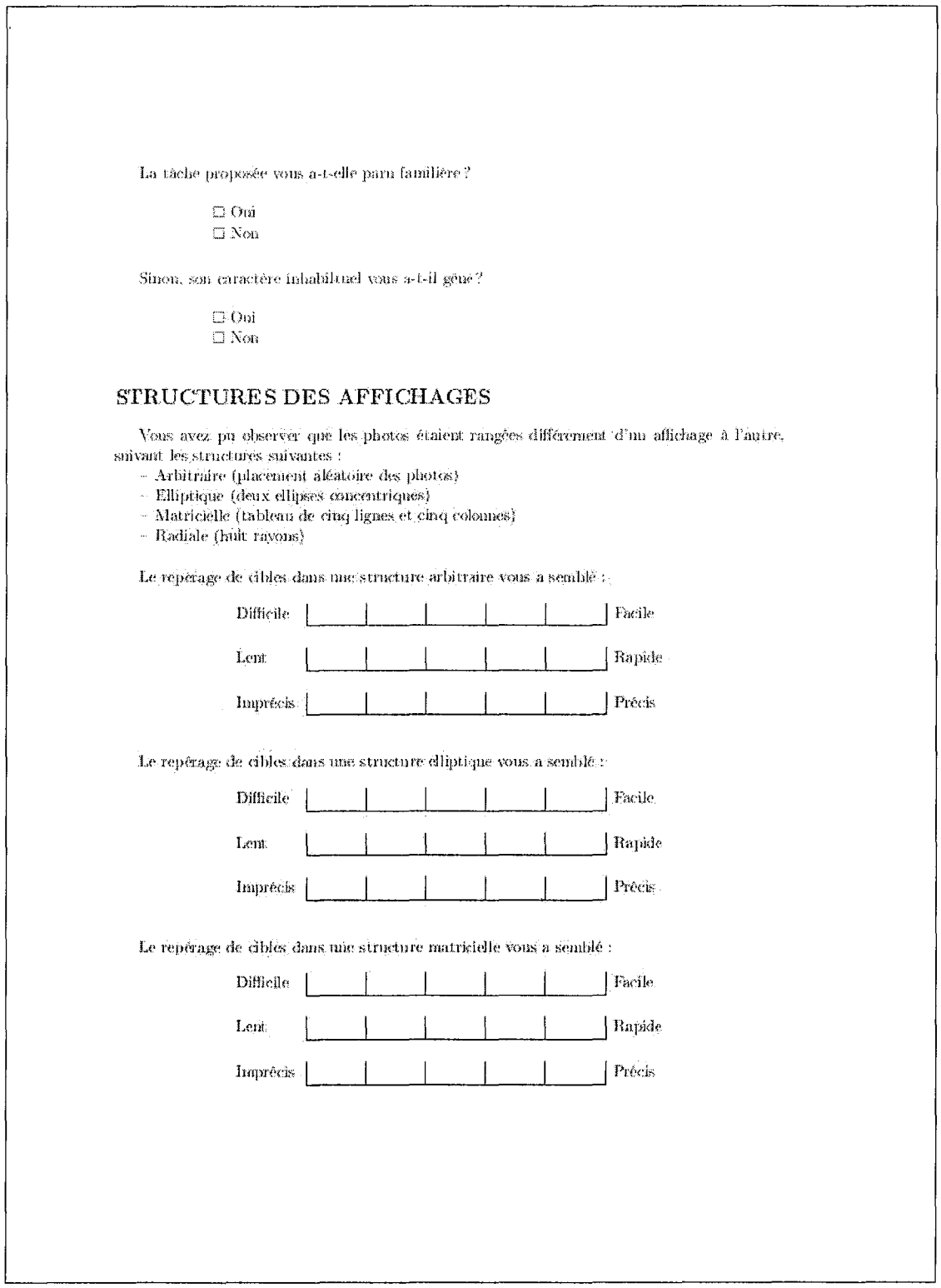
Il vous est arrivé d'hésiter entre plusieurs cibles :

- ☐ Jamais
- ☐ Moins de cinq fois
- ☐ Entre cinq et 15 fois
- ☐ Plus de 15 fois

Il vous est arrivé de choisir une cible au hasard :

- ☐ Jamais
- ☐ Une ou deux fois
- ☐ Entre trois et cinq fois
- ☐ Plus de cinq fois

FIG. B.4 – Toisième questionnaire : Évaluation de la condition multimodale (page 1).



La tâche proposée vous a-t-elle paru familière ?

☐ Oui
☐ Non

Sinon, son caractère inhabituel vous a-t-il gêné ?

☐ Oui
☐ Non

STRUCTURES DES AFFICHAGES

Vous avez pu observer que les photos étaient rangées différement d'un affichage à l'autre, suivant les structures suivantes :

- Arbitraire (placement aléatoire des photos)
- Elliptique (deux ellipses concentriques)
- Matricielle (tableau de cinq lignes et cinq colonnes)
- Radiale (huit rayons)

Le repérage de cibles dans une structure arbitraire vous a semblé :

Difficile _____ Facile

Lent _____ Rapide

Imprécis _____ Précis

Le repérage de cibles dans une structure elliptique vous a semblé :

Difficile _____ Facile

Lent _____ Rapide

Imprécis _____ Précis

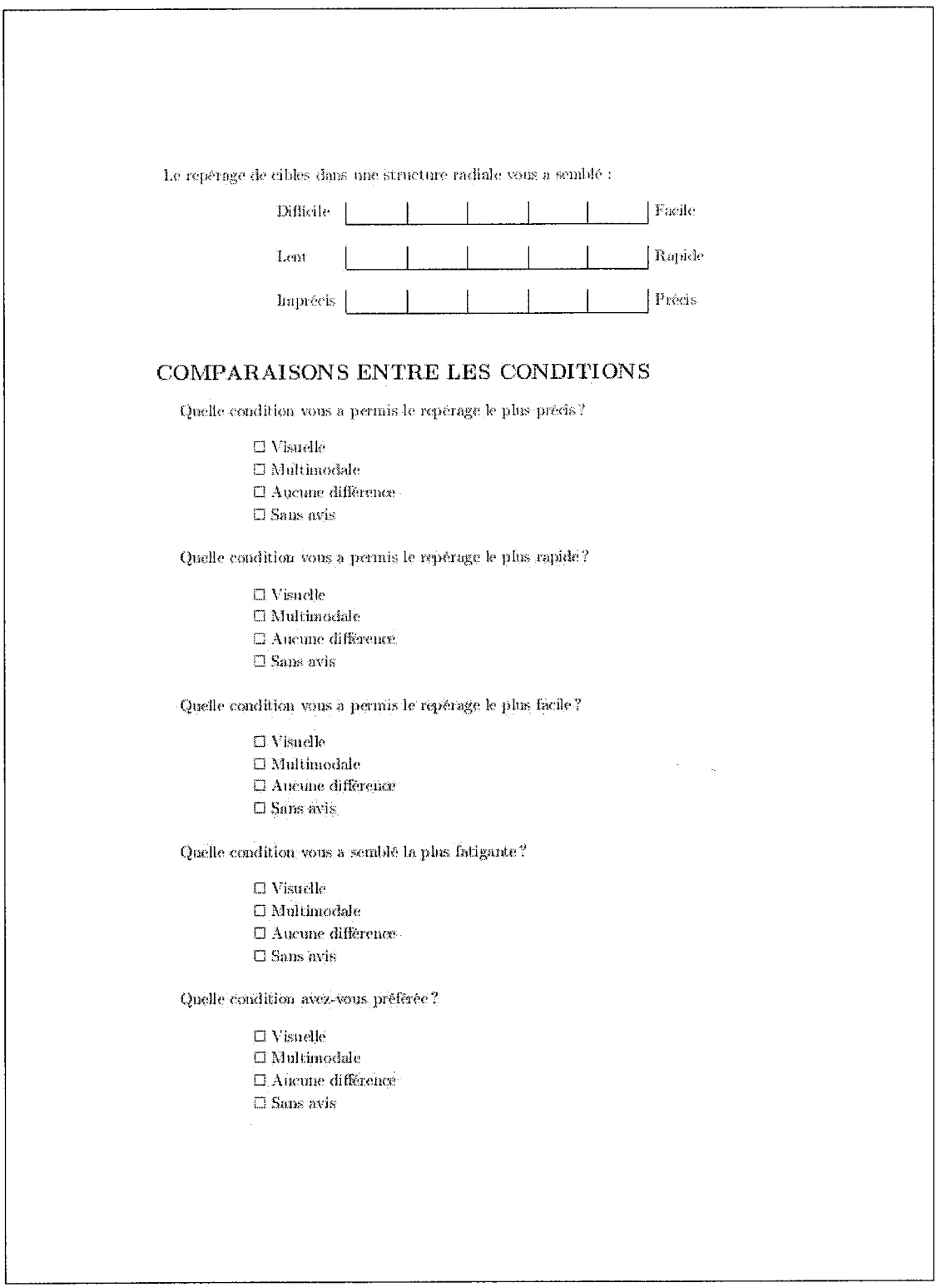
Le repérage de cibles dans une structure matricielle vous a semblé :

Difficile _____ Facile

Lent _____ Rapide

Imprécis _____ Précis

FIG. B.5 – Troisième questionnaire : Évaluation de la condition multimodale (page 2).



Le repérage de cibles dans une structure radiale vous a semblé :

Difficile ____|____|____|____|____| Facile

Lent ____|____|____|____|____| Rapide

Imprécis ____|____|____|____|____| Précis

COMPARAISONS ENTRE LES CONDITIONS

Quelle condition vous a permis le repérage le plus précis ?

- ☐ Visuelle
- ☐ Multimodale
- ☐ Aucune différence
- ☐ Sans avis

Quelle condition vous a permis le repérage le plus rapide ?

- ☐ Visuelle
- ☐ Multimodale
- ☐ Aucune différence
- ☐ Sans avis

Quelle condition vous a permis le repérage le plus facile ?

- ☐ Visuelle
- ☐ Multimodale
- ☐ Aucune différence
- ☐ Sans avis

Quelle condition vous a semblé la plus fatigante ?

- ☐ Visuelle
- ☐ Multimodale
- ☐ Aucune différence
- ☐ Sans avis

Quelle condition avez-vous préférée ?

- ☐ Visuelle
- ☐ Multimodale
- ☐ Aucune différence
- ☐ Sans avis

FIG. B.6 – Toisième questionnaire : Évaluation de la condition multimodale (page 3).

Variable : temps de sélection des cibles (ms)				
Condition	Groupe	Images	Moyenne (ms)	Nombre d'observations
PV	1	1 à 30	7630	360
PV	1	31 à 60	6184	360
PV	1	61 à 90	7008	360
PV	1	91 à 120	6917	360
PV	2	1 à 30	4028	360
PV	2	31 à 60	4752	360
PV	2	61 à 90	4614	360
PV	2	91 à 120	4256	360
PM	1	1 à 30	1935	360
PM	1	31 à 60	2031	360
PM	1	61 à 90	1772	360
PM	1	91 à 120	1678	360
PM	2	1 à 30	1748	360
PM	2	31 à 60	1628	360
PM	2	61 à 90	1618	360
PM	2	91 à 120	1569	360

FIG. B.7 – Évolution des temps moyens de sélection.

Variable : nombre d'erreurs				
Condition	Groupe	Images	Nombre d'erreurs	Nombre d'observations
PV	1	1 à 30	20	360
PV	1	31 à 60	19	360
PV	1	61 à 90	20	360
PV	1	91 à 120	17	360
PV	2	1 à 30	25	360
PV	2	31 à 60	20	360
PV	2	61 à 90	19	360
PV	2	91 à 120	10	360
PM	1	1 à 30	4	360
PM	1	31 à 60	5	360
PM	1	61 à 90	2	360
PM	1	91 à 120	6	360
PM	2	1 à 30	17	360
PM	2	31 à 60	21	360
PM	2	61 à 90	10	360
PM	2	91 à 120	13	360

FIG. B.8 – Évolution de la précision des sélections.

Table des figures

Liste des tableaux

Bibliographie

[Ahlberg *et al.*, 1992] C. Ahlberg, C. Williamson et B. Shneiderman. Dynamic queries for information exploration: an implementation and evaluation. Dans *Proceedings of ACM Conf. on Human Factors in Computing Systems CHI'92 (May 3-7 1992, Monterey, USA)*, rédacteurs P. Bauersfeld, J. Bennett et G. Lynch, pages 619–626. New York:ACM Press, 1992.

[Althoff *et al.*, 2001] F. Althoff, G. McGlaun, G. Spahn et M. Lang. Combining multiple input modalities for virtual reality navigation - a user study. Dans *Proceedings of 9th International Conf. on Human-Computer Interaction HCI International 2001(August 5-10 2001, New Orleans, Louisiana, USA)*, rédacteurs M. Smith, G. Salvendy, D. Harris et R.J. Koubek, pages 47–49. Mahwah, NJ:Lawrence Erlbaum Associates, 2001.

[André et Rist, 1993] E. André et T. Rist. The design of illustrated documents as a planning task. Dans *Intelligent Multimedia Interfaces*, rédacteur M.T. Maybury, pages 94–116. Menlo Park (CA):AAAI/MIT Press, 1993.

[André, 1997] E. André. WIP and PPP : A comparison of two multimedia presentation systems in terms of the standard reference model. *Computer Standards and Interfaces*, 18(6-7):555–564, 1997.

[Aumont, 2001] J. Aumont. *L'image (Deuxième édition)*. NATHAN, 2001.

[Baber, 2001] C. Baber. Computing in a multimodal world. Dans *Proceedings of 1st International Conf. on Universal Access in Human-Computer Interaction UAHCI'01 HCI International 2001(August 5-10 2001, New Orleans)*, rédacteur C. Stephanidis, volume 3. Mahwah (NJ):Lawrence Erlbaum Associates, 2001.

[Baudel et Braffort, 1993] T. Baudel et A. Braffort. Reconnaissance de gestes de la main en environnement réel. Dans *Actes du Colloque international 'L'interface des mondes réels et virtuels' (Mars 1993, Montpellier, France)*, pages 207–216, 1993.

[Bernsen, 1993] N.-O. Bernsen. Modality theory: supporting multimodal interface design. Dans *Proceedings from the ERCIM Workshop on Multimodal Human-Comuter Interaction (Nancy, November 1993)*, pages 13–23, 1993.

[Bernsen, 1994] N.-O. Bernsen. Foundations of multimodal representations, a taxonomy of representational modalities. *Interacting with computers*, 6:347–371, 1994.

[Bertin, 1981] J. Bertin. *Graphics and graphics information-processing.* Berlin:Walter de Gruyter, 1981.

[Bertin, 1983] J. Bertin. *Semiology of graphics: diagrams, networks, maps.* Madison, Wisconsin:The University of Winsconsin Press, 1983.

[Blois *et al.*, 1999] V. Blois, S. Vermandel, J. Charlier, F. Leclerc, C. Altuzarra, B. Guery, D. Beaune, H. Djeddi, S. Defoort-Dhelemmes, M. Goudemand et P. Thomas. Trois recherches à propos d'un appareil de commande par le regard comme outil de communication en réanimation pédiatrique. Résultats préliminaires. *Les cahiers du REIRPR*, 11:32–46, 1999.

[Bolt, 1980] R.A. Bolt. Put-That-There: voice and gesture at the graphics interface. *Computer Graphics*, 14(3):262–270, 1980.

[Bouchard et Cyr, 2000] S. Bouchard et C. Cyr. *Recherche psychosociale : Pour harmoniser recherche et pratique.* Sainte-Foy (Québec, Canada):Presses de l'Université du Québec, 2000.

[Bourget, 1992] M.L. Bourget. *Conception et réalisation d'une interface de dialogue personne-machine multimodale.* PhD thesis, Institut National Polytechnique, Grenoble, 1992.

[Burhans *et al.*, 1995] D.T. Burhans, R. Chopra et R.K. Srihari. Domain specific understanding of spatial expressions. Dans *Proceedings of the forteenth International Joint Conference on Artificial Intelligence IJCAI-95 (19th August 1995, Montréal, Canada)*, rédacteur Cris Melish, pages 33–40. Morgan Kaufman, 1995.

[Byrne *et al.*, 1999] M.D. Byrne, B.E. John, N.S. Wehrle et D.C. Crow. The tangled web we wove: a taskonomy of www use. Dans *Proceedings of ACM Conf. on Human Factors in Computing Systems CHI'99 (May 15-20 1999, Pittsburgh, PA, USA)*, rédacteurs M.G. Williams, M.W. Altom, K. Ehrlich et W. Newman, pages 544–551. New York:ACM Press, 1999.

[Cadet *et al.*, 2002] C. Cadet, L. Charles et J.-L. Galus. *La communication par l'image.* Paris:NATHAN, 2002.

[Carbonell et Kieffer, 2002] N. Carbonell et S. Kieffer. Do oral messages help visual exploration? Dans *Proceedings of International CLASS Workshop on Natural, Intelligent and Effective Interaction in Multimodal Dialogue Systems (June 28-29 2002, Copenhagen, DK)*, pages 27–36, 2002.

[Carbonell et Kieffer, To appear] N. Carbonell et S. Kieffer. Do oral messages help visuel search? Dans *Natural, intelligent and effective interaction in multimodal dialogue systems*, rédacteurs N.O. Bernsen, J.V. Kuppevelt et L. Dybkjaer, page 25 pages. Kluwer Academic Publishers, To appear.

[Card et Mackinlay, 1997] S.K. Card et J. Mackinlay. The structure of the information visualization design space. Dans *Proceedings of IEEE Symposium on Information Visualization InfoVis'97 (October 19-24 1997, Phoenix, AZ)*, rédacteurs J. Dill et N. Gershon, pages 92–99. IEEE Computer Society Press, 1997.

[Catinis et Caelen, 1995] L. Catinis et J. Caelen. Analyse du comportement multimodal de l'usager humain dans une tâche de dessin. Dans *Septièmes journées sur l'ingénierie de l'Interaction Homme-Machine IHM'95 (Octobre 1995, Toulouse)*, pages 123–129. Toulouse:Cépaduès-Éditions, 1995.

[Chelazzi, 1999] L. Chelazzi. Serial attention mechanisms in visual search: A critical look at the evidence. *Psychological Research*, 62(2-3):195–219, 1999.

[Christensen, 1997] L.B. Christensen. *Experimental Methodology (7th ed.)*. Boston:Allyn and Bacon, 1997.

[Chuah et Roth, 1996] M. Chuah et S. Roth. On the semantics of interactive visualizations. Dans *Proceedings of IEEE Symposium on Information Visualization (InfoVis'96)(October 28-29 1996, San Francisco, CA, USA)*, pages 29–36. IEEE Computer Society Press, 1996.

[Coutaz *et al.*, 1995] J. Coutaz, L. Nigay, D. Salber, A. Blandford, J. May et R. Young. Four easy pieces for assessing the usability of multimodal interaction: the CARE properties. Dans *Proceedings of 5th IFIP International Conference on Human-Computer Interaction INTERACT'95 (June 25-29 1995,Lillehammer, Norway)*, rédacteurs K. Nordby, P. Helmersen, D. Gilmore et S. Arnesen, pages 115–120. Boston, MA:Kluwer Academic Publishers, 1995.

[Coutaz et Caelen, 1991] J. Coutaz et J. Caelen. A taxonomy for multimedia and multimodal user interfaces. Dans *First ERCIM Workshop on Multimodal Human-Computer Interaction (November 1991,Lisbon)*, pages 143–148, 1991.

[Cribbin et Chen, 2001a] T. Cribbin et C. Chen. Exploring cognitive issues in visual information retrieval. Dans *Proceedings of Eighth IFIP TC 13 Conf. on Human Computer Interaction, INTERACT 2001 (July 9-13 2001, Tokyo, Japan)*, rédacteur M. Hirose, pages 166–173. NCP, 2001.

[Cribbin et Chen, 2001b] T. Cribbin et C. Chen. A study of navigation strategies in spatial-semantic visualizations. Dans *Proceedings of 9th International Conf. on Human-Computer Interaction HCI International 2001 (August 5-10 2001, New Orleans, Louisiana, USA)*, rédacteurs M. Smith, G. Salvendy, D. Harris et R.J. Koubek, volume 1, pages 948–952. Mahwah, NJ:Lawrence Erlbaum Associates, 2001.

[De Vries et Johnson, 1997] G. De Vries et G.I. Johnson. Spoken help for a car stereo: an exploratory study. *Behaviour and Information Technology*, 16(2):79–87, 1997.

[Diederich *et al.*, 2003] A. Diederich, H. Colonius, D. Bockhorst et S. Tabeling. Visual-tactile spatial interaction in saccade generation. *Experimental Brain Research*, 148(3):328–337, February 2003.

[Doll et Home, 2001] T.J. Doll et R. Home. Guidelines for developing and validating models of visual search and target acquisition. *Optical Engineering*, 40(9):1776–1783, 2001.

[Dollinger et Hoyer, 1996] S. Dollinger et W. Hoyer. Age and skill differences in the processing demands of visual inspection. *Applied Cognitive Psychology*, 10:225–239, 1996.

[Drury et Clement, 1978] C.G. Drury et M.R. Clement. The effect of area, density, and number of background characters on visual search. *Human Factors*, 20:597–602, 1978.

[Drury, 1992] C.G. Drury. Inspection performance. Dans *Handbook of Industrial Engineering, 2nd Edition*, rédacteur Gavriel Salvendy, chapitre 88, pages 2283–2314. New York:John Wiley and Sons Inc., 1992.

[Duncan et Humphreys, 1989] J. Duncan et G. Humphreys. Visual search and stimulus similarity. *Psychological Review*, 96:433–458, 1989.

[Ehrenmann *et al.*, 2001] M. Ehrenmann, R. Zollner, S. Knoop et R. Dillmann. Sensor fusion approaches for observation of user actions in programming by demonstration. Dans *Proceedings of the IEEE International Conference on Multi Sensor Fusion and Intergration (MFI) (Baden-Baden, Germany, 19-22 August 2001)*, volume 1, pages 227–232, 2001.

[Engelkamp, 1992] J. Engelkamp. Modality and modularity of the mind. Dans *Actes du 5ème colloque de l'ARC 'Percevoir, Raisonner, Agir - Articulation de Modèles Cognitifs' (March 24-26 1992, Nancy, France)*, pages 321–343, 1992.

[Faraday et Sutcliffe, 1997] P. Faraday et A. Sutcliffe. Designing effective multimedia presentations. Dans *Proceedings of ACM Conf. on Human Factors in Computing Systems CHI'97 (March 22-27 1997, Atlanta, Georgia, USA)*, rédacteur S. Pemberton, pages 272–278. New York:ACM Press, 1997.

[Findlay et Gilchrist, 1998] J. Findlay et I. Gilchrist. Eye guidance and visual search. Dans Underwood [1998], chapitre 13, pages 295–312.

[Frank, 1998] A. Frank. Formal models for cognition - taxonomy of spatial location description and frames of reference. Dans *Spatial Cognition - An Interdisciplinary Approach to Representing and Processing Spatial Knowledge*, rédacteurs C. Freksa, C. Habel et K.F. Wender, volume 1, pages 293–312. Berlin:Springer Verlag, 1998.

[Furnas et Bederson, 1995] G.W. Furnas et B.B. Bederson. Space-scale diagrams: understanding multiscale interfaces. Dans *Proceedings of ACM Conf. on Human Factors in Computing Systems CHI'95 (May 7-11 1995, Denver, Colorado, USA)*, rédacteurs I. Katz, R. Mack, L. Marks, M.B. Rosson et J. Nielsen, pages 234–241. New York:ACM Press, 1995.

[Giraudet, 2000] G. Giraudet. *Mise en évidence de la flexibilité du système visuel.* PhD thesis, École des Hautes Études en Sciences Sociales (EHESS), spécialité Sciences Cognitives, 2000.

[Gramopadhye et Madhani, 2001] A.K. Gramopadhye et K. Madhani. Visual search and visual lobe size. Dans *Proceedings of Fourth International Conference on Visual Form - Lecture Notes in Computer Science*, rédacteurs C. Arcelli et et al, volume 2059, pages 525–531. Berlin:Springer Verlag, 2001.

[Guyomard *et al.*, 1995] M. Guyomard, D. Le Meur, S. Poignonnec et J. Siroux. Experimental work for the dual usage of voice and touch screen for a cartographic application. Dans *Proceedings of the ESCA Tutorial and Research Workshop on Spoken Dialogue Systems (30 May-2 June 1995, Vigso, Denmark)*, pages 153–156, 1995.

[Harisson et Vicente, 1996] B.L. Harisson et K.J. Vicente. An experimental evaluation of transparent menu usage. Dans *Proceedings of ACM Conf. on Human Factors in Computing Systems CHI'96 (April 13-18 1996, Vancouver, British Columbia, Canada)*, rédacteurs M. Tauber, V. Belloti, R. Jeffries, J.D. Mackinlay et J. Nielsen, pages 391–398. New York:ACM Press, 1996.

[Hauptmann et McAvinney, 1993] A.G. Hauptmann et P. McAvinney. Gestures with speech for graphic manipulation. *International Journal of Man-Machine Studies*, 38:231–249, 1993.

[Henderson et Hollingworth, 1998] J.M. Henderson et A. Hollingworth. Eye movements during scene viewing: an overview. Dans Underwood [1998], chapitre 12, pages 269–293.

[Huges *et al.*, 1996] H.C. Huges, G. Nozawa et F. Kitterle. Global precedence, spatial frequancy channels, and the statistics of natural images. *Journal of Cognitive Neuroscience*, 8:197–230, 1996.

[Itti *et al.*, 1998] L. Itti, C. Koch et E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259, 1998.

[Itti et Koch, 1999] L. Itti et C. Koch. Learning to detect salient objects in natural scenes using visual attention, 1999.

[Itti et Koch, 2000] L. Itti et C. Koch. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40(10-12):1489–1506, 2000.

[Itti, 2000] L. Itti. http://ilab.usc.edu/bu/. Site Internet: Bottom-Up Visual Attention Home Page, 2000.

[Jacob, 1993] R.J.K. Jacob. Eye-gaze computer interfaces: what you look is what you get. *IEEE Computer*, 26(7):65–67, July 1993.

[Kalyuga *et al.*, 1999] S. Kalyuga, P. Chandler et J. Sweller. Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology*, 13:351–371, 1999.

[Kieffer et Carbonell, 2003] S. Kieffer et N. Carbonell. Assistance orale à la recherche visuelle dans des affichages complexes. Dans *Actes de la 15ème Conférence Francophone sur l'Interaction Homme-Machine (25-28 Novembre 2003, Caen, France)*, pages 126–133, 2003.

[Kramer *et al.*, 2001] A.F. Kramer, N.D. Cassavaugh, D.E. Irwin, M.S. Peterson et S. Hahn. Influence of single and multiple onset distractors on visual search for singleton targets. *Perception & Psychophysics*, 63(6):952–968, September 2001.

[Krause, 1997] J. Krause. Multimodal interaction for mobile environments. *RIS - Review of Information Science, 2(2)*, 1997.

[Lamping *et al.*, 1995] J. Lamping, R. Rao et P. Pirolli. A focus+context technique based on hyperbolic geometry for visualizing large hierarchies. Dans *Proceedings of ACM Conf. on Human Factors in Computing Systems CHI'95 (May 7-11 1995, Denver, Colorado, USA)*, rédacteurs

I. Katz, R. Mack, L. Marks, M.B. Rosson et J. Nielsen, pages 401–408. New York:ACM Press, 1995.

[Levy *et al.*, 1996] E Levy, J. Zacks, B. Tversky et D. Schiano. Gratuitous graphics ? putting preferences in perspective. Dans *Proceedings of ACM Conf. on Human Factors in Computing Systems CHI'96 (April 13-18 1996, Vancouver, British Columbia, Canada)*, rédacteurs M. Tauber, V. Belloti, R. Jeffries, J.D. Mackinlay et J. Nielsen, pages 42–49. New York:ACM Press, 1996.

[Livny *et al.*, 1997] M. Livny, R. Ramakrishnan, K. Beyer, G. Chen, D. Donjerkovic, S. Lawande, J. Myllymaki et K. Wenger. Devise: integrated querying and visual exploration of large datasets. *SIGMOD Record*, 26(2):301–312, 1997.

[Lokuge et Ishizaki, 1995] I. Lokuge et S. Ishizaki. Geospace: an interactive visualization system for exploring complex information spaces. Dans *Proceedings of ACM Conf. on Human Factors in Computing Systems CHI'95 (May 7-11 1995, Denver, Colorado, USA)*, rédacteurs I. Katz, R. Mack, L. Marks, M.B. Rosson et J. Nielsen, pages 409–413. New York:ACM Press, 1995.

[Léger *et al.*, 2003] L. Léger, D. Chene, T. Baccino et C. Tijus. The effect of semantic relatedness and typicality upon visual detection of a target. Dans *Proceedings of the 25th Annual Meeting of Conference of the Cognitive Science Society*, pages 716–721. Boston, NJ:LEA, 2003.

[Mackay *et al.*, 1998] W. Mackay, A.-L. Fayard, L. Frobert et L. Médini. Reinventing the familiar: exploring an augmented reality design space for air traffic control. Dans *Proceedings of ACM Conf. on Human Factors in Computing Systems CHI'98 (April 18-23 1998, Los Angeles, USA)*, rédacteurs C-M. Karat, A. Lund, J. Coutaz et J. Karat, pages 558–565. New York:ACM Press, 1998.

[Maybury, 1993] M.T. Maybury. *Intelligent Multimedia Interfaces.* Memlo Park, (CA):AAAI/MIT Press., 1993.

[Maybury, 2001] M.T. Maybury. Universal multimedia information access. Dans *Proceedings of 1st International Conf. on Universal Access in Human-Computer Interaction UAHCI'01 HCI International 2001(August 5-10 2001, New Orleans)*, rédacteur C. Stephanidis, volume 3, pages 382–386. Mahwah (NJ):Lawrence Erlbaum Associates, 2001.

[Micromégas, 2003] Micromégas. Approches multi-échelles pour la navigation dans les masses de données familières. Rapport technique, Action Concertée Incitative - Masses de données, 2003.

[Mignot et Carbonell, 1996] C. Mignot et N. Carbonell. Commande orale et gestuelle: étude empirique. *Technique et science informatiques*, 15(10):1399–1428, 1996.

[Mukherjea *et al.*, 1995] S. Mukherjea, J. Foley et S. Hudson. Visualization complex hypermedia networks through multiple hierarchical views. Dans *Proceedings of ACM Conf. on Human Factors in Computing Systems CHI'95 (May 7-11 1995, Denver, Colorado, USA)*, rédacteurs I. Katz, R. Mack, L. Marks, M.B. Rosson et J. Nielsen, pages 331–337. New York:ACM Press, 1995.

[Mulken *et al.*, 1999] S. Mulken, E. André et J. Müller. An empirical study of the trustworthiness of life-like interface agents. Dans *Proceedings of the 8th International Conf. on Human-Computer Interaction HCI International 1999 (August 22-27 1999, Munich, Germany)*, rédacteur J. Bullinger, H.-J. & Ziegler, pages 152–156. Mahwah (NJ):Lawrence Erlbaum Associates, 1999.

[Münz, 2001] S. Münz. http://fr.selfhtml.org/introduction/hypertexte/definitions.htm. Site Internet : SELFHTML - Aides à la navigation, 2001.

[Nigay et Coutaz, 1993] L. Nigay et J. Coutaz. A design space for multimodal systems: concurrent processing and data fusion. Dans *Proceedings of Conf. on Human Factors in Computing Systems INTERCHI'93 (April 24-29, 1993, Amsterdam, The Netherlands)*, rédacteurs S. Ashlund, K. Mullet, A. Henderson, E. Hollnagel et T. White, pages 172–178. New York:ACM Press & Addison Wesley, 1993.

[Oviatt *et al.*, 1997] S. Oviatt, A. DeAngeli et K. Kuhn. Integration and synchronization of input modes during multimodal human-computer interaction. Dans *Proceedings of ACM Conf. on Human Factors in Computing Systems CHI'97 (March 22-27 1997, Atlanta, Georgia, USA)*, rédacteur S. Pemberton, pages 415–422. New York:ACM Press, 1997.

[Paivio, 1977] A. Paivio. Images, propositions and knowledge. Dans *Images, perception and knowledge [Western Ontario Studies in the Philosophy of Science, 8]*, rédacteur J.M. Nicholas, pages 47–71. Dordrecht, Netherlands:Reidel, 1977.

[Pan et McKeown, 1996] S. Pan et K. R. McKeown. Spoken language generation in a multimedia system. Dans *Proceedings of ICSLP '96*, volume 1, pages 374–377, Philadelphia, PA, 1996.

[Pelz *et al.*, 2001] J. Pelz, M. Hayhoe et R. Loeber. The coordination of eye, head, and hand movements in a natural task. *Experimental Brain Research*, 139(3):266–277, 2001.

[Perkins, 1995] R. Perkins. The interchange online network: simplifying information access. Dans *Proceedings of ACM Conf. on Human Factors in Computing Systems CHI'95 (May 7-11 1995, Denver, Colorado, USA)*, rédacteurs I. Katz, R. Mack, L. Marks, M.B. Rosson et J. Nielsen, pages 558–565. New York:ACM Press, 1995.

[Pirolli *et al.*, 2000] P. Pirolli, S.K. Card et M.M. Van Der Wege. The effect of information scent on searching information visualizations of large tree structures. Dans *Proceedings of the International Working Conference on Advanced Visual Interfaces AVI'2000 (May 23-26 2000, Palermo, Italy)*, pages 161–172. New York:ACM Press, 2000.

[Pirolli et Card, 1995] P. Pirolli et S.K. Card. Information foraging in information access environments. Dans *Proceedings of ACM Conf. on Human Factors in Computing Systems CHI'95 (May 7-11 1995, Denver, Colorado, USA)*, rédacteurs I. Katz, R. Mack, L. Marks, M.B. Rosson et J. Nielsen, pages 51–58. New York:ACM Press, 1995.

[Plaisant *et al.*, 2002] C. Plaisant, J. Grosjean et B.B. Bederson. Space tree: supporting exploration in large node link tree, design evolution and empirical evaluation. Dans *Proceedings*

of INFOVIS 2002, IEEE Symposium on Information Visualization (October 2002, Boston, USA), pages 57–64, 2002.

[Plesniak et Ravikanth, 1998] W. Plesniak et P. Ravikanth. Coincident display using haptics and holographic video. Dans *Proceedings of ACM Conf. on Human Factors in Computing Systems CHI'98 (April 18-23 1998, Los Angeles, USA)*, rédacteurs C-M. Karat, A. Lund, J. Coutaz et J. Karat, pages 304–311. New York:ACM Press, 1998.

[Price et Humphreys, 1989] C.J Price et G.W. Humphreys. The effects of surface detail on object categorization and naming. *Quarterly Journal of Experimental Psychology*, 41A:797–828, 1989.

[Rasmussen, 1986] J. Rasmussen. Information processing and human-machine interaction: an approach to cognitive engineering. Dans *North-Holland series in system science and engineering*, rédacteur North-Holland, chapitre 12. Amsterdam, Netherlands:Elsevier, 1986.

[Robbe *et al.*, 2000] S. Robbe, N. Carbonell et P. Dauchy. Expression constraints in multimodal human-computer interaction. Dans *Proceedings of International Conference on Intelligent User Interfaces IUI 2000 (January 9-12, 2000, New Orleans, Louisiana, USA)*, rédacteur H. Lieberman, pages 225–229. New York:ACM Press, 2000.

[Robert, 1988] M. Robert. Validité, variables et contrôle. Dans *Fondements de la recherche scientifique en psychologie*, rédacteur M. Robert. St-Hyacinthe, Québec:Edisem, 1988.

[Robertson *et al.*, 1998] G. Robertson, M. Czerwinski, K. Larson, D.C. Robbins, D. Thiel et M. van Dantzich. Data mountain: Using spatial memory for document management. Dans *Proceedings of the 11th Annual Symposium on User Interface Software and Technology UIST 1998 (November 1-4, 1998, San Francisco, California, USA)*, pages 153–162. New York:ACM Press, 1998.

[Rosenthal et Jacobson, 1968] R. Rosenthal et L. Jacobson. *Pygmalion in the classroom.* New York:Holt, Rinehart and Winston, 1968.

[Rosenthal, 1976] R. Rosenthal. *Experimenter effect in behavioral research.* New York:Halsted Press, 1976.

[Shipman *et al.*, 1995] F. Shipman, C. Marshall et T. Moran. Finding and using implicit structure in human-organised spatial layouts of information. Dans *Proceedings of ACM Conf. on Human Factors in Computing Systems CHI'95 (May 7-11 1995, Denver, Colorado, USA)*, rédacteurs I. Katz, R. Mack, L. Marks, M.B. Rosson et J. Nielsen, pages 346–353. New York:ACM Press, 1995.

[Shneiderman, 1983] B. Shneiderman. Direct manipulation: a step beyond programming languages. *IEEE Computer*, 16:57–69, 1983.

[Shneiderman, 1996] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. Dans *Proceedings of IEEE Visual Languages 1996 (September 3-6, Boulder, Colorado, USA)*, pages 336–343. IEEE Computer Society Press, 1996.

[Siroux *et al.*, 1997] J. Siroux, M. Guyomard, F. Multon et C. Remondeau. Multimodal references in georal tactile. Dans *Proceedings of EACL'97 Workshop (July 11th 1997, Madrid, Spain)*, pages 39–43, 1997.

[Sutcliffe et Patel, 1996] A. Sutcliffe et U. Patel. 3D or not 3D: is this nobler in the mind? Dans *Proceedings of HCI'96 People and Computers XI (August 20-23 1996, London, UK)*, rédacteurs M. A. Sasse, R. J. Cunningham et R. L. Winder, pages 79–94. London, UK:Springer-Verlag, 1996.

[Tanriverdi et Jacob, 2000] V. Tanriverdi et R.J.K. Jacob. Interacting with eye movements in virtual environments. Dans *Proceedings of ACM Conf. on Human Factors in Computing Systems CHI'00 (April 1-6 2000, The Hague, Amsterdam)*, rédacteurs T. Turner, G. Szwillus, M. Czerwinski et F. Paterno, pages 265–272. New York:ACM Press, 2000.

[Treisman et Gormican, 1988] A. Treisman et S. Gormican. Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, 95:15–48, 1988.

[Underwood, 1998] rédacteur G. Underwood. *Eye guidance in reading and scene perception.* North Holland Elsevier Science Ltd-Oxford, 1998.

[Van Diepen *et al.*, 1998] M.J. Van Diepen, M. Wampers et G. d'Ydewall. Functional division of the visual field: moving masks and moving windows. Dans Underwood [1998], chapitre 15, pages 337–355.

[Van Diepen *et al.*, 1999] M.J. Van Diepen, L. Ruelens et G. d'Ydewall. Brief foveal masking during scene perception. *Acta Psychologica*, 101:91–103, 1999.

[Vernier et Nigay, 2000] F. Vernier et L. Nigay. Interfaces multimodales : composition et caractérisation des modalités de sortie. Dans *Actes de la conférence ERGO-IHM 2000 (3-6 Octobre 2000, Biarritz, France)*, rédacteurs D.L. Scapin et E. Vergison, pages 203–210. CRT ILS & ESTIA, Bidart, 2000.

[Wang *et al.*, 2000] Y. Wang, Z. Liu et J.-C. Huang. Multimedia content analysis-using both audio and visual clues. *Signal Processing Magazine, IEEE*, 17(6):12–36, November 2000.

[Ware, 2004] rédacteur C. Ware. *Information Visualization. Perception for design ($2^{ème}$ édition).* North Holland Elsevier Science Ltd-Oxford, 2004.

[Yu et Brewster, 2003] W. Yu et S. Brewster. Evaluation of multimodal graphs for blind people. *Universal Access in the Information Society*, 2(2):105–124, 2003.

Mademoiselle **KIEFFER Suzanne**

DOCTORAT de l'UNIVERSITE HENRI POINCARE, NANCY 1

en INFORMATIQUE

VU, APPROUVÉ ET PERMIS D'IMPRIMER N° 1072

Nancy, le 19 juillet 2005

Le Président de l'Université

LE PRÉSIDENT

J.P. FINANCE

Université Henri Poincaré, Nancy I
24-30 rue Lionnois - B.P. 3069 - 54013 NANCY Cédex
Tél. : 03 83 68 20 00 - Fax : 03 83 68 21 00

Résumé

Ce travail porte sur la conception d'une nouvelle forme d'interaction Homme-Machine: la multimodalité parole+présentation visuelle en sortie du système. Plus précisément, l'étude porte sur l'évaluation des apports potentiels de la parole, en tant que mode d'expression complémentaire du graphique, lors du repérage visuel de cibles au sein d'affichages 2D interactifs.

Nous avons adopté une approche expérimentale pour déterminer l'influence d'indications spatiales orales sur la rapidité et la précision du repérage de cibles, et évaluer la satisfaction subjective d'utilisateurs potentiels dans cette forme d'assistance à l'activité d'exploration visuelle.

Les différentes études réalisées ont montré d'une part que les présentations multimodales facilitent et améliorent les performances des utilisateurs pour le repérage visuel, en termes de temps et de précision de sélection des cibles. Elles ont montré d'autre part que les stratégies d'exploration visuelle des affichages, en l'absence de messages sonores, dépendent de l'organisation spatiale des informations au sein de l'affichage graphique.

Abstract

This work is about the design of a new form of Human-Computer Interaction: the speech+visual presentation combination as an output multimodality. More precisely, it deals with the valuation of the potential benefits from the speech, as a graphical expression mode, during visual target detection tasks in 2D interactive layouts.

We used an experimental approach to show the influence of oral messages, including the spatial localization of the target in the layout, on users speed and accuracy. We also valued the subjective satisfaction of the users about this assistance to visual exploration.

Three experimental studies showed that multimodal presentations of the target facilitate and improve users performances, considering both speed and accuracy. They also showed that visual exploration strategies, without any oral message, depend on the spatial organization of the informations displayed.