

RAPPORT DE STAGE DE MASTER 2 ATIAM

Rôle de l'audition dans la perception de l'espace

Indiana WOLLMAN

Equipe *Espaces Acoustiques et Cognitifs* – IRCAM – Paris

1 mars 2010 - 7 juillet 2010

Sous la direction de Isabelle Viaud Delmon (Isabelle.Viaud-Delmon(at)ircam.fr)
Olivier Warusfel (Olivier.Warusfel(at)ircam.fr)

Remerciements

Je tiens à remercier mes encadrants Isabelle Viaud-Delmon et Olivier Warusfel pour m'avoir accueillie dans leur équipe de recherche et m'avoir proposé un sujet de Master qui m'a passionnée.

Je suis particulièrement reconnaissante à Thibaut Carpentier, Van Nguyen et Joseph Sanson pour leurs disponibilités, les nombreuses discussions que nous avons eues et l'aide qu'ils m'ont apporté.

Un grand merci à Emmanuel Flety et Gérard Bertrand pour la réalisation des irremplaçables charentaises augmentées.

Toute ma reconnaissance à l'Ircam comme Institution dont j'ai apprécié le caractère unique au monde. Et un clin d'œil à Carlos Agon qui a été mon premier contact à l'Ircam.

Table des matières

1 - Introduction - Vers la création d'un monde virtuel auditif évolutif.....	3
2 - Cadre général : La perception multi-sensorielle de l'espace.....	5
2.1 - Etudes des processus mis en jeu lors de l'apprentissage d'un espace.....	5
2.2 - Intégrer l'audition dans un protocole d'étude de la mémoire spatiale.....	6
2.2.1 - Le corps sonore dans l'espace sonore.....	7
2.2.2 - Avatar sonore lié aux pas du sujet : sonification d'indice idiothétique.....	8
3 - Techniques de réalité virtuelle auditive.....	9
3.1 - Dispositif de captation gestuelle.....	9
3.2 - Contrôle et traitement de signal.....	9
3.3 - Techniques de reproduction sonore spatiale immersive et interactive.....	12
3.3.1 Synthèse binaurale.....	12
3.3.2 Mesures de HRTF/Individualisation : Etat de l'art.....	13
3.3.3 Pistes considérées pour la synthèse de sources en basse élévation.....	15
4 - Déroulement pratique de l'expérience.....	19
4.1 - Participants.....	19
4.2 - Protocole expérimental.....	19
5 - Résultats.....	24
5.1 - Analyse de la phase d'apprentissage et du « Probe ».....	24
5.2 - Analyse de la phase de test.....	30
6 - Conclusion générale.....	38
7 - Bibliographie.....	39

Introduction

Vers la création d'un monde virtuel auditif évolutif

L'intérêt grandissant pour les environnements dits immersifs et interactifs, y compris dans le domaine artistique incite à reposer les questions de perception spatiale auditive dans un cadre plus large tenant compte de l'interaction entre nos différentes voies sensorielles, notamment visuelles, auditives et proprioceptives. L'étude de l'interaction audition-proprioception apparaît essentielle pour le développement raisonné de nouvelles médiations sonores basées sur l'interaction corps/audition/espace.

Une série d'expériences s'appuyant sur un paradigme classique utilisé en neurosciences pour étudier la mémoire spatiale, le « Morris Water Maze » (Morris 1981), a déjà été effectuée dans l'équipe Espaces Acoustiques et Cognitifs au sein de laquelle je réalise mon stage. Les travaux déjà réalisés au sein de l'équipe ont principalement été menés dans le cadre de projets nationaux ou européens et ont bénéficié de l'apport de chercheurs invités issus du domaine des neurosciences.

Le projet que j'ai mené pendant ces quatre mois s'inscrit dans la continuité de ces travaux. L'objectif principal de mon stage est d'étudier l'interaction audition/corps/espace dans un dispositif de reproduction sonore immersif - environnement interactif excluant la modalité visuelle - et de développer de nouveaux paradigmes d'interaction avec un contenu sonore. Il s'agit donc de mettre en place une expérience de psychologie expérimentale permettant de tester la mémoire auditive cognitive spatiale.

Ce projet exploite ainsi les technologies de restitution 3D et de capture du mouvement existantes, et permet, à travers la mise en place de scènes de réalité virtuelles auditives, d'étudier les paramètres auditifs pertinents pour l'établissement d'une carte spatiale cognitive dans une tâche de localisation.

Nous décrirons tout d'abord le cadre général dans lequel s'inscrit l'expérience de psychologie cognitive basée sur la création d'un monde virtuel auditif évolutif et incluant la mise en place d'un avatar sonore.

Nous présenterons ensuite les différents outils technologiques et les investigations d'acoustique virtuelle auxquels nous avons fait appel pour mettre en place cette expérience.

Nous détaillerons dans un troisième temps le protocole expérimental tel que nous l'avons fait passer à un groupe de volontaires sains.

Enfin, nous décrirons les résultats de l'expérience et présenterons une première analyse de ceux-ci.

2 - Cadre général : La perception multi-sensorielle de l'espace

2.1 Etudes des processus mis en jeu lors de l'apprentissage d'un espace

La tâche de navigation dans un labyrinthe d'eau, le "Water Maze", est une expérience très couramment utilisée dans le domaine des neurosciences pour étudier l'apprentissage et la mémoire spatiale. Cette expérience a été mise au point par le neuro-scientifique Richard G. Morris en 1981 [1] pour montrer que des lésions de l'hippocampe - structure du cerveau jouant un rôle important dans la mémoire, la navigation et l'orientation dans l'espace - entraînent un affaiblissement de l'apprentissage spatial [2].

Dans le paradigme classique, un animal - généralement un rat - est placé dans une piscine contenant une "plateforme de repos", cachée à quelques millimètres en dessous de la surface de l'eau, tel que représenté sur la *figure 1* ci-dessous. La couleur de l'eau étant rendue opaque par l'ajout de poudre de lait, l'animal doit nager au départ sans pistes pour retrouver cette plateforme cachée. Les seuls repères qu'il peut utiliser sont des repères visuels, se trouvant hors de la piscine. Une fois la plateforme trouvée, l'essai est répété plusieurs fois - la plateforme demeurant toujours à la même place - et à mesure que l'animal se familiarise avec le labyrinthe, le temps passé pour retrouver la cible diminue. Cette amélioration des performances provient du fait que l'animal a appris à localiser la plateforme par rapport aux indices visuels qu'il aura jugés pertinents. Après suffisamment d'essais, l'animal est capable de nager directement sur la plateforme cible, quel que soit le point d'entrée dans le labyrinthe.

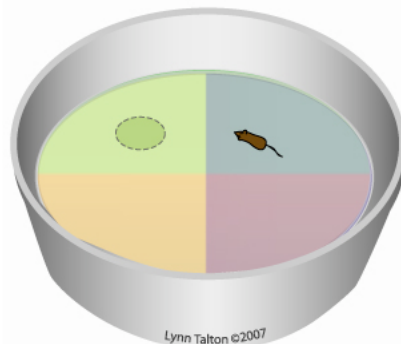


figure 1 : Morris Water Maze

- La plateforme cachée se situe en haut à gauche
- La partition de la zone de navigation en quarts est schématisée à l'aide des 4 couleurs

Au cours des différentes navigations, plusieurs paramètres comme le temps total passé pour atteindre la plateforme, la distance totale parcourue, ou encore la trajectoire décrite sont enregistrés.

Cette expérience a été revisitée de nombreuses fois depuis, et l'adaptation du test à des sujets humains a été facilitée par la mise en place d'environnements de réalité virtuelle, impliquant exclusivement la modalité visuelle. La navigation dans le Water Maze virtuel s'effectue donc à l'aide d'un joystick. Il s'agit toujours pour les participants de se repérer dans un espace au départ inconnu, dans le but de retrouver une plateforme cible, non visible à l'entrée dans le labyrinthe.

La différence des performances entre genres - distinguant les hommes et les femmes - sur des sujets humains a été un des sujets de préoccupation premiers dans l'étude des tâches

de localisation dans un Water Maze virtuel. Il a en effet été prouvé que les capacités d'apprentissage spatial sont différentes chez les mâles et femelles mammifères et cette constatation méritait d'être étudiée au sein de la population humaine. Dans cette voie, plusieurs études telles que celles menées par Sandstrom et al. [3] et Astur et al. [4] ont montrées qu'il existe des différences certaines entre les deux sexes, non seulement sur les caractéristiques de navigation mais aussi sur les performances enregistrées. La première étude citée montre que les sujets humains féminins utilisent principalement des indices visuels précis alors que les sujets masculins utilisent non seulement ces indices visuels mais également les informations fournies par la géométrie de l'espace de navigation. La seconde étude montre qu'il existe une différence affirmée entre les sexes dans l'apprentissage spatial virtuel : les hommes retrouvent bien plus rapidement la cible que les femmes. Ces résultats ont été les premiers à démontrer l'efficacité et l'utilité du Morris Water Maze pour étudier la mémoire spatiale chez l'homme.

Un article plus poussé de ces mêmes auteurs [5] a prouvé que l'expérience du Morris Water Maze n'est pas interchangeable avec n'importe quel autre test de mémoire spatiale. En effet, la comparaison de l'expérience du Water Maze avec une autre utilisant un autre type de dispositif connu de tests d'évaluation de la mémoire spatiale, démontrent que les deux paradigmes mettent en jeu différentes stratégies d'orientation pendant la tâche effectuée. Les performances comparatives des deux sexes ne sont pas les mêmes selon qu'il s'agissent de l'une ou l'autre des expériences.

Hamilton et al. [6] ont également mis en place un Water Maze virtuel testable par des sujets humains et les résultats obtenus dans cette tâche de localisation étant très similaires à ceux trouvés avec les rats, ils sont arrivés à la conclusion que les processus impliqués dans l'apprentissage cognitif spatial sont régis par le même type de mécanismes chez l'homme et chez le rongeur.

Par ailleurs, l'analyse des stratégies utilisées par des sujets pour se repérer dans une tâche de localisation fait partie des facteurs à considérer pour l'investigation de la mémoire cognitive spatiale. Une étude plus récente a mis en évidence les différentes stratégies d'orientation utilisées dans cette tâche de localisation. Ainsi, Kallai [7] et al. ont montré qu'il existe trois stratégies de recherche dans une tâche de navigation de type labyrinthe virtuel, prédisant les performances spatiales des sujets.

Les différentes expériences chez l'homme conduites depuis la mise en place de l'expérience fondatrice du Water Maze de Richard G. Morris ont donc centré l'étude de la mémoire spatiale exclusivement autour de l'utilisation de la *modalité visuelle* dans une tâche de localisation. De plus, les environnements virtuels utilisés pour mesurer les performances spatiales des participants ne sont pas des environnements virtuels immersifs.

2.2 Intégrer l'audition dans un protocole d'étude de la mémoire spatiale

Une série d'expériences basées sur l'expérience du Morris Water Maze et permettant de s'intéresser aux aspects multisensoriels de la mémoire spatiale humaine a déjà été menée au sein de l'équipe Espaces Sonores et Cognitifs de l'Ircam et c'est dans la continuité de ces travaux que mon stage s'inscrit. Le travail de l'équipe est, lui, centré sur l'utilisation de la *modalité auditive* dans un environnement excluant la modalité visuelle. J'ai dû mettre en place une expérience de psychologie expérimentale mettant en évidence la contribution de la *modalité auditive* aux représentations spatiales dans un environnement virtuel *immersif*, exclusivement sonore.

Le dispositif expérimental auquel j'ai eu recours permet aux volontaires de déambuler librement dans le noir dans un espace de 30 m² tout en étant immergés auditivement dans une scène virtuelle mise à jour en temps réel selon leurs déplacements : les sujets sont équipés d'un casque audio sans fil délivrant le paysage sonore, et portent sur leur tête des marqueurs de position utiles au suivi des déplacements. Un système de tracking sans fil - vidéo infra-rouge avec des marqueurs passifs - est utilisé pour transmettre au système de rendu sonore les coordonnées de la tête du sujet. Le système de rendu sonore utilise une base de HRTFs synthétisant les indices directionnels et simulant les caractéristiques acoustiques de la pièce expérimentale pour reproduire un environnement acoustique réaliste.

Sur la base du Water Maze d'origine, le but de l'expérience est d'analyser comment un participant est capable de retrouver 'la source cible' - équivalent de la plateforme de repos - dans l'espace sonore dans lequel il est invité à évoluer. Cette source est intuitivement « cachée » car elle ne se déclenche que lorsque le sujet passe dessus. Elle est donc 'cachée' à l'audition jusqu'au moment où l'auditeur la retrouve.

L'environnement auditif du participant est composé de stimuli auditifs constants positionnées tout autour de la zone définie pour la marche. Le participant est donc placé dans une situation de navigation par l'audition et par l'action : aucun périphérique de contrôle ou autre joystick n'est nécessaire pour réaliser cette expérience. L'interaction qui existe entre l'audition et les mouvements du participant - c'est-à-dire entre deux modalités sensorielles, auditive et idiothétique (sensations proprioceptives et vestibulaires) - réside dans l'asservissement du contenu et de l'organisation spatiale de la scène sonore aux mouvements du sujet. Les avantages attendus d'une telle interaction sont qu'elle repose sur des modalités sensorielles tridimensionnelles, donc utiles à la mémoire cognitive spatiale. Par ailleurs, nos représentations spatiales sont intimement liées aux processus d'apprentissage engageant le corps. C'est donc grâce à la maîtrise conjointe de ces deux modalités, audition et action, que nous pourrions synthétiser et contrôler des espaces sonores interactifs.

Nous avons par ailleurs cherché à développer des avatars sonores pouvant potentiellement améliorer les tâches de localisation dans des environnements virtuels auditifs. Dans de telles tâches, on peut en effet imaginer que la perception de la distance entre une source fixe située dans l'espace proximal et une source dynamique attachée à une partie du corps, représente un indice améliorant la précision du mouvement. Le système de suivi de mouvements par caméras infrarouges utilisé permet ainsi d'attacher un objet sonore à un segment corporel du sujet. Il est intéressant d'analyser si le retour auditif des mouvements du sujet dans l'espace constitue une augmentation sensorielle permettant d'affiner le comportement sensorimoteur pour l'exploration du monde virtuel.

2.2.1 Le corps sonore dans l'espace sonore

Le travail d'expérimentation réalisé lors de ce projet est centré sur la notion d'avatar sonore, c'est-à-dire sur la manifestation auditive de la présence de l'utilisateur dans le monde virtuel exclusivement sonore dans lequel il est amené à évoluer. Le concept d'avatar sonore permet une interactivité bi-directionnelle entre le monde virtuel et le sujet. Ce dernier ne se contente pas de se mouvoir en fonction de l'espace sonore dans lequel il est immergé, mais ses mouvements peuvent également modifier l'espace sonore environnant.

Introduire un avatar sonore dans la scène auditive revient donc à spatialiser un signal sonore lié et produit par le déplacement d'un des segments corporels du sujet, par exemple un bras, les pieds, la tête... la conception de l'avatar devant résulter d'un compromis entre la facilité avec laquelle on peut le repérer et ainsi l'interpréter, et la minimisation de sa surcharge auditive (masquage). Le fait d'utiliser des avatars sonores nécessite donc de

maîtriser la nature particulière du son, notamment ses rapports au temps et à l'espace, pour mettre en place les paramètres caractéristiques de sa manifestation sonore.

Au final, l'utilisation d'un avatar sonore permet d'étudier les liens qui existent entre les actions motrices et l'anticipation auditive.

2.2.2 Avatar sonore lié aux pas du sujet : sonification d'indices idiothétiques

La nouveauté de l'expérience réalisée pendant mon stage a été d'intégrer un nouvel avatar sonore dans l'Auditory Water Maze : la diffusion des bruits de pas du sujet au casque lui permet d'avoir un retour sonore de son évolution dans l'espace.

La sonification des pas est en effet naturelle dans une situation de navigation dans un environnement sonore virtuel car il s'agit d'une partie du corps produisant effectivement du son dans les situations de marche réelles. Les pieds d'une personne représentent ainsi simplement deux nouvelles sources sonores placées à environ 1m60 en dessous du sujet pour se conformer à la réalité.

Les différentes situations de tests que nous avons mises en place pendant l'expérience nous ont permis de tester si le retour sonore des pas du sujet aidait à la « sensation de présence ». Plusieurs situations pouvaient être envisagées pour aborder cette question.

Notre préoccupation principale a été de créer des distorsions entre la sensation réelle qu'a le sujet en se déplaçant et le retour sonore modifié qu'il entend de ses bruits de pas. Un exemple de distorsion serait de modifier simplement la *réverbération des pas* dans un test, ce qui pourrait suggérer que l'environnement total a été modifié. Cette modification n'affecte pourtant pas la structure géométrique des indices auditifs - la scène sonore serait composée des mêmes sources, positionnées au même endroit de l'espace - mais seulement l'espace dans lequel ces indices s'inscrivent. Nous avons donc cherché à mettre en oeuvre une tâche expérimentale nous permettant de mesurer l'habileté des êtres humains à se repérer dans un monde virtuel sonore où des distorsions perceptives peuvent déranger la perception unifiée du corps. Il s'agit de savoir si le sujet est capable à s'adapter au conflit entre les informations proprioceptives fournies par la marche et les informations auditives spatiales ou s'il s'en retrouve véritablement désorienté. Dans ce dernier cas, les capacités du sujet à retrouver la source cible dans l'environnement virtuel diminueraient significativement comparé à une situation normale où l'on n'introduit pas de distorsion.

Cette sonification a soulevé une série de questions qui a demandé de tester en tout premier lieu le résultat perceptif de la situation pour que nous nous rendions compte de la crédibilité du rendu et de son intérêt pour des questions expérimentales. L'intégration de l'avatar dans l'espace s'est fait donc en parallèle d'expériences perceptives, et de vérifier que l'on ne saturait pas la perception auditive d'informations.

Finalement, c'est grâce à l'inclusion cet avatar sonore que nous avons pu étudier l'interaction corps/audition/espace dans le dispositif de reproduction sonore immersif. Il est clair que l'exploration de ces différentes modalités d'interaction nécessite la création de modules de traitement des signaux issus de la capture spatiale. Les données récupérées étant transmises au synthétiseur sonore pour établir le retour auditif en temps-réel.

3. Techniques de réalité virtuelle auditive

Au moins deux dispositifs sont obligatoires pour mettre en place une expérience de réalité virtuelle : un dispositif de captation gestuelle qui permet de rendre compte en temps réel de la position et des mouvements du sujet et des techniques de reproduction sonore spatialisée 3D. Une interface de contrôle permet d'implémenter les étapes de traitement faisant le lien entre ces deux technologies.

3.1 Dispositif de captation gestuelle

Différents dispositifs de caméras permettant de suivre la position d'objet ou de personne dans l'espace sont aujourd'hui disponibles et couramment utilisés pour la réalité virtuelle.

Huit caméras infrarouge constituent le système de capture video que l'on a utilisé pendant l'expérience et convertissent les mouvements de l'utilisateur pour les traiter en temps réel grâce au logiciel Max/MSP. Ce système de cameras infra-rouges est couplé à des marqueurs passifs de type (billes rétro réfléchissantes) portés par les sujets. Deux marqueurs distincts ont été utilisés pendant l'expérience pour nous permettre de suivre des mouvements de corps différents en fournissant les positions du sujet sous forme de 6 coordonnées, les trois directions de l'espace, et les trois angles d'inclinaison par rapport au trois directions de l'espace. Un des marqueurs est placé sur la tête du sujet pour nous permettre de suivre la trajectoire globale du sujet dans l'aire de tracking ; les rotations de la tête n'ont pas d'influence sur cette trajectoire mais sont importantes pour asservir en temps réel les indices de localisation exploités par le système de rendu sonore. L'autre est porté à la ceinture, dans le dos du sujet et nous permet de suivre la position du bassin, donc des jambes et des pieds - il est très rare que lors de la marche, les pieds tournent sans le bassin - Ce marqueur est donc utile pour le rendu sonore des sons de pas. Cet accrochage au niveau de la ceinture permet donc de s'affranchir des mouvements (rotation en particulier) de tête qui sont enregistrés avec le premier marqueur cité.

Ces marqueurs sont assez performants en termes de précision spatiale et zone de couverture, mais sont tributaires de la taille du sujet qui participe à l'expérience : plus le sujet est grand plus nous rencontrons des problèmes de tracking lors des expériences, car les caméras perdent périodiquement la trace des marqueurs placés sur le sommet de la tête du sujet et dans son dos. Le sujet trop grand pouvait passer dans un « trou » de couverture vidéo au niveau de sa tête, et également masquer les marqueurs de son dos lors de la marche avant en sortie de zone.

3.2 Contrôle et Traitement de signal

La gestion globale de l'expérience comprend des tâches de construction de l'environnement sonore, de définition des interactions et de contrôle de leurs évolutions au cours du déroulement de l'expérience. Les paramètres de spatialisation des sources sonores composant la scène virtuelle doivent être rafraîchis en temps réel avec une latence inférieure au seuil de perception. Ceci nous a amené à traiter l'interaction le plus directement possible entre les modules de capture spatiale et ceux de synthèse sonore. Nous avons donc réalisé une interface sous Max/MSP permettant d'effectuer à la fois la partie de traitement du signal et la partie de contrôle de l'expérience menée.

Cette interface repose sur la librairie logicielle de spatialisation temps réel Spat, développé pour être l'outil de traitement de signal de Max/MSP. C'est ce logiciel SPAT qui génère la spatialisation des sons envoyés au système de rendu sonore utilisé.

Cette interface permet également de gérer les informations de tracking issues des caméras. Les positions des sujets sous la forme de 6 coordonnées (comme présentées au paragraphe précédent) dans l'espace de tracking prédéfini sont ainsi transmises au SPAT. Elles sont traitées pour transmettre au système de rendu sonore les sons à diffuser en fonction des coordonnées de la tête du sujet. Parallèlement, le temps total mis par les sujets pour retrouver la cible et leurs trajectoires sont systématiquement enregistrés dans des fichiers texte, grâce au contrôle des données de captation gestuelle par Max/MSP : toutes les 10ms, les positions des deux marqueurs portés par le sujet sont relevées sous forme de 6 coordonnées.

Cette interface possède pour finir des outils d'édition de scène sonore. Il était en effet possible de suivre à l'écran le déplacement des sujets dans la scène sonore. Grâce au contrôle du dispositif de captation vidéo effectué au sein de l'interface, la représentation imagée du sujet sur l'écran correspond à ses mouvements en temps réel.

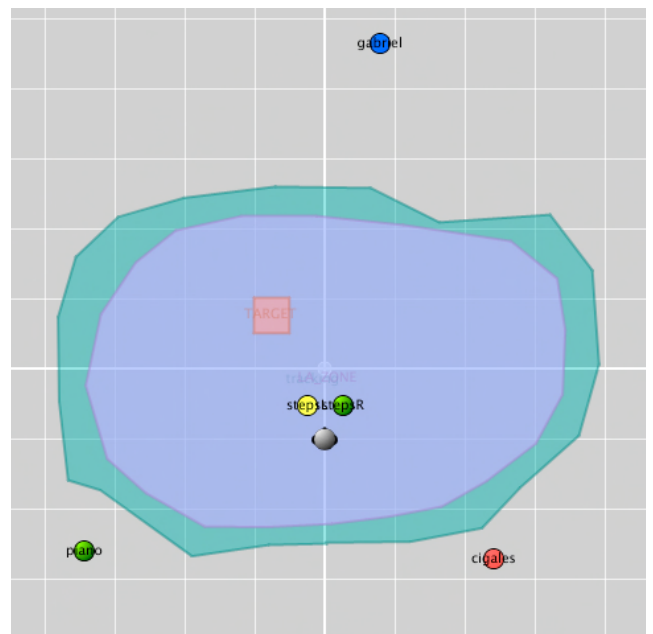
Spatialisation des sources sous Max/MSP

Dans notre expérience, cinq sons sont spatialisés pour le rendu sonore. En effet, selon qu'on tourne la tête d'un côté ou d'un autre, qu'on avance dans une direction ou une autre, le système de rendu sonore devait permettre au participant se repérer facilement et donc localiser précisément certaines sources sonores. Il s'agit de :

- Trois sources définissant le paysage sonore statique - sons respectivement de piano, de cigales, et de voix - sont positionnés en bordure de la zone de tracking, comme présenté ci-après. Les sources sont positionnées de telle sorte que leur distance au centre de la cible est identique et égale à 4 m. Les angles entre ces trois sources sont respectivement 160° (sources « piano »/ « voix »), 120° (sources « voix »/ « cigales »), 80° (sources « cigales »/ « piano »). Les positions des sources définissant le paysage sonore restent inchangées entre le début et la fin de l'expérience, on parlera alors de sources statiques.
- Les deux sources de sons de pas (pied gauche/pied droit) liées à l'avatar - sons de marche sur de la paille - sont positionnées au niveau de chacun des pieds gauche et droite du sujet, et leurs positions sont déduites de la position du marqueur porté à la ceinture en bas du dos du sujet. Elles sont placées à 1,2 m en dessous de la position en altitude du marqueur porté par le dos du sujet. Le pied gauche est placé à 20 cm sur la gauche de la position traquée du marqueur, et le pied droit est placé à 20 cm sur sa droite, permettant de respecter l'écart existant entre nos deux jambes. Les deux sources sont toutes deux placées à 50 cm devant le sujet.

figure 2 : *Visualisation de la scène de contrôle réalisée avec Max/MSP*

- La zone de tracking est en vert et la zone de recherche en bleu/violet
- La source cible est modélisée par le carré rouge.
- Les trois ronds bleu/rouge/vert situés en dehors de la zone sont les trois sources statiques.
- L'auditeur est en gris et les deux sources modélisant ses pieds sont les ronds jaune et vert



Trois autres sons sont utilisés dans cette expérience mais ne nécessitant pas de spatialisation pour le rendu sonore, ils ne sont pas traités par le Spat.

- Le bruit du vent en sortie de zone qui avertit de sujet qu'il a quitté l'aire de tracking
- La source cible (son de sifflet) a une position fixe durant toute l'expérience, et est placée aux coordonnées (-0.75 , 0.75) par rapport au centre de la zone.
- Les sons de galets utilisés dès que le sujet n'est pas en situation directe de test. Ils sont utilisés pour que le sujet ne puisse jamais entendre les bruits de la salle depuis l'instant où il y entre, jusqu'à ce qu'il en ressorte en fin d'expérience.

Finalement, la synthèse d'une scène sonore s'appuie sur des techniques de traitement du signal et ont abouti au développement de la librairie logicielle de spatialisation temps réel Spat et de modules dédiés à la synthèse binaurale sur casque adaptée aux applications immersives.

Dispositif de sonification des pas : conception des « chaussures augmentées »

Pour permettre au sujet d'entendre ses pas, nous avons fabriqué des paires de chaussures « sensibles à la pression » exercée par le pied lors de la marche. Au cours du déplacement, c'est tout naturellement que l'utilisateur lui-même déclenche les sons de pas synthétiques grâce aux capteurs de pression mécanique FSR (Force Sensitive Resistance) judicieusement placés sous ses semelles permettant de déclencher l'échantillon sonore. Les capteurs sont placés sous les talons des pieds gauche et droit, dans les semelles de chacune des deux chaussures. L'appui généré par le talon lorsque le sujet marche provoque la compression du capteur présent dans la semelle. La visualisation d'une augmentation de la pression sous un talon partant de 0 est ainsi synonyme de pas.

Les sons de pas synthétiques déclenchés par la marche sont contrôlés en temps réel dans la plateforme Max/MSP.



figure 3 : Chaussures Augmentées

Les quatre capteurs sont directement reliés à un boîtier émetteur wifi, la *Wise Box*, porté à la ceinture, comme montré en *figure 1*. La *Wise Box* est un outil développé par Emmanuel Flety (IRCAM) pour le contrôle gestuel. Ce dispositif est connecté par wifi à un

ordinateur portable et permet d'envoyer toutes les 5ms une valeur de tension provenant de chacun des capteurs à l'interface réalisée sous Max/MSP.



figure 4 : Câblage des chaussons - WiseBox

Un patch a été spécialement réalisé pour visualiser les variations de tension des capteurs FSR et nous a permis de fixer le seuil de tension significatif d'un pas ; un système de déclenchement à hystérésis a été implémenté. Les sons de pas sont donc spatialisés à l'aide du SPAT avant d'être envoyés au système de rendu sonore.

3.3 Techniques de reproduction sonore spatiale immersive et interactive

Il existe actuellement différentes technologies de reproduction pour restituer un champ sonore cohérent spatialement, c'est-à-dire un champ dans lequel l'évolution temporelle des indices acoustiques associés à la perception auditive spatiale, et utilisés pour la restitution sonore, est régie par les déplacements de l'auditeur. Les techniques holophoniques sont utilisées pour les applications collectives et la technique binaurale pour les applications individuelles. L'expérience menée pendant mon stage porte sur la perception individuelle et nécessite une grande mobilité de la part du sujet. Elle s'est appuyée sur la synthèse binaurale, technique de diffusion 3D pour casque.

3.3.1 Synthèse binaurale

La synthèse binaurale est un procédé de spatialisation des sons qui est basé sur le filtrage des signaux par des filtres reproduisant les propriétés de la propagation acoustique entre une source située en un point donné de l'espace et les conduits auditifs d'un auditeur. Ces filtres linéaires, chacun associé à une direction de l'espace, sont définis par les fonctions de transfert acoustiques (HRTF pour Head Related Transfer Function) qui correspondent à toutes les transformations subies par une onde sonore entre la source sonore considérée et les oreilles de l'auditeur : diffraction par la tête, réflexions sur le buste, résonances des pavillons. Ces filtres portent en eux l'expression de tous les indices physiques de localisation nécessaires pour une perception de l'espace par le système auditif. La synthèse binaurale utilise ainsi ces filtres pour sculpter les signaux à présenter aux tympons de l'auditeur afin de lui procurer l'illusion d'une scène sonore réaliste. Cette technique de reproduction sonore 3D nécessite l'utilisation d'un casque d'écoute.

3.3.2 Mesures de HRTF/Individualisation : Etat de l'art

Les HRTF étant très liées à la morphologie de la tête et des pavillons, la spatialisation n'est correctement assurée que si ces filtres sont bien adaptés à l'auditeur. Les techniques de mesures des HRTF en chambre anéchoïque se basent sur l'utilisation de deux microphones miniatures introduits dans les oreilles du sujet et d'un haut-parleur se déplaçant tout autour de la tête du sujet -placée au centre du dispositif - et décrivant un échantillonnage spatial fin. On procède à des mesures de réponses impulsionnelles relatives à la tête (HRIR) ; les HRTF sont les transformées de Fourier de ces HRIR.

Remarque

Les données brutes obtenues lors des mesures de HRTF sur les sujets sont le résultat d'une convolution : on mesure en fait la réponse impulsionnelle de l'oreille mais également celle du haut parleur diffusant et celle du micro enregistreur placé dans l'oreille. Pour supprimer les contributions des haut parleurs et des micros, les données ont ensuite été traitées par égalisation en champ diffus (il s'agit de réaliser une déconvolution en utilisant une HRIR mesurée en champ diffus). Le champ diffus est habituellement mesuré en chambre réverbérante mais une méthode alternative peut être employée : elle consiste à estimer la HRIR diffuse par une moyenne spatiale pondérée de tous les différents points de mesures. Toutes les mesures de HRTF ont donc été égalisées de la même façon.

Les HRTF contiennent les principaux indices physiques de localisation nécessaires pour une perception de l'espace par le système auditif : la différence interaurale en amplitude (ILD) et la différence interaurale en temps d'arrivée (ITD).

Les deux oreilles n'étant bien évidemment pas positionnées au même endroit, il existe une différence de trajet acoustique entre la source et chaque oreille, et donc un retard, résumé par le concept unique d'ITD, pour toute source située en dehors du plan médian. De plus, la tête d'un auditeur agit comme un obstacle face à une onde acoustique incidente. Il en résulte des différences d'intensité entre les sons captés à chaque oreille, qui dépendent de la position de la source. C'est ce que modélise l'ILD. La captation par les oreilles d'une même onde sonore donne lieu à ces deux indices dits 'interauraux' tel que représenté sur la figure ci-contre:

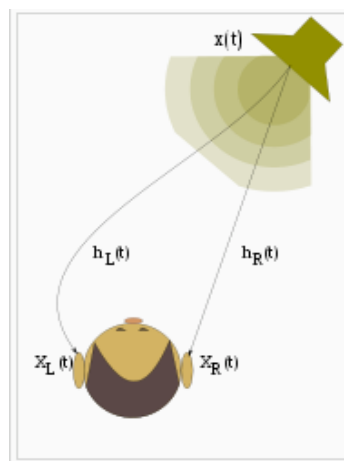


figure 5 : Approche système du problème acoustique.

Les phénomènes acoustiques modifient le signal entre sa génération à la source et sa captation aux oreilles et sont modélisés par des filtres linéaires invariants dans le temps h_L et h_R , qui dépendent de la position de la source dans l'espace

La source émet un signal $x(t)$ et les signaux reçus aux oreilles gauche et droite sont x_L et x_R . Dans le cas d'une source fixe, les phénomènes de propagation observés entre la source et les oreilles de l'auditeur peuvent être modélisés par deux systèmes linéaires invariants, caractérisables par leurs réponses impulsionnelles $h_L(t)$ et $h_R(t)$ telles que :

$$x_{L,R} = h_{L,R} * x(t) \text{ soit sous forme fréquentielle : } X_{L,R}(j\omega) = H_{L,R} \cdot X(j\omega)$$

où H_L et H_R sont les fonctions de transfert traduisant les phénomènes acoustiques subis par le signal x entre la source et l'oreille (HRTF).

Par ailleurs, les phénomènes de réflexions et de diffraction subis par l'onde sonore lors de son interaction avec le torse et la tête engendrent d'un point de vue signal, un filtrage : des creux et des pics dépendant de la direction de la source viennent entacher le spectre du signal d'origine. Ce sont ces colorations spectrales que l'on nomme indices spectraux. Ces indices sont dits monauraux - au sens où ils apparaissent aux tympons indépendamment pour chaque oreille - et sont notamment responsables de notre faculté de repérage de l'angle d'élévation. On parlera alors d'indices spectraux monauraux de localisation.

Les mesures de HRTF réalisées suivant la méthode de mesure de réponse impulsionnelle décrite précédemment rendent compte des deux indices acoustiques binauraux mentionnés et des indices spectraux. Cependant, les mesures exhaustives des HRTF sont longues, coûteuses et inconfortables, car elles doivent être réalisées pour de nombreuses directions de l'espace - séance de mesure durant 1 à 2 heures avec la tête bloquée pour un sujet - en chambre anéchoïque. Ces contraintes ont rendu difficile, voire impossible la mise en oeuvre de mesures individuelles de HRTF dans le cadre de l'expérience, qui devait débiter un mois et demi après mon arrivée dans l'équipe. Il a donc fallu investiguer les pistes disponibles pour s'affranchir des mesures de HRTF individuelles. Beaucoup d'études traitant des questions de *pseudo* individualisation des HRTF se retrouvent dans la littérature.

L'impact de l'utilisation de HRTF non-individuelles a fait l'objet de nombreuses études [8][9]. Il ressort qu'en termes de localisation, une dégradation des performances est observable pour une majorité de sujets. L'azimut des sources virtuelles reste généralement bien estimé, alors que l'élévation est nettement moins bien perçue. Wenzel et al. [10] montrent ainsi que les confusions haut/bas et avant/arrière augmentent considérablement, les sources positionnées à l'avant étant généralement perçues à l'arrière.

Il est cependant communément admis que l'utilisation des HRTF d'un bon localisateur est un bon choix pour réduire les artefacts perceptifs liés à la synthèse binaurale non-individuelle. Le choix de HRTF non-individuelles constituant un bon compromis universel doit se faire, comme l'ont proposé Moller et al. [11] d'après une évaluation préliminaire de toutes les HRTF d'une base de données par un large échantillon d'individus. Si la base de données contient les mesures de HRTF d'une grande partie de la population, présentant un large éventail de morphologies, alors il est probable qu'un nouvel auditeur y trouve un ensemble de HRTF qui lui convienne. L'équipe d'Espaces Acoustiques et Cognitifs a constitué une base de données de HRTF (80 personnes se sont prêtées aux mesures) et a mené des tests psychoacoustiques - évaluation sur tests d'écoute - pour sélectionner les jeux de HRTF les plus performants pour les sujets en termes de localisation. Deux sujets élus - sujets 'typiques' - dont les HRTF procurent au plus grand nombre une spatialisation satisfaisantes sont ressortis de cette étude. Les mesures ont été effectuées en balayant 11 points d'élévation variant sur l'intervalle $[-40^\circ; 90^\circ]$. La sphère balayée n'est pas totale à cause du dispositif expérimental : il n'est pas possible d'effectuer des mesures de HRTFs sur la partie de basse de la sphère car les hauts parleurs atteignent déjà le sol pour les mesures à -40° . 650 mesures de HRTF ont été enregistrées sur des sujets réels.

D'autre part, la situation de synthèse dynamique temps-réel dans laquelle les indices acoustiques synthétisés sont asservis aux mouvements du sujet réduisent l'importance d'un recours à des HRTFs individuelles.

Par ailleurs, des mesures sur tête artificielle ont été réalisées en chambre anéchoïque à l'Ircam, toujours dans le cadre de travaux menés par l'équipe Espaces Acoustiques et Cognitifs. Le modèle de tête artificielle utilisé est le modèle KEMAR, présenté ci-dessous. La tête artificielle ne souffrant pas d'être placée la tête vers le bas, nous disposons de 1722 points de mesures de HRTF, d'où la présence d'une banque de données balayant l'élévation $[-90^\circ, 40^\circ]$.



figure 6 : modèle de tête artificielle - mannequin KEMAR

Une autre façon de faire de la synthèse binaurale consiste à utiliser des modèles de tête, ce qui permet par là même de s'affranchir des mesures de HRTF. Les modèles classiques utilisés pour reproduire le champ sonore 3D avec la synthèse binaurale sont les modèles à tête sphérique rigide [12],[13]. Ces modèles se basent sur des HRTF contenant les deux indices acoustiques binauraux les plus importants pour la localisation en azimuth : l'ILD et l'ITD. Ces modèles ne sont néanmoins pas suffisants pour la localisation en élévation.

3.3.3 Pistes considérées pour la synthèse de sources en basse élévation

Les sons des pas utilisés pendant l'expérience étant diffusés au casque et représentant par définition une source sonore positionnée constamment à la verticale des sujets, nous avons cherché à réaliser une synthèse binaurale des sons de pas qui retranscrive correctement les sons de sources en basse élévation. Plusieurs méthodes ont été considérées pour réaliser cette synthèse : la prédiction des HRTF par une méthode de concaténation, l'utilisation de mesures de tête artificielle seule avec adaptation de l'ITD, et l'implémentation d'un modèle complet de tête/torse.

prédiction/concaténation des mesures de HRTFs tête cible/tête artificielle

Nous avons tout d'abord envisagé de réaliser une "interpolation" des mesures réalisées sur les têtes de sujets réels : il s'agit d'ajuster la partie spectrale des HRTF. En effet, nous disposons des mesures faites sur une tête artificielle (tête+tronc) sur toute la sphère entourant les oreilles de la tête et des mesures réalisées sur des têtes de sujets réels, balayant la portion de l'espace pour des angles d'élévation compris dans l'intervalle $[-40^\circ, +90^\circ]$. Toutes ces mesures étant déjà égalisées, l'idée était donc de concaténer les mesures obtenues sur les têtes de sujets réels avec les mesures obtenues avec la tête artificielle sur la partie basse élévation $[-90^\circ; -40^\circ]$. On choisirait ensuite par exemple comme tête réelle cible celle qui a été jugée la

plus adaptée par les auditeurs lors des tests d'écoute évoqués dans la partie précédente.

utiliser des mesures de tête artificielle seule avec adaptation de l'ITD par mise à l'échelle des maxima

Une autre solution a été d'utiliser les HRTFs venant d'un mannequin (KEMAR), le même pour tous les sujets de l'expérience, et d' "individualiser" ces HRTF en adaptant la partie temporelle des HRIR au sujet qui se prête à l'expérience. Cela revient à adapter l'ITD, qui représente un des indices les plus pertinents pour la reconnaissance spatiale, à chacun des sujets, en mesurant le rayon de sa tête avant l'expérience. L'ITD peut en effet être bien approché d'après une modélisation sphérique de la tête d'un auditeur tenant compte d'une part de la différence de marche entre les deux oreilles de l'auditeur (modèle de Woodworth [14]) et d'autre part de la différence de phase qui existe entre les signaux captés à chaque oreille et provenant d'une onde plane sinusoïdale [15] :

$$ITD = a/c (\sin(\theta) + \theta) \text{ pour les hautes fréquences}$$

$$ITD = 3a/c.\sin(\theta) \text{ pour les basses fréquences}$$

avec θ , l'azimut de la source et c la célérité du son dans l'air

On réalise alors cette adaptation d'ITD en réalisant une "mise à l'échelle des maxima d'ITD", c'est-à-dire que les valeurs des maxima l'ITD utilisés pour les HRTF de la synthèse binaurale sont égales à celles des maxima d'ITD calculés à partir du périmètre de la tête du sujet. Cette méthode d'adaptation d'ITD a finalement été écartée car nous avons trouvé que les maxima d'ITD pour l'ensemble de la base de données de HRTF ne variaient pas beaucoup (l'écart maximal trouvé étant de $3/44100 = 7.10^{-5}$ s, ce qui est très inférieur au "JND" - Just Noticeable Difference d'ITD - de l'ordre de 10 à 20 μ s).

Implémentation du modèle dit du 'bonhomme de neige'

Nous avons pour finir implémenté un modèle de synthèse binaurale fournissant des indices acoustiques de localisation en basse élévation. Nous avons pour cela cherché un modèle de tête/tronc, prenant en compte toutes les transformations subies par une onde sonore provenant d'une source positionnée en basse élévation.

En nous appuyant sur l'article de V. R. Algazi, R. O. Duda et D. M. Thompson, « The use of head-and-torso models for improved spatial sound synthesis » [16], nous avons implémenté le dit 'snowman model', modèle d'un 'bonhomme de neige' en ce qu'il modélise une tête et un torse par deux sphères rigides. L'effet de ces sphères sur l'onde sonore peut être approximé par une série de modules simples de traitement de signal : deux modules indépendant modélisant l'ITD, et deux modules représentant les filtrages. Les délais temporels modélisent l'arrivée du son sur la sphère de tête, et la réflexion des hautes fréquences sur le torse. Les filtres modélisent eux d'une part l'atténuation due à la tête des sons provenant de sources positionnées du côté controlatéral de la tête, et d'autre part l'atténuation due à la présence du torse pour des sources positionnées à basse élévation.

Bien qu'il soit connu depuis longtemps que le torse a une influence non-négligeable sur les HRTF, les études réalisées sur l'importance psycho-acoustique de ces effets sont restés sans conclusion. L'effet le plus connu est celui de la localisation d'une source positionnée au dessus du sujet, les épaules de ce dernier jouant le rôle de réflecteur. Dans une étude précédente, Agazi et al. [17] ont montré l'importance de la réflexion sur le torse comme indice d'élévation lorsque la source n'est pas positionnée dans le plan médian. En général, le délai temporel le plus long qui existe entre le son direct et son réfléchi est celui qui est mesuré lorsque la source est positionnée au dessus de la tête. A mesure que l'élévation de la source diminue, le délai temporel diminue, jusqu'à devenir éventuellement nul lorsque la droite

reliant la source à l'oreille est tangente au torse. L'ensemble de ces rayons tangents définit ce que l'on appelle le 'cône d'ombre du torse', comme présenté sur la figure suivante:

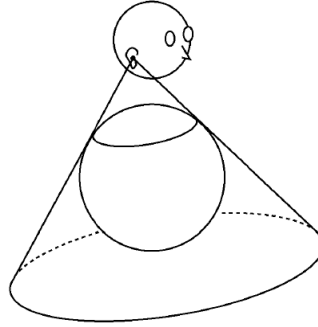


figure 7 : 'cône d'ombre du torse',
défini comme l'ensemble des droites partant de l'oreille, tangentes au torse

Une source positionnée en dehors du 'cône d'ombre du torse' entraîne une réflexion sur le torse modélisé par une sphère.

Le modèle du bonhomme de neige est défini par trois paramètres : le rayon de la tête (a), le rayon du torse (b) et la taille du cou (h). Les valeurs utilisées pour l'implémentation du modèle sont celles qui correspondent au modèle de tête sphérique KEMAR ($a = 8.7 \text{ cm}$; $b = 16.9 \text{ cm}$; $h = 5.3 \text{ cm}$).

Le modèle de filtrage par bonhomme de neige est développé en se basant sur le modèle de filtrage existant pour une simple sphère et en analysant la géométrie combinée des deux sphères qui composent le bonhomme de neige. Le modèle de tête sphérique à l'origine proposé par Brown et Duda [12] ,[13] consiste à mettre en cascade un filtre et un module de délai temporel. Il est représenté ci-après :

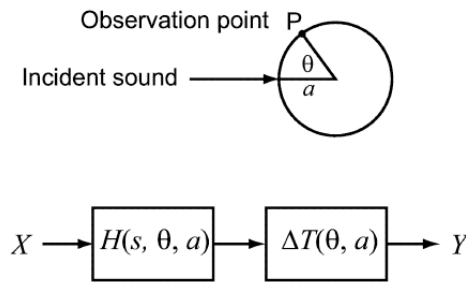


figure 8 : Diagramme en bloc d'un modèle de filtrage
pour les HRTF calculées sur une sphère isolée

En combinant les contributions des deux sphères individuelles, approximées par les filtres décrits précédemment, on obtient un modèle convenable pour reproduire un champ sonore 3D avec la synthèse binaurale, dont des sources sont en basse élévation.

La figure (a) suivante illustre l'alternance du modèle entre réflexion sur le torse quand la source est hors du cône d'ombre généré par le torse et non-réflexion lorsqu'elle est à l'intérieur de ce cône. Par des considérations géométriques, il est possible de déterminer facilement l'instant de transition. La figure (b) représente les deux provenances possibles d'un rayon parvenant aux oreilles : provenant soit d'un cheminement direct soit d'une première réflexion sur le torse. La figure (c) modélise le cheminement d'une onde émanant d'une

source présente à l'intérieur du cône d'ombre du torse. Celle-ci passe donc d'abord autour du torse avant d'atteindre la tête et donc l'oreille.

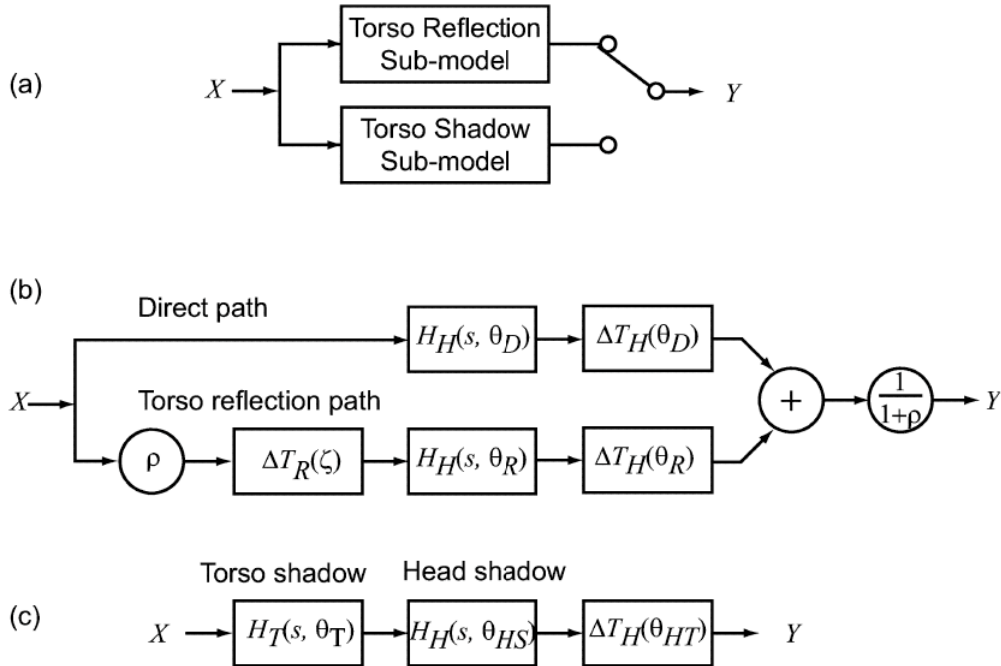


figure 9 : Diagramme en bloc du modèle de filtrage pour le “snowman”

Cette méthode de synthèse de HRTF se basant sur l'utilisation d'un modèle tête/torse a fourni des résultats satisfaisants, mais les tests sont restés très informels.

Cependant, n'ayant pas pu le tester en situation réelle pour valider sa pertinence, nous avons finalement choisi d'utiliser les mêmes HRTF pour tous les sujets, HRTF provenant des mesures réalisées avec la tête artificielle KEMAR.

Le couplage de la technique binaurale avec un dispositif de suivi de la position et de l'orientation de la tête nous a permis d'asservir la scène sonore diffuse aux mouvements de l'auditeur. La mise en jeu de cette boucle de rétroaction entre proprioception et audition est déterminante pour l'adhésion de l'auditeur à la scène sonore virtuelle qui lui est présentée puisqu'elle permettait de garantir la congruence entre les variations des indices acoustiques binauraux et ses mouvements propres.

4. Déroulement pratique de l'expérience

4.1 Participants

Vingt personnes (douze hommes et huit femmes), en majorité étudiants, ont passé l'expérience. Chacun d'eux a reçu une indemnisation de 10 euros pour 1h d'expérience. La moyenne d'âge des sujets est 25 ans.

Les participants ont été informés des consignes de l'expérience dans la phase de préparation précédant leur passage.

4.2 Protocole expérimental

Le test se décompose en plusieurs étapes.

Questionnaires à remplir

Il a été demandé aux sujets de remplir avant l'expérience deux questionnaires concernant leur aptitude à s'immerger dans des scènes de réalité virtuelle. Deux autres questionnaires sont également remplis en fin d'expérience, pour vérifier qu'il y a concordance entre les séries de réponses données en pré et post expérience.

Par ailleurs, durant l'expérience, nous avons posé à deux reprises la même série de questions aux sujets, dans un ordre différent pour éviter les automatismes de réponses, concernant leur « conscience du corps ». Il s'agit d'analyser si les modifications opérées sur le retour sonore de leur pas modifient également leur sensation de présence au cours de l'expérience.

La première série de questions suit la première phase de l'expérience, et la seconde le premier test.

Préparation des sujets

Avant de débiter l'expérience, nous équipons les sujets d'écouteurs sans fils et des chaussons augmentés cablés au boîtier qu'ils portent à la ceinture, sans leur révéler l'utilité du dispositif utilisé, (voir figure ci-contre). Nous leur expliquons simplement que les chaussons qu'ils portent permettent de coupler leurs pas à leur évolution dans l'environnement sonore. Sur les écouteurs et sur la ceinture sont fixés des ensembles de 5 marqueurs permettant de localiser leur position dans l'espace.

A l'entrée dans la pièce, les sujets sont invités à s'entraîner à déambuler dans le noir et à écouter le paysage auditif dans lequel ils sont immergés et grâce auquel il leur a été demandé de prendre des repères. Après cette étape de familiarisation, les sujets sont immergés dans une nouvelle scène spatiale auditive, qui a une configuration et une délimitation précises.

L'expérience commence alors, et se décompose en trois parties : la phase d'apprentissage, la phase de tests, le « probe ». Pendant tout le déroulement de l'expérience, la



figure 10 : Sujet équipé

pièce reste plongée dans le noir et les sujets entendent constamment un paysage sonore.

Première phase : ‘Apprentissage’

La première phase de l’expérience est consacrée à l’apprentissage, c’est-à-dire que les sujets sont immergés six fois de suite dans la même scène sonore, créant un paysage auditif complexe : les trois sources sonores statiques (le piano, la voix et les cigales) forment la scène sonore statique ; les sons de pas du sujet viennent s’ajouter à cette scène sonore comme présenté sur la figure suivante

figure 11 : Visualisation de la scène d’apprentissage (réalisée avec Max/MSP)

- La zone de tracking est en vert et la zone de recherche en bleu/violet
- La cible est modélisée par le carré rouge.
- Les trois ronds bleu/rouge/vert situés en dehors de la zone sont les trois sources statiques.
- L’auditeur est en gris et les deux sources modélisant ses pieds sont les ronds jaune et vert



Ces sources étant spatialisées, le paysage sonore doit leur servir à se repérer. A chaque fois que les participants sortent de la zone de recherche, un souffle de vent se déclenche dans leur casque, signifiant qu’ils ont quitté l’aire de tracking et qu’ils doivent y retourner. Au sein de cette scène spatiale auditive, leur tâche est de trouver une source sonore cachée : cette cible sonore est localisée dans une petite zone que les participants trouvent en explorant la scène spatiale auditive. Lorsqu’ils passent sur cette petite zone, un bruit de sifflet se déclenche dans leur casque. Dès qu’ils la quittent, le sifflet s’arrête. La tâche des participants est de se diriger le plus rapidement et le plus précisément possible sur cette cible. L’essai se finit automatiquement dès qu’ils ont trouvé la cible.

Les sujets réalisent six essais. A chaque nouvel essai, les sujets sont guidés dans le noir vers un nouveau point d’entrée dans la scène, pour que le point d’entrée n’ait pas d’incidence sur les résultats. Les sujets ont 3 minutes pour trouver la cible. S’ils ne l’ont pas trouvé au bout de ce temps-là, l’essai est automatiquement arrêté.

Après les 6 essais utiles pour l’apprentissage - il a été montré que l’on nécessite de se déplacer 6 fois dans la même scène avant de converger directement sur la cible, témoignage d’une bonne mémorisation spatiale du lieu - la première série de questions lui est posée. Le sujet peut alors passer à la phase de tests.

Seconde phase : ‘Tests’

La seconde phase de l’expérience est une phase de « tests ». Certains éléments de la scène spatiale auditive sont modifiés mais les sujets ont toujours suffisamment d’indices pour

pouvoir se repérer, de la même façon qu'ils le faisaient au cours de l'apprentissage. Le sujet a toujours pour but de retrouver la cible cachée qui, elle, ne change pas de place. Les sujets prennent, ou non, conscience des modifications sonores opérées.

L'idée est donc de modifier certains aspects de la configuration sonore pour tenter de mettre en évidence leur contribution respective dans la représentation spatiale que le sujet se fait. Chacun des 4 tests mis en oeuvre dans le protocole est immédiatement suivi d'une "ré-immersion" dans la scène sonore que le sujet a mémorisée dans la phase d'apprentissage pour effectuer une sorte de stabilisation de la mémoire et des repères utilisés.

Le premier test réalisé concerne la perception auditive du corps. La configuration de la scène sonore est identique à celle de l'apprentissage. Les bruits de pas sont toujours déclenchés par la marche du sujet mais les sons diffusés au casque sont délocalisés sur sa gauche et non là où il se trouve, de sorte à induire un conflit spatial entre la position du corps et le retour sonore du corps, comme présenté sur la figure suivante :

figure 12 : Visualisation de la scène du test 1
(réalisée avec Max/MSP)

- La zone de tracking est en vert et la zone de recherche en bleu/violet
- La cible est modélisée par le carré rouge.
- Les trois ronds bleu/rouge/vert situés en dehors de la zone sont les trois sources statiques.
- L'auditeur est en gris et les deux sources modélisant ses pieds sont les ronds jaune et vert, ici délocalisés sur sa gauche



Un retour sonore de pas délocalisés par rapport aux informations spatiales auditives provenant de la scène sonore calculées en fonction de la position de la tête pourrait potentiellement générer un phénomène de type autoscopique - sortie de corps - Ce test a nécessité de spatialiser le retour sonore des pas, pour les localiser à 1,5m sur la gauche, 60cm devant le sujet. Ici, dans la mesure où le sujet reconnaît le bruit des pas comme les siens, on introduit un conflit entre les informations idiothétiques produites par le corps en interaction avec l'espace, avec le retour sonore de l'action, sans modifier l'espace. A l'issue de ce second test, la seconde série de questions concernant la « conscience du corps » est posée au sujet.

Après un "ré-apprentissage" de la scène de départ, le sujet réalise le second test dans lequel les sons des sources sonores statiques de la scène sont modifiés : les trois sons réels précédemment utilisés sont remplacés par des sons purement synthétiques (se référer à la **figure** pour la représentation visuelle de la scène, avec des noms de sources statiques différents). Il s'agit de tester si le codage spatial est plus important que le codage sémantique dans une tâche de repérage spatial. En d'autres termes, on cherche à analyser l'implication de l'assignation de 'sens' aux sons (codage sémantique) dans le repérage spatial auditif. Dans ce test, on modifie simplement la composante sémantique des indices mais la structure géométrique des indices - à savoir l'agencement spatial de la scène - reste inchangé.

Dans le troisième test, la scène spatiale définie par les trois sources statiques est identique à celle de l'apprentissage, mais le son du retour sonore des pas est modifié dans une certaine zone de l'espace parcouru. En effet, la zone tracking est divisée en deux : sur une moitié de l'espace, le retour sonore des pas est tel que le sujet l'a découvert pendant l'apprentissage, c'est-à-dire qu'il entend les sons de marche sur la paille ; sur l'autre moitié, le sujet entend des sons de marche sur le gravier.

figure 13 : Visualisation de la scène du test 3
(réalisée avec Max/MSP)

- La zone de tracking est en vert et la zone de recherche en bleu/violet, la partie grisée est la zone des sons de gravier, l'autre celle des sons de paille
- La cible est modélisée par le carré rouge.
- Les trois ronds bleu/rouge/vert situés en dehors de la zone sont les trois sources statiques.
- L'auditeur est en gris et les deux sources modélisant ses pieds sont les ronds jaune et vert, ici délocalisés sur sa gauche



En modifiant le son du retour sonore des pas dans une zone de l'espace, on suggère que l'environnement total a été modifié.

Dans le quatrième test, on supprime le retour sonore des pas, ie on supprime les sources correspondants au pas. La configuration de la scène spatiale reste, elle, inchangée, comme présenté ci-après :

figure 14 : Visualisation de la scène du test 4
(réalisée avec Max/MSP)

- La zone de tracking est en vert et la zone de recherche en bleu/violet.
- La cible est modélisée par le carré rouge.
- Les trois ronds bleu/rouge/vert situés en dehors de la zone sont les trois sources statiques.
- L'auditeur est en gris.
(les sources modélisant les pieds sont absentes)



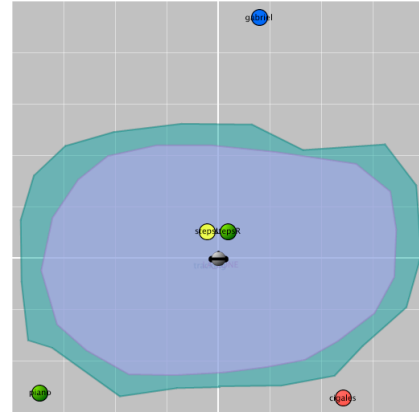
Ce test nous permet de mettre en évidence l'impact du retour sonore corporel dans la représentation mentale de l'espace. On teste ici si les sujets sont perturbés, même inconsciemment, par cette suppression.

« Probe »

La dernière partie de l'expérience nous permet de confirmer que la scène auditive virtuelle a été mémorisée et que ce n'est pas au hasard si le sujet trouve la cible : on enlève la source cible de la scène sonore - sans que les sujets ne le sachent - pour vérifier que les sujets cherchent vraiment la cible dans la zone où elle était placée jusqu'alors.

figure 15 : Visualisation de la scène du
« Probe » (réalisée avec Max/MSP)

- La zone de tracking est en vert et la zone de recherche en bleu/violet.
- Les trois ronds bleu/rouge/vert situés en dehors de la zone sont les trois sources statiques.
- L'auditeur est en gris.
- La cible est absente



L'essai a une durée fixe, il s'arrête automatiquement au bout de 2mn. Il est suffisamment long pour réaliser une analyse des trajectoires de convergence, de la portion de l'espace le sujet passe le plus de temps.

Le recueil des données se fait en respectant l'anonymat.

4. Analyse des résultats

Les données de 14 sujets sur les 20 qui se sont présentés à l'expérience ont été retenues pour l'analyse. Les multiples problèmes techniques rencontrés - liés au tracking mais également à l'utilisation de casque sans fil pour le rendu binaural - au cours de certaines expériences ne nous ont en effet pas permis d'analyser les données de 6 des 20 sujets. Nous avons finalement analysé les résultats de 5 femmes et 9 hommes.

Nous avons formulé certaines hypothèses concernant les performances des sujets dans les différentes configurations de tests, et toutes les analyses que nous avons réalisées se sont appuyées sur l'étude de la variable « temps ». Nous disposons en effet de 14 fichiers texte par sujet - 6 pour l'apprentissage, 7 pour les tests suivis des 'ré-immersions' et 1 pour le probe - donnant la valeur de ses coordonnées dans l'espace de tracking toutes les 50 ms. Nous pouvons donc retracer toutes les trajectoires suivies par les sujets au cours des différents tests, et donc avoir accès aux temps passés dans différentes zones de l'espace qui nous ont paru remarquables.

Les premières analyses ont été réalisées sous *Matlab*. Les analyses statistiques ont été réalisées à l'aide du logiciel *Statistica*.

4.1 Analyse de la phase d'apprentissage et du « Probe »

Analyse descriptive

La première analyse qu'il est nécessaire de conduire pour valider l'expérience concerne la phase d'apprentissage. S'il n'est pas clair que les sujets ont mémorisé la localisation de la source dans la scène sonore, ils n'ont pas pu se construire des repères auditifs utiles pour naviguer dans les scènes de tests.

Le temps total passé pour retrouver la cible dans les 6 essais de la phase d'apprentissage, depuis l'entrée du sujet dans la scène sonore jusqu'à sa stabilisation sur la cible, est une des données les plus parlantes qu'il a nous a fallu considérer pour mettre en évidence la mémoire spatiale. Si l'apprentissage se fait correctement, les sujets doivent converger de plus en plus rapidement sur la cible au cours des différents passages, et le temps total de l'essai doit donc diminuer significativement.

Les deux graphes présentés ci-après mettent ainsi en évidence la mémorisation spatiale des 14 participants dans la tâche de localisation de la phase d'apprentissage.

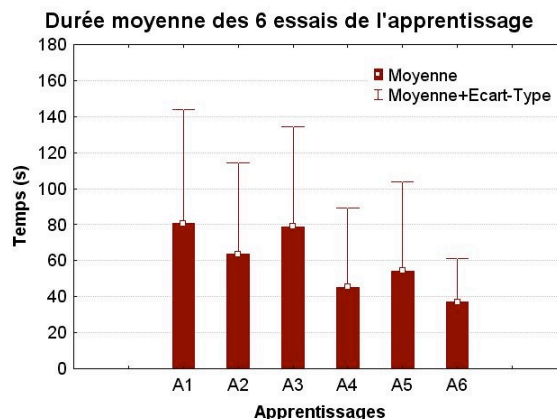
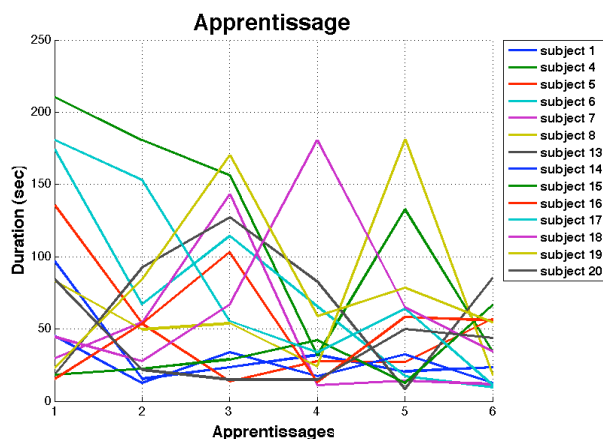
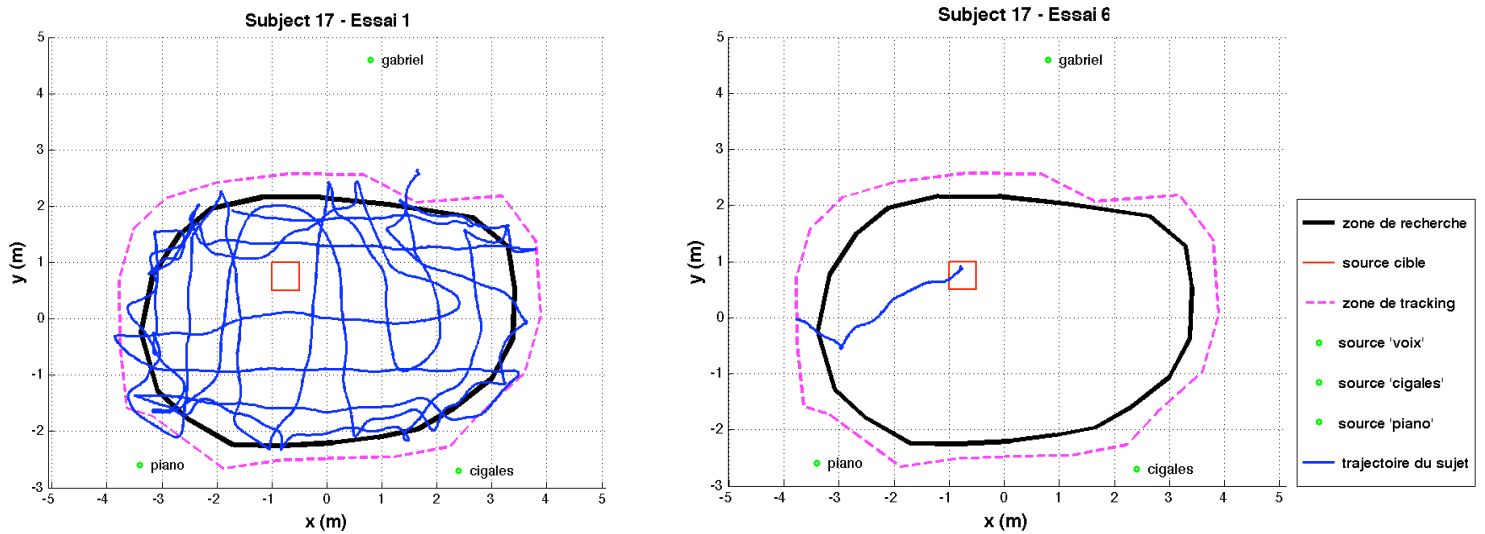


figure 16a : Performances des sujets au cours de la phase d'apprentissage
figure 16b : Répartition des temps moyennés sur l'ensemble des résultats pour chaque essai d'apprentissage

Sur la *figure 16a*, la majorité des temps totaux passés par les sujets à retrouver la cible au cours des 6 essais réalisés diminue et la dispersion des résultats se resserre au cours des essais pour finir nettement plus petite au 6^{ème} et dernier essai qu'au premier. Sur la *figure 16b*, c'est la moyenne sur tous les sujets du temps passé à chaque essai qui est représenté en barre ; l'écart-type y figure également. On note également une diminution du temps moyen passé et une diminution de l'écart-type (l'écart-type mesuré pour le premier essai est 63.46 et celui mesuré pour le dernier essai de l'apprentissage, essai 6, est 23.97).

La trajectoire décrite par les participants au cours des 6 essais nous permet également de rechercher des marques d'apprentissage spatial. En effet, après le premier essai qui ne sert finalement qu'à l'exploration de la scène puisque les sujets trouvent la cible que par hasard, la trajectoire suivie est un témoignage de l'utilisation de repères auditifs dans la scène sonore pour l'orientation, et de la convergence vers la localisation mémorisée de la source cible. Il en est ressorti que les sujets se sont principalement aidés d'un axe sonore (source « piano »/source « voix ») qu'ils ont balayé plusieurs fois jusqu'à retomber sur la cible.

L'enregistrement des coordonnées spatiales de sujets dans la scène sonore nous permet de tracer les trajectoires décrites au cours d'un essai. On obtient par exemple ce genre de résultats en analysant les trajectoires d'un sujet particulièrement consciencieux :



figures 17 a et b : Trajectoires d'un sujet pendant les essais 1 (17a) et 6 (17b) de la phase d'apprentissage

Sur la *figure 17a*, on remarque que lors du premier essai, le sujet a quadrillé la quasi-totalité de l'espace de recherche de manière méthodique, sans malheureusement tomber sur la source cible. Sur la *figure 17b* modélisant l'essai 6, soit après 4 autres essais, il converge très rapidement sur l'emplacement de la cible, ce qui est synonyme d'apprentissage spatial.

C'est grâce à l'obtention de ces trajectoires qu'il est possible de réaliser une « carte spatiale » dans cette tâche de localisation.

Nous avons effectué un partitionnement de l'aire de tracking sous *Matlab* en petites zones carrées de 0.1m^2 . A chaque fois qu'un point de la trajectoire d'un sujet se trouve dans un des petits carrés du quadrillage spatial, la valeur du temps passé dans ce petit carré augmente de 50 ms. L'outil *ColorMap* de *Matlab* permet de représenter cette augmentation du temps à l'aide d'une échelle de couleurs - du bleu foncé pour 0 seconde passée dans le carré jusqu'au rouge pour le maximum de temps passé en secondes - associées à des valeurs numériques en secondes.

En superposant les trajectoires des 14 sujets pour un essai de l'apprentissage donné, nous obtenons une carte des différentes parties de la scène sonore les plus visitées. Ainsi, on obtient les cartes suivantes, pour les 6 essais de l'apprentissage.

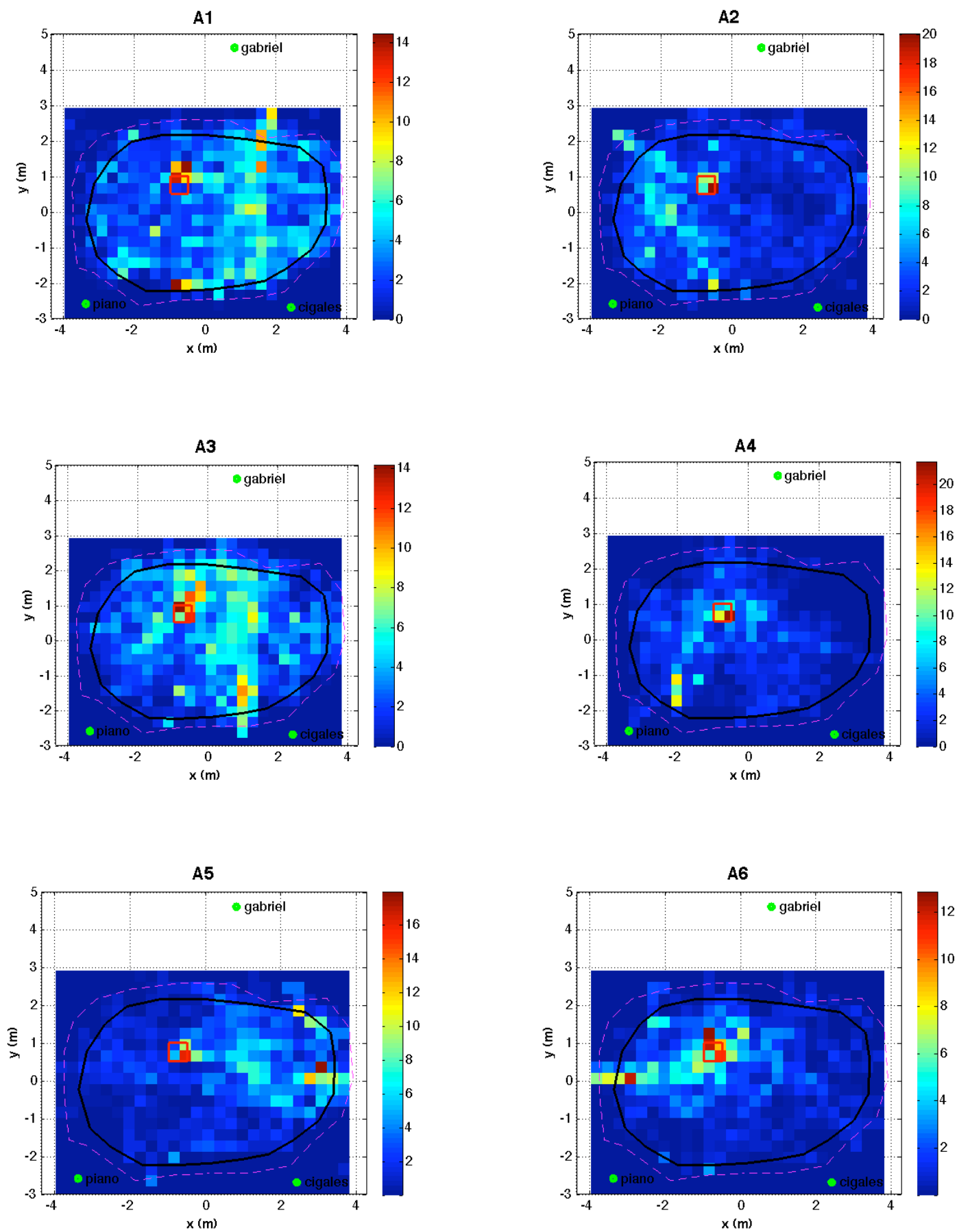


figure 18 : Cartes spatiales des 6 essais de la phase d'Apprentissage

Les figures présentées ci-dessus sont les représentations visuelles de l'apprentissage spatial. En effet, la carte associée au premier essai présente une distribution très diluée, aucune partie de l'espace de recherche ne ressort particulièrement car toutes les parties ont toutes été autant balayées. Il ne s'agit en effet que d'une première exploration de la scène sonore. On observe par la suite un re-centrage de la zone de recherche, l'aire n'étant plus parcourue en totalité - le bleu foncé ressort - la diminution du temps total passé est également visible.

La dernière phase de l'expérience, le « Probe » mérite d'être analysée en même temps que les essais de l'apprentissage car il s'agit d'un essai réalisé dans la même scène sonore que pendant l'apprentissage - outre le fait que l'on ait supprimé la source cible de la scène sonore - et il a la particularité d'être long - il dure 2mn pour tous les sujets- Ceci permet de bien mettre en évidence la mémorisation de l'emplacement de la cible puisque, ne trouvant pas la cible, les sujets tournent autour de la zone apprise. Avec la même méthode de tracé que pour les 6 essais de l'apprentissage, nous avons réalisé une carte spatiale de cette tâche de localisation. Le résultat est sans appel :

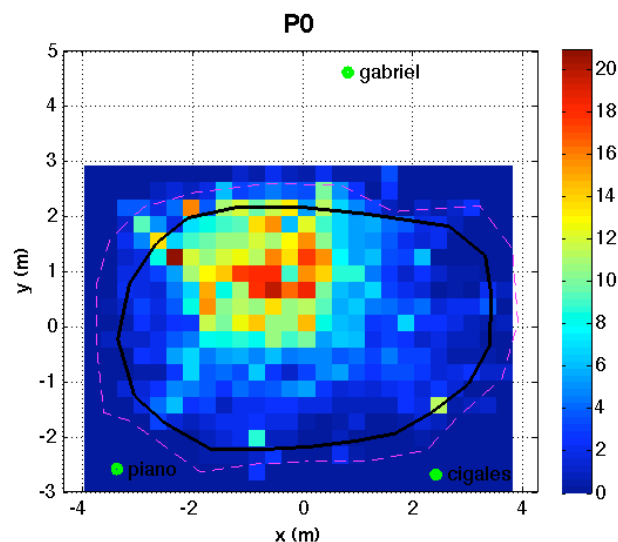


figure 19 : Carte spatiale du « Probe »

On peut observer que la recherche est cette fois réellement centrée sur la position de la cible.

Il est clair que les performances de tous les sujets ne sont pas les mêmes pendant l'expérience : certains sont très rapides et réalisent de très bonnes performances, d'autres le sont moins et auraient sans doute bénéficié de quelques essais de plus dans la phase d'apprentissage. L'essai du « Probe » est donc probablement encore plus explicite que le dernier essai de l'apprentissage car tous les sujets sans distinction ont passé 2mn dans la scène sonore. La représentation à l'aide d'une échelle de couleur dépendant du temps passé n'est ainsi pas tributaire de la rapidité des sujets mais se base simplement sur leurs performances de repérage cognitif. Il en ressort une carte spatiale plus « lissée ».

Analyse statistique

Une courte analyse statistique concernant l'ensemble de ces essais a été menée en parallèle de l'analyse descriptive pour vérifier la validité des observations.

Ces analyses statistiques ont comparé des données de temps passé dans différentes zones prédéfinies de l'espace. Il a été choisi de découper la zone ellipsoïdale de recherche en

8 « parts » d'aire identique et égale à $1/8^{\text{ème}}$ de la zone de recherche. La zone de recherche ayant recouvrant $24,74 \text{ m}^2$, les 8 petites zones font $3,09 \text{ m}^2$ - il est ensuite possible d'obtenir le découpage de la zone en 4 quarts par addition de 2 parts d'aire égale à $1/8^{\text{ème}}$ de la zone de recherche - La première « part » est définie de telle sorte que l'ouverture du rayon liant le barycentre de la zone de recherche au centre de la cible balaye une aire égale à $3,09 \text{ m}^2$. Le premier quart correspond alors au quartier contenant la source cible.

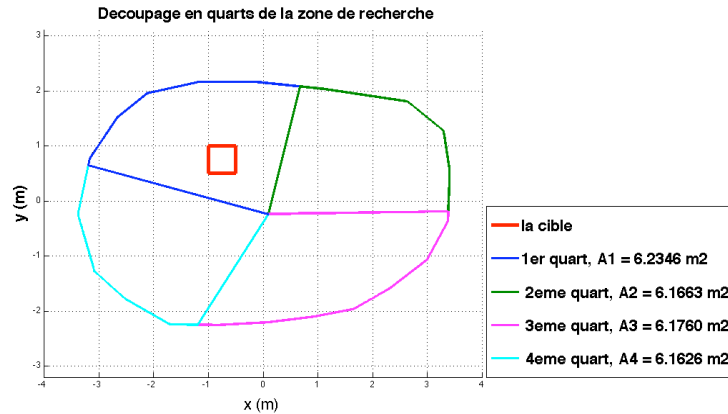


figure 20: Découpage de la zone de recherche de $24,74 \text{ m}^2$ en 4 quarts d'environ 6 m^2

Les temps passés par les 14 sujets dans ces différentes zones ont été relevés pour être analysés. Nous nous sommes basés sur le test de Wilcoxon pour réaliser cette analyse statistique. Le test de Wilcoxon est une alternative non-paramétrique au test de t de Student. Les tests non-paramétriques ont été développés pour traiter des problèmes dont on ne connaît pas les paramètres de la distribution de la variable étudiée dans la population. En d'autres termes, les méthodes non-paramétriques ne se fondent pas sur l'estimation de paramètre décrivant la distribution de la variable étudiée dans la population (comme la moyenne ou l'écart-type). Ce test est conçu pour tester une hypothèse sur la position (médiane) d'une distribution de population. Il implique d'utiliser des couples *appariés*, c'est-à-dire des données obtenues dans différentes conditions mais au sein d'une même population. Puisque dans notre cas, les mêmes sujets réalisent les essais d'apprentissage puis les tests, sous des conditions différentes, ce test est approprié.

Le test de Wilcoxon tel que présenté sur le logiciel *Statistica* renvoie la valeur de p. Dans un test statistique, la valeur p est le plus petit niveau auquel on rejette l'hypothèse nulle. Un résultat est dit *statistiquement significatif* lorsqu'il est improbable qu'il puisse être obtenu par un simple hasard. On utilise généralement un seuil de probabilité de 5% : si le résultat avait moins de 5% de chances d'être obtenu par hasard, alors il est jugé significatif.

L'analyse statistique du « Probe » par quartier, avec le test de Wilcoxon, montre que le premier quart (Q1) est significativement différent des trois autres, et que les 2^{ème} et 3^{ème} quarts (Q2 et Q3) ne sont pas significativement différents, ce qui signifie qu'ils occupent la même fonction dans la représentation cognitive de l'espace (voir table 1).

	Q1	Q2	Q3	Q4
Temps passé en secondes en moyenne(écart-type)	56.4 (23.9)*	15.8(16.4)**	12.26(10.1)**	42(15)

*différence significative entre Q1 et Q2 ($Z(14)=3.04$, $p < 0.05$), Q1 et Q3 ($Z(14)= 3.23$, $p<0.05$), Q1 et Q4 ($Z(14)= 2.04$, $p < 0.05$)

**différence non significative entre Q2 et Q3 ($Z(14)= 0.72$, $p > 0.05$), Q2 et Q4 ($Z(14)= 2.29$ $p < 0.05$)

***différence significative entre Q3 et Q4 ($Z(14)=3.1$, $p < 0.05$)

table 1 : Résultats du test de Wilcoxon sur les temps passés dans les quarts pendant le « Probe »

En se basant sur les pourcentages de temps total passé dans les différentes zones, l'analyse statistique des différents essais de l'apprentissage et du « Probe » montre que le premier quart - contenant la cible - dans le « Probe » n'est pas significativement différent de celui de l'essai 6 :

	Q1 (Essai 6)	Q1 (Probe)
Temps passé en pourcentage du temps total de l'essai (écart-type)	44.19 (15.68)*	52.1 (21.3)

*différence non significative entre Q1(Essai 6) et Q1(Probe) ($Z(14)=1.22$, $p < 0.05$)

table 2 : Résultats du test de Wilcoxon sur les pourcentages de temps passés dans les 1ers quarts de l'essai 6 et du « Probe »

De plus, le temps passé dans ce premier quart pendant l'essai 6 étant significativement différent du temps passé dans les autres quarts, on peut donc considérer qu'un apprentissage s'est fait dès le 6^{ème} essai de la première phase.

4.2 Analyse de la phase de test

L'établissement d'une carte spatiale pour les 4 situations de tests a également été réalisé. Il s'agit de voir si les distributions spatiales de temps sont concentrées sur la cible , à l'image du dernier essai de l'apprentissage, ou si elles sont plus uniformes spatialement, ce qui indique que toute la scène sonore a été ré-explorée. On pourrait imaginer que dans ce dernier cas, le sujet déboussolé par une modification de la scène a cru qu'il s'agissait d'une toute nouvelle configuration sonore et a donc parcouru à nouveau la totalité de la zone de recherche.

Analyse descriptive du test 1

On obtient ainsi pour le test 1 - où le rendu sonore des pas a été délocalisé sur la gauche des sujets - la carte suivante, utilisant le même code de couleur que sur les cartes présentées précédemment :

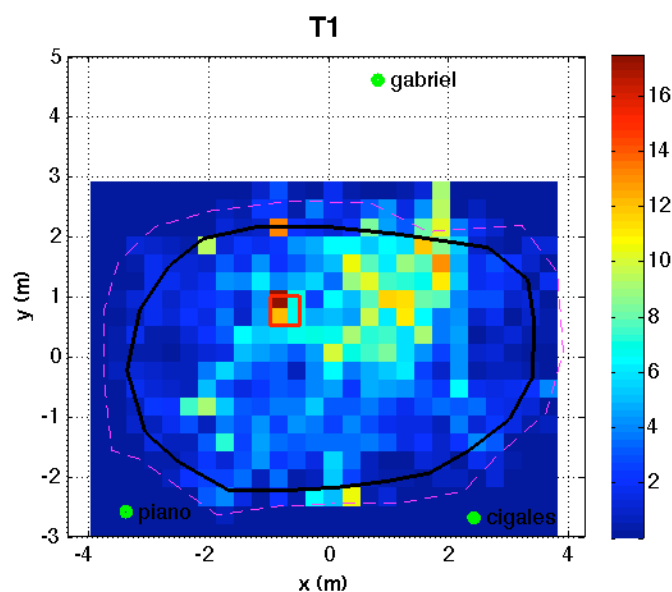


figure 21 : Carte spatiale du test 1

On observe sur cette carte une distribution assez dense sur la droite de la cible, ce qui pourrait laisser supposer que le retour sonore des pas délocalisés à gauche ont conduit le sujet à se positionner toujours plus à droite de toute la scène sonore (sources statiques + cible) que ce qu'il faisait dans la phase d'apprentissage. L'analyse statistique des temps passés dans les quartiers de la scène sonore vient confirmer cette constatation : le 2^{ème} quart de la zone est significativement différent des 3^{ème} et 4^{ème} quarts (voir table 3).

	Q1	Q2	Q3	Q4
Temps passé en secondes en moyenne(écart-type)	22.1 (16.9)	24.5 (28.9)*	13.1 (12.7)	14 (16.8)

*différence non significative entre Q2et Q1 ($Z(14)=0.53$, $p > 0.05$)

différence significative entre Q2 et Q3 ($Z(14)=2.29$, $p < 0.05$)

différence significative entre Q2 et Q4 ($Z(14)= 2.35$, $p < 0.05$)

table 3 : Résultats du test de Wilcoxon sur les temps passés dans les quarts du test 1

Une autre analyse montre que le test 1 est significativement différent du dernier essai de l'apprentissage (essai 6) ce qui prouve une baisse des performances, consécutives aux modifications sonores effectuées.

Il a donc été intéressant de reconstituer la même carte en ne considérant non pas la localisation de l'axe des sujets (donnée par le marqueur porté sur la tête) mais celle de leur pas (donné par le marqueur porté dans le dos).

L'idée est de voir si cette distribution est plus "diluée" que celle de l'axe des sujets - les pas étant une espèce de "satellite" orienté dans n'importe quelle direction - ou si au contraire cette distribution est davantage concentrée - les pas ayant dans ce cas un rôle "pivot" -

Il est de plus intéressant de regarder si cette distribution est davantage centrée sur la cible.

En considérant la localisation des pas, nous obtenons la carte suivante :

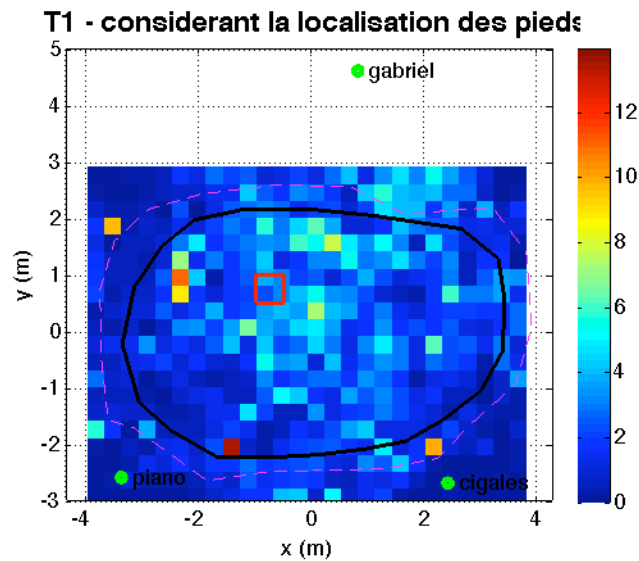


figure 22 : Carte spatiale du test 1, réalisée à partir de la localisation des pieds des sujets « délocalisés sur leur gauche »

Il apparaît que la distribution est nettement plus uniforme que la précédente, ce qui écarte l'hypothèse que les pas jouent le rôle de pivot dans l'expérience. Cependant, on peut avancer que les problèmes de tracking que nous avons rencontrés concernant le marqueur du dos lié à la position des pieds du sujet, y sont pour beaucoup dans la carte obtenue. En effet, lorsque les caméras perdent la trace du marqueur, l'enregistrement des coordonnées spatiales du marqueur reste constamment le même jusqu'à ce que les caméras repère à nouveau le marqueur. Il s'agit souvent d'une perte de tracking aux frontières de la zone de tracking, d'où le temps affiché passé en sortie de zone. Il s'ensuit que le fichier d'enregistrement des positions contient beaucoup de positions consécutives identiques. La carte spatiale tracée en fonction de la position des pieds est alors erronée puisque le temps effectivement passé au point de perte de tracking était bien plus court que celui affiché par les couleurs de l'échelle de temps adoptée.

Un exemple de perte de suivi de position lors du passage de l'un des sujets dans ce test est présenté ci-après:

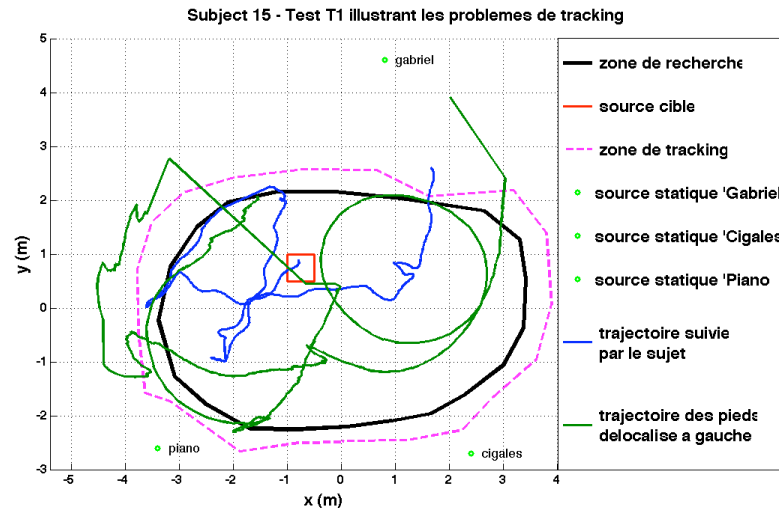


figure 23 : Illustration du problème de tracking du marqueur porté dans le dos et lié à l'avatar sonore des pieds dans le test 1

Sur la *figure 23*, la trajectoire bleue représente la trajectoire effectivement suivie par le sujet au cours de son passage, et la trajectoire verte représente les positions des pieds « délocalisés sur le gauche du sujet », donc sur la gauche de sa trajectoire réelle, au cours de ce même passage. On remarque qu'il existe différents points d'accroche du marqueur de position des pieds, et souvent en sortie de zone, là où les problèmes de tracking sont les plus fréquents. On notera par exemple la perte de tracking aux coordonnées ($x = -3.19$, $y = 2.77$) et la longueur du « saut » effectué par le suivi de position de ce marqueur pour être que le marqueur soit à nouveau repéré par les caméras.

Nous aurions espéré pouvoir réaliser une carte spatiale centrée sur la position des pieds dans ce test 1 pour l'ensemble des participants, mais nous sommes bien malheureusement dépendants des performances de la technologie utilisée.

C'est à la suite de ce test que nous avons posé les 6 questions concernant la sensation de présence. Pour un grand nombre de sujets, ce sont ces questions concernant explicitement les pas qui les ont fait prendre conscience de la présence de ces sons.

Analyse descriptive du test 2

On obtient pour le test 2 - où les sons des sources statiques ont été remplacés par des sons synthétiques - la carte suivante, utilisant le même code de couleur que sur les cartes présentées précédemment :

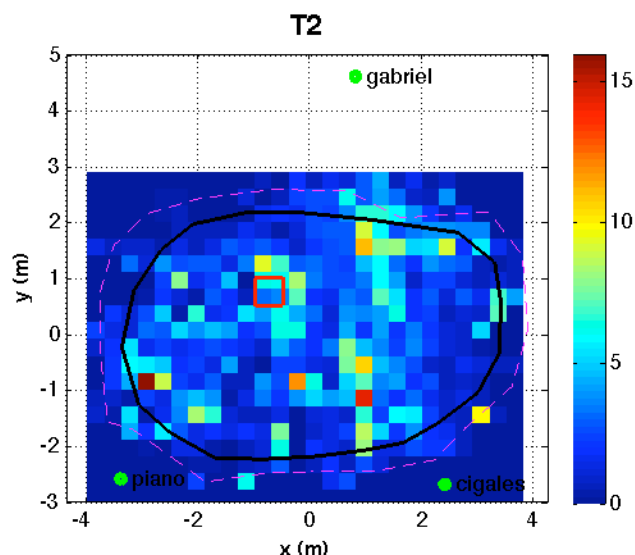


figure 24 : Carte spatiale du test 2

On remarque que la distribution des temps est « diluée » spatialement. Ceci prouve que les participants ont complètement ré-exploré la scène sonore : il semble qu'ils aient été très déroutés par le changement brusque d'environnement sonore, bien que la structure géométrique de la scène soit restée identique.

Les repères utilisés dans cette expérience par les sujets pour la localisation spatiale reposeraient donc peut-être sur un codage plus sémantique que uniquement spatial. Dans cette expérience, on pourrait penser que l'un a été préféré à l'autre, en l'occurrence le codage sémantique a été préféré au codage spatial. Il reste néanmoins que l'alternance entre les deux types de codage une fois les repères construits est manifestement une tâche difficile.

Analyse statistique du test 2

	Q1	Q2	Q3	Q4
Temps passé en secondes en moyenne(écart-type)	14.7 (9.6)*	18.34 (24)**	16.9 (17.1)***	16.1 (24.7)

*différence non significative entre Q1et Q2 ($Z(14)=0.41$, $p > 0.05$)

différence non significative entre Q1 et Q3 ($Z(14)=0.03$, $p > 0.05$)

différence non significative entre Q1 et Q4 ($Z(14)= 0.4$, $p > 0.05$)

**différence non significative entre Q2et Q3 ($Z(14)=0.03$, $p > 0.05$)

différence non significative entre Q2 et Q4 ($Z(14)=0.09$, $p > 0.05$)

***différence non significative entre Q3et Q4 ($Z(14)=0.78$, $p > 0.05$)

table 4 : Résultats du test de Wilcoxon sur les temps passés dans les quarts du test 2

Aucun des quarts n'est significativement différent des autres pendant ce test. Toutes les zones ont été parcourues de la même manière, comme lors d'une première exploration de scène sonore.

Analyse descriptive du test 3

On obtient pour le test 3 - où la zone de recherche a été divisée en deux zones de sons de pas - la carte spatiale suivante, utilisant le même code de couleur que sur les cartes présentées précédemment :

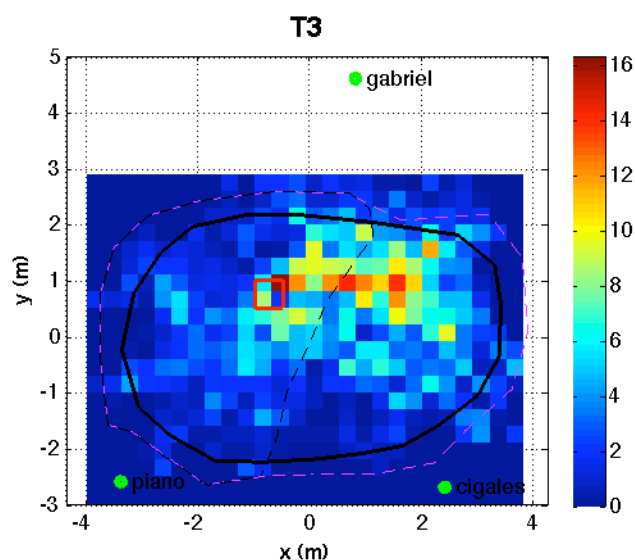


figure 25 : Carte spatiale du test 3

La « frontière » entre les deux zones de sons de pas est représentée en pointillés sur la carte.

On observe que les sujets sont longtemps restés le long de cette frontière, proche de la source « voix », la distribution étant plus concentrée à cet endroit. Le retour sonore des pas avait donc été intégré comme un élément constitutif de la scène, même s'il n'a pas pu être utilisé comme indice spatial utile au repérage de la cible : les sons de marche sur la paille sont en effet représentatifs de l'espace d'évolution de la phase d'apprentissage, alors que les sons de marche sur les graviers représentent un nouvel espace sonore. De plus, la réverbération du lieu étant portée par la réverbération de ces sons de pas, les sons de pas émis portent en eux non seulement les propriétés sémantiques mais également les propriétés acoustiques du lieu mémorisé. Les participants ont mis du temps à dépasser franchement la frontière car le retour sonore des sons de pas ayant changé, cela suggérait que l'environnement total avait changé et que la cible ne pouvait pas s'y trouver.

Analyse statistique du test 3

Finalement, l'analyse statistique par test de Wilcoxon comparant les deux zones de sons de pas ne permet pas de mettre en évidence qu'une des deux zones a été plus parcourue que l'autre ; ce résultat est compréhensible car les sujets sont restés à la frontière entre les deux. Cependant, une analyse statistique sur les temps passés dans les quarts de la zone de recherche montre que les 1^{er} et 2^{ème} quarts ne sont pas significativement différents entre eux, mais significativement différents des 3^{ème} et 4^{ème} quarts.

	Q1	Q2	Q3	Q4
Temps passé en secondes en moyenne(écart-type)	18.33 (9.1)*	28.1 (30.6)**	12 (16.5)***	12.8 (13)

*différence non significative entre Q1 et Q2 ($Z(14)=0.22$, $p > 0.05$)

différence significative entre Q1 et Q3 ($Z(14)=2.23$, $p < 0.05$)

différence significative entre Q1 et Q4 ($Z(14)=1.6$, $p < 0.05$)

**différence significative entre Q2 et Q3 ($Z(14)=2.8$, $p < 0.05$)

différence non significative entre Q2 et Q4 ($Z(14)=1.16$, $p > 0.05$)

***différence non significative entre Q3 et Q4 ($Z(14)=0.8$, $p > 0.05$)

table 5 : Résultats du test de Wilcoxon sur les temps passés dans les quarts du test 3

Une autre analyse montre que le test 3 est significativement différent du dernier essai de l'apprentissage (essai 6) ce qui prouve une baisse des performances, consécutives aux modifications sonores effectuées.

Analyse descriptive du test 4

On obtient pour le test 4 - où les sons de pas ne font plus partie de la restitution sonore diffusée au casque du sujet- la carte spatiale suivante, utilisant le même code de couleur que sur les cartes présentées précédemment :

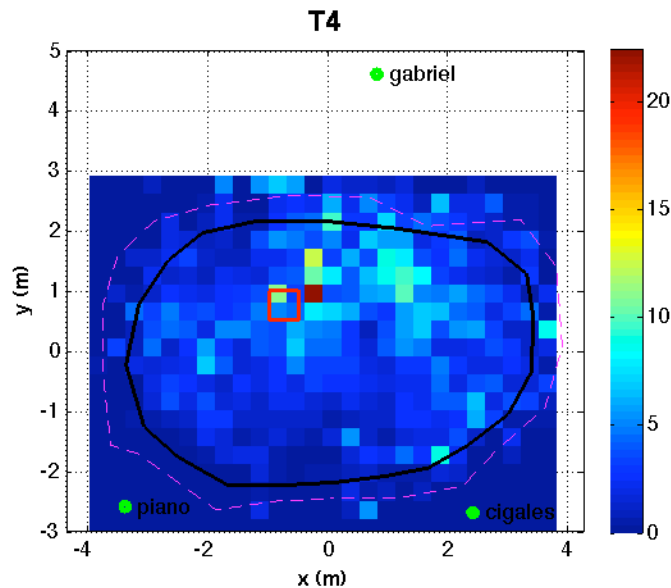


figure 26: Carte spatiale du test 4

Cette carte représentant une distribution relativement dispersée, nous en déduisons que la durée du test et la répartition spatiale des sujets dans la scène sont dues à des effets de désorientation lié à la suppression du retour sonore des pas ; la configuration de la scène sonore statique restant elle, identique à celle de la phase d'apprentissage. Ce test confirme que les sujets ont une conscience même partielle de leur présence dans le monde sonore et du fait que leurs mouvements peuvent modifier cet espace sonore environnant. Dès la suppression de

ce retour sonore, changement dont les sujets prennent ou non conscience, la difficulté de la tâche augmente par rapport à la phase d'apprentissage.

Analyse statistique du test 4

L'analyse statistique réalisée à l'aide du test de Wilcoxon pour comparer les temps passés dans les différents quarts de la zone de recherche montre que ces quarts sont globalement équivalents. Le tableau suivant donne les résultats de l'analyse :

	Q1	Q2	Q3	Q4
Temps passé en secondes en moyenne(écart-type)	18.33 (9.1)*	28.1 (30.6)**	12 (16.5)***	12.8 (13)

*différence non significative entre Q1et Q2 ($Z(14)=0.22$, $p > 0.05$)

différence non significative entre Q1 et Q3 ($Z(14)=1.28$, $p > 0.05$)

différence non significative entre Q1 et Q4 ($Z(14)= 1.47$, $p > 0.05$)

**différence significative entre Q2et Q3 ($Z(14)=2.29$, $p < 0.05$)

différence non significative entre Q2 et Q4 ($Z(14)=0.9$, $p > 0.05$)

***différence non significative entre Q3et Q4 ($Z(14)=0.56$, $p > 0.05$)

table 6 : Résultats du test de Wilcoxon sur les temps passés dans les quarts du test 3

Les sujets avaient en effet choisi d'utiliser certains repères auditifs pour la localisation, mais tous avaient en tête ces bruits de paille même s'ils n'y prêtaient que peu d'attention. En supprimant ce retour sonore, la convergence sur la cible est moins nette, tous les quarts de la zone de recherche sont à nouveau balayés.

Remarque générale

Une analyse statistique réalisée avec le test de Wilcoxon nous a permis de comparer deux à deux les apprentissages et les tests/ré-immersion dont les portes d'entrée dans la scène sonore sont les mêmes. En effet, on peut se demander si le n° de porte d'entrée n'a pas une incidence sur les résultats. On constate par exemple que les acquisitions de ré-immersion 7 et 8 suivant respectivement les tests 1 et 2 sont assez "laborieuses" comparativement aux acquisitions de la phase d'apprentissage précédentes 4, 5.

Les résultats de cette analyse montrent que les tests/ré-immersions et les essais de la phase d'apprentissage dont les portes d'entrée dans la scène sont les mêmes, sont pour certains, significativement différents, pour d'autres non. Autrement dit, il en ressort que une durée d'acquisition longue est due à un effet de "désorientation" liée à la difficulté de la tâche et non à une porte d'entrée plus "difficile".

Conclusion générale

Les expériences menées mettent en évidence la capacité humaine à construire une représentation mentale de l'organisation spatiale d'une scène, sur la seule base de l'interaction auditive - informations auditives fournies par l'espace sonore - et idiothétique au cours d'une action motrice.

Ce type de situation fait instinctivement penser à l'étude de la perception chez les sujets aveugles mais elle n'est testée pour l'instant dans l'équipe que sur des sujets normo-voyants. La condition de déprivation visuelle imposée - il est nécessaire d'isoler sa contribution pour étudier la modalité auditive car le canal visuel est bien connu pour avoir le rôle principal dans une tâche de localisation- permet de mettre en évidence que chez l'Homme, les indices auditifs participent automatiquement au repérage spatial. La construction d'une carte cognitive spatiale chez les sujets intervient même en l'absence de construction d'un équivalent visuel.

Les résultats de ces expériences fournissent donc des indices précieux sur la contribution exceptionnelle de la modalité auditive aux représentations spatiales. Ces expériences suggèrent en particulier que le retour auditif du corps participe au repérage spatial.

Par la suite, il sera donc possible de réaliser l'expérience du Water Maze en utilisant les modèles de synthèse binaurale développés pendant mon stage, en particulier le « snowman model » en condition d'expérience réelle, ce qui représenterait une alternative intéressante aux mesures de HRTF individuelles. Concernant l'avatar sonore lié au pas du sujet, il serait intéressant d'analyser l'influence de l'*anticipation* et du *retard* du retour sonore des pas sur les performances du sujet dans une tâche de localisation. L'utilisation du « gesture follower » - suivi de geste - développé par l'équipe Temps Réel de l'Ircam est une solution que nous avons envisagée durant mon stage mais l'utilisation de ce logiciel n'a pas été suffisamment testée avant l'expérience, par manque de temps, pour que nous puissions l'inclure dans le protocole.

Finalement, la conception d'environnements interactifs excluant la modalité visuelle soulève des questions scientifiques stimulantes et présente un large potentiel d'applications innovantes dans des domaines aussi divers que le jeu, les installations artistiques et la musique.

Bibliographie

- [1] - R. G. M. Morris (1981) « Spatial Localization Does Not Require the Presence of Local Cues » **Learning and Motivation** volume 12 n°2, pp 239-260.
- [2] - R. G. M. Morris, P. Garrud, J. N. P. Rawlins & J. O'Keefe (1982) « Place navigation impaired in rats with hippocampal lesions » **Nature**. volume 297, pp 681 - 683
- [3] - Noah J. Sandstrom, Jordy Kaufman, Scott A. Huettel (1998) « Males and females use different distal cues in a virtual environment navigation task », **Cognitive Brain Research** volume 6, pp 351–360
- [4] - Robert S. Astur a, Jennifer Tropp, Simona Sava, R. Todd Constable, Etan J. Markus (2004) « Sex differences and correlations in a virtual Morris water task, a virtual radial arm maze, and mental rotation » **Behavioural Brain Research** volume 151, pp 103–115
- [5] - Robert S. Astur , Maria L. Ortiz , Robert J. Sutherland (1998) « A characterization of performance by men and women in a virtual Morris water task: A large and reliable sex difference » **Behavioural Brain Research** , volume 93, pp 185–190
- [6] - Derek A. Hamilton, Ira Driscoll, Robert J. Sutherland (2002) « Human place learning in a virtual Morris water task: some important constraints on the flexibility of place navigation » **Behavioural Brain Research**, volume 129, pp 159–170
- [7] - Janos Kalliaia, Tamas Makany, Kazmer Karadi, William J. Jacobs (2005) « Spatial orientation strategies in Morris-type virtual water task for humans », **Behavioural Brain Research**, volume 159, pp 187–196
- [8] - Paul M. Hofman, Jos G.A. Van Riswick and A. John Van Opstal (1998) « Relearning sound localization with new ears », **Nature**, volume 1 no 5, pp 417-421
- [9] - Mark R. Aderson, Durand R. Begault, Elisabeth M. Wenzel (2001) « Direct Comparison of the Impact of Head Tracking, Reverberation, and Individualized Head-Related Transfer Functions on the Spatial Perception of a Virtual Speech Source », **J. Audio Eng. Soc.**, volume 49, no.10
- [10] - Elisabeth M. Wenzel, Marianne Arruda and Doris J. Kistler, Frederic L. Wightman (1993) « Localization using nonindividualized head-related transfer Functions », **J. Acoust. Soc. Am.**, volume 94, no. 1, pp 111-123
- [11] - H. Moller, C. Jensen, D. Hammerhoi, M. Sorensen (1996) « Selection of a typical human subject for binaural recording » **Acta Acustica**, n° 82, p 215
- [12] - J. Blauert (1997) « Spatial Hearing » (revised edition) **MIT Press**, Cambridge, MA
- [13] - R.O. Duda, W.L. Martens (1998) « Range of dependence of the response of a spherical head model », **J. Acoust. Soc. Am.**, volume 104, n° 5, pp 3048-3058
- [14] - R. Woodworth, H. Schlosberg (1954) « Experimental Psychology », **Methew**, p12
- [15] - G. F. Kuhn (1977) « Model for the interaural time difference in the azimuthal plane » **J. Acoust Soc. Am.**, volume 62, n°1, pp 157-167
- [16] - V. Ralph Algazi, Richards O. Duda, Dennis M. Thompson (2002 October 5-8) « The use of head-and-torso models for improved spatial sound synthesis », **Proceedings of the 113th Convention of the Audio Engineering Society**
- [17] - V. R. Algazi, C. Avendano, R. O. Duda (2002) « Elevation localization and head-related transfer function analysis at low frequencies », **J. Acoust Soc. Am.**, volume 109, n° 3, pp 1110-1122