Sonification binaurale pour l'aide à la navigation

Gaëtan Parseihian

Thèse de doctorat de l'Université Pierre et Marie Curie

École Doctorale SMAER

Spécialité : Acoustique

Présentée par

Gaëtan PARSEIHIAN

Pour obtenir le titre de
DOCTEUR de L'UNIVERSITÉ PARIS 6

Sujet :

Sonification binaurale pour l'aide à la navigation

préparée

au Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur
(CNRS UPR 3251)

Soutenue le 23 octobre 2012 devant le jury composé de

Christophe d'Alessandro	Directeur de thèse
Durand R. Begault	Rapporteur
Brian F.G. Katz	Co-directeur de thèse
Richard Kronland Martinet	Rapporteur
Jean-Dominique Polack	Examinateur
Patrick Susini	Examinateur

Les recherches originales ont ceci de particulier qu'elles se situent ordinairement en marge des disciplines reconnues et cataloguées officiellement. Tandis que la phonétique, l'acoustique, l'électronique, se partagent le champ d'investigation des spécialistes et que, d'autre part, les conservatoires pratiquent scrupuleusement l'empirisme musical traditionnel, le fossé reste ouvert entre le domaine des sciences expérimentales et celui de l'expérience esthétique.

Pierre Schaeffer

Remerciements

Je voudrais tout d'abord remercier Brian François Grégory Katz pour ces années passées dans la confiance et l'enthousiasme, pour son aide dans les moments difficiles de la thèse et du projet NAVIG ainsi que pour les qualités et défauts qu'il a su me transmettre. Merci à Christophe d'Alessandro d'avoir accepté d'être victime de la loi 1984 (qui dit que pour être habilité à diriger une thèse il faut en avoir encadré plusieurs) et donc d'avoir encadré Brian sur mon encadrement tout en tentant même des fois de me cadrer.

Je remercie vivement Durand R. Begault (venu de Californie pour l'occasion) et Richard Kronland Martinet pour leur rapports détaillés ainsi que Patrick Susini et Jean-Dominique Polack pour avoir participé à l'évaluation de ce travail.

Ce travail n'aurait pas eu lieu sans Markus Noisternig qui m'a dirigé vers Brian lorsque je trainais dans les couloirs de l'IRCAM cherchant désespérément une bourse de thèse. Markus, j'espère pouvoir faire la même chose pour toi un jour en t'envoyant un bon étudiant !

Je tiens à remercier toutes les personnes qui ont contribué, de près ou de loin, à la réalisation de ce travail par leurs conseils, leurs remarques ou en ayant participé aux différents tests perceptifs que j'ai menés durant la thèse. Les camarades de thèse très proches du LIMSI : Tifanie pour les pauses/discussions, les expéditions à la piscine, ... Maitre Kemar pour les disputes. Le hip-hop qui tache ? La bonne humeur excessive et surjouée et pour tout le reste. Popo aka Paulo, le mec le plus posé du labo. Les autres collègues de l'équipe AA ainsi que ceux du LIMSI. Un petit big up à mes stagiaires Marie, Victor, Mathias et Simon. Ça m'a fait plaisir de travailler avec vous. Merci aux collègues du projet NAVIG avec une mention spéciale pour Adrien : les semaines d'intégrations étaient bien cool grâce à toi !!! Merci à Malika, Mathieu et Lourdes.

Merci à Lulu, pour le travail et le reste. Tu vois malgré mes moqueries incessantes sur ton boulot, je n'ai pas été imperméable à l'ergonomie et tu as bien inspiré mes travaux ! À Lise et Emilien !!! On a tous fini par la finir notre thèse. Pas mal, hein ?! À Aurelie et Jojo mes chères colocs, vous avez été d'un soutient sans faille durant deux ans, j'espère que les coloquements perdureront ! À mes trois compères de Brane Project : Beb, Clément et Raph !

Enfin ce travail n'aurait pas eu lieu sans le soutien de ma famille et les multiples relectures de documents et du manuscrit des différents côtés (mon père, ma mère, ma belle soeur, ...). Un grand merci à ma petite mamounette qui m'a permis d'aller jusque là.

Résumé

Dans cette thèse, nous proposons la mise en place d'un système de réalité augmentée fondé sur le son 3D et la sonification, ayant pour objectif de fournir les informations nécessaires aux non-voyants pour un déplacement fiable et sûr. La conception de ce système a été abordée selon trois axes.

L'utilisation de la synthèse binaurale pour générer des sons 3D est limitée par le problème de l'individualisation des HRTF. Une méthode a été mise en place pour adapter les individus aux HRTF en utilisant la plasticité du cerveau. Évaluée avec une expérience de localisation, cette méthode a permis de montrer les possibilités d'acquisition rapide d'une carte audio-spatiale virtuelle sans utiliser la vision.

La sonification de données spatiales a été étudiée dans le cadre d'un système permettant la préhension d'objet dans l'espace péripersonnel. Les capacités de localisation de sources sonores réelles et virtuelles ont été étudiées avec un test de localisation. Une technique de sonification de la distance a été développée. Consistant à relier le paramètre à sonifier aux paramètres d'un effet audio, cette technique peut être appliquée à tout type de son sans nécessiter d'apprentissage supplémentaire.

Une stratégie de sonification permettant de prendre en compte les préférences des utilisateurs a été mise en place. Les « morphocons » sont des icônes sonores définis par des motifs de paramètres acoustiques. Cette méthode permet la construction d'un vocabulaire sonore indépendant du son utilisé. Un test de catégorisation a montré que les sujets sont capables de reconnaître des icônes sonores sur la base d'une description morphologique indépendamment du type de son utilisé.

Mots clés : réalité augmentée, son 3D, sonification, aide à la navigation, perception spatiale, plasticité auditive.

Abstract

This manuscript presents an augmented reality system based on 3D sound and sonification whose aim is to provide navigation assistance for visually impaired users. The design of this system has been addressed in three ways.

First, 3D sound generation via binaural synthesis has limitations due to the problem of the need for HRTF individualisation. A new method based on brain plasticity is established to adapt individuals to HRTFs using an audio-kinaesthetic platform, reversing the standard paradigm. This method has shown the potential for a rapid adaptation of the auditory system to virtual auditory cues without the use of vision.

Second, spatial data sonification is investigated in the context of a system for locating and grasping objects in the peripersonnel space. Sound localization performance was examined by comparing real and virtual sound sources. On the basis of the results, a distance sonification method is developed with the aim of improving user performance. Rather than employing sonification by sound synthesis, the proposed sonification method varies parameters of an audio effect that is applied to a base sound. This method allows the user to select and change the base sound without requiring additional learning.

Finally, we present the concept of a new method of sonification designed to answer end-user needs in terms of aesthetics and sonification customization. “Morphocons” are short audio units whose aim is the construction of a sound vocabulary based on the temporal evolution of sound. An identification test highlights the efficiency of morphocons for conveying the same information with various types of sounds.

Keywords: augmented reality, 3D sound, sonification, navigation aid, spatial perception, auditory plasticity.

Table des matières

Table des figures

Liste des tableaux

Chapitre 1

Introduction

1.1 Contexte

Ces dernières années, le développement des ordinateurs personnels et de nombreux autres dispositifs électroniques a permis aux domaines du son spatialisé et de la sonification de faire leur apparition dans de multiples applications commerciales. Issues de la recherche en acoustique et en interfaces homme-machine, et premièrement utilisées dans des applications comme les jeux vidéos ou l'informatique, ces technologies commencent à émerger dans les systèmes permettant d'assister un utilisateur dans la réalisation d'une tâche spécifique.

Parmi ces systèmes, les dispositifs d'aide à la navigation pour les non-voyants sont des candidats idéals à l'utilisation de ces deux technologies. Ces dispositifs, dont le but est de permettre à l'utilisateur de se déplacer d'un endroit à un autre d'une manière sûre et fiable, nécessitent la transmission d'informations cartographiques et visuelles. Pour cela, il est nécessaire d'utiliser une modalité sensorielle accessible au non-voyant. Un retour auditif paraît plus approprié qu'un retour haptique car il permet plus facilement de transmettre des informations complexes. Pourtant la majorité des non-voyants y est réfractaire. Les trois principales raisons de ce rejet sont : un retour sonore susceptible de masquer l'environnement sonore ambiant, les instructions vocales souvent longues et perturbantes et les sons non vocaux jugés trop désagréables par les utilisateurs pour une utilisation quotidienne.

Le but du travail présenté dans cette thèse, et effectué dans le cadre du projet ANR-NAVIG [1], est de concevoir un dispositif de guidage auditif permettant de transmettre des informations à l'utilisateur en utilisant, d'une part des sons 3D (par l'intermédiaire de la synthèse binaurale), et d'autre part la sonification (technique visant à transmettre des informations sous la forme de sons non vocaux). La combinaison de ces deux technologies devant permettre :

1. Le projet ANR-NAVIG (Navigation Assistée par VIsion artificielle et Gnss) a pour but de permettre aux utilisateurs non-voyants de se déplacer vers une destination voulue, de façon fiable et sûre, sans interférer avec leur comportement de déplacement habituel. En plus de l'aide au déplacement et à l'orientation, le dispositif doit permettre de localiser et saisir des objets en champ proche sans nécessité de les pré-équiper avec un composant électronique. Les différents éléments du système NAVIG sont détaillés dans l'annexe B.

– de guider l'utilisateur avec des sons simulant des trajectoires sonores virtuelles (plutôt qu'en décrivant le trajet verbalement),
– d'indiquer à l'utilisateur la position d'un objet ou d'un lieu en plaçant une source sonore virtuelle à son emplacement.

L'objectif final vise à transmettre des informations visuelles ou cartographiques en superposant à l'environnement réel des éléments sonores virtuels.

1.2 Objectifs

Nous pensons que l'utilisation de la réalité augmentée, par l'intermédiaire de la sonification binaurale générée sur des casques stéréophoniques osseux (ne masquant pas les sons ambiants), peut contribuer à diminuer les critiques émises sur l'utilisation du son par les utilisateurs non-voyant et aider à mettre au point un système qui puisse satisfaire leurs besoins tant du point de vue de l'efficacité que du confort d'utilisation.

Cependant, la synthèse binaurale nécessite l'acquisition d'un grand nombre de mesures, appelées HRTF (pour Head Related Transfer Function), qui doivent être effectuées sur chaque individu. Ces mesures nécessitent un dispositif lourd et onéreux et il est compliqué d'utiliser des HRTF individuelles dans le cadre d'un projet commercial. Les sons 3D sont donc générés avec des HRTF non-individuelles entraînant ainsi des dégradations au niveau des performances de localisation qui peuvent se révéler handicapantes dans le contexte d'un système d'aide à la navigation. De plus, même dans le cas idéal de localisation de sons réels, les performances du système auditif sont loin d'être aussi précises que les performances du système visuel. Ces performances, si elles sont bien documentées pour les sources lointaines et qu'elles peuvent être jugées acceptables pour une aide à la navigation, ont été peu étudiées dans le cas de sources proches et peuvent présenter des dégradations entraînant des problèmes de localisation. Le son 3D risque par conséquent d'être jugé inutile si les informations qu'il transmet sont jugées floues ou erronées. La sonification quant à elle, bien qu'elle offre de grandes possibilités pour la transmission d'informations sous forme sonore, est souvent basée sur des sons de synthèse qui peuvent être jugés non naturels et désagréables par les utilisateurs. L'absence de prise en compte de critères esthétiques pour la mise en place du guidage sonore risque de mener à un rejet du système par les utilisateurs.

Afin de mettre en place un système efficace mais surtout utilisable dans la vie courante, il apparaît nécessaire de trouver des solutions à ces problèmes. Basée sur ce constat, la problématique de cette thèse est la suivante : est-il possible d'utiliser le son 3D et la sonification dans un dispositif d'aide à la navigation pour les non-voyants ?

L'objectif de ce manuscrit est d'apporter des éléments de réponse à cette problématique en l'abordant selon trois axes principaux :

- Amélioration du rendu binaural avec des HRTF non-individuelles
  ⇒ *Comment régler le problème de l'individualisation des HRTF ?*

- Amélioration des indices de localisation par l'utilisation de la sonification

  ⇒ *Quelles sont les capacités de localisation et de saisie d'objets sonores dans l'espace péri-personnel ?*

  ⇒ *Est-il possible d'améliorer les indices de perception de la distance en utilisant la sonification ?*

- Sonification personnalisable par l'utilisateur

  ⇒ *Comment mettre en place des méthodes de sonification permettant de satisfaire les critères esthétiques de tous les utilisateurs ?*

1.3 Organisation du manuscrit

Ce manuscrit est structuré autour des trois axes présentés précédemment. Ces axes ont été traités de façon quasiment autonome et peuvent être abordés par le lecteur indépendamment les uns des autres. Ils présentent des travaux réalisés dans le cadre du projet mais qui peuvent être utilisés dans de nombreuses autres applications.

Le chapitre 2 présente une revue de la littérature des principales thématiques abordées dans cette thèse : le son 3D, la sonification et les systèmes d'aide à la navigation pour les non-voyants. La partie sur le son 3D présente les différents indices acoustiques permettant la localisation auditive, le principe de la synthèse binaurale ainsi que les performances de localisation auditive avec des sources sonores réelles et virtuelles. La revue de littérature sur la sonification introduit ce concept de transmission d'information apparu avec les nouvelles technologies audio, donne ses grandes fonctions ainsi que les techniques de sonifications les plus communes. Enfin, la dernière partie de ce chapitre détaille les différents types de dispositifs d'aide à la navigation pour les non-voyants.

Le chapitre 3 aborde la question de l'amélioration du rendu binaural avec des HRTF non-individualisées. Après avoir décrit les dégradations engendrées par l'utilisation d'HRTF non-individuelles et les méthodes existantes pour adapter les HRTF au sujet, nous introduisons une nouvelle manière de considérer le problème de l'individualisation. S'inspirant des expériences sur la plasticité cérébrale, nous explorons dans ce chapitre les possibilités d'adapter le système auditif des sujets à des HRTF non-individuelles. Cette méthode, basée sur un jeu audio-kinesthésique, est décrite puis évaluée sur plusieurs jours en alternant les sessions d'adaptation et des tests de localisation sonore. Les travaux présentés dans ce chapitre ont été publiés dans [Katz et Parseihian, 2012] et [Parseihian et Katz, 2012b].

L'objectif du chapitre 4 est d'explorer les possibilités d'amélioration des indices de localisation par l'utilisation de la sonification. La première partie de ce chapitre étudie les performances de localisation du système auditif en champ proche pour des sons réels, en fonction de la main utilisée pour le pointage. À partir de ces résultats, une méthode de sonification est mise en place pour améliorer les indices de perception de la distance. Cette méthode est basée sur un mapping de paramètres d'effets audio. Trois métaphores de sonification de la distance ont été réalisées selon

cette méthode. Ces métaphores sont évaluées avec un test de localisation et comparées aux performances de localisation de sources virtuelles. La méthode de sonification de la distance décrite dans la deuxième partie de ce chapitre a été présentée dans [Parseihian *et al.*, 2012].

La mise en place de méthodes de sonification permettant de satisfaire les critères esthétiques de tous les utilisateurs est traitée dans les chapitres 4 et 5. La méthode mise en place dans le chapitre 4 consiste à faire varier les paramètres acoustiques du son en fonction de la distance en utilisant des effets audio. Cette méthode permet d'appliquer la sonification à tout type de sons, tout en gardant le même mapping entre la distance et les indices acoustiques ajoutés. Le chapitre 5 introduit un nouveau type de signaux auditifs permettant de créer un vocabulaire sonore indépendant du type de son utilisé lorsqu'il est nécessaire de transmettre plusieurs informations sous la forme de messages sonores. Ces signaux, appelés *morphocons* sont des petites entités sonores dont la construction est faite sur la base de descriptions de l'évolution temporelle du son. Un vocabulaire de *morphocons* est mis en place dans le cadre du projet NAVIG pour donner des informations à l'utilisateur sur la présence de points de repère, de points d'intérêt ou de points sur la trajectoire à suivre pendant la navigation en extérieur. Trois palettes de *morphocons* ont été créées à partir de ce vocabulaire et validées avec un test de catégorisation. La présentation du concept des *morphocons* et son application au projet NAVIG pour la mise en place d'un vocabulaire sonore a été publiée dans [Parseihian et Katz, 2012a].

L'annexe A présente les travaux relatifs à l'analyse des besoins utilisateurs en terme d'informations à donner et sur la manière de les transmettre. L'annexe B quant à elle, présente l'architecture du système NAVIG et décrit brièvement les principaux éléments du système. Pour plus d'informations sur le système NAVIG, le lecteur pourra se réferer à [Katz *et al.*, 2012a, Katz *et al.*, 2012b, Kammoun *et al.*, 2012].

Chapitre 2

Sonification 3D et aide aux non-voyants : état de l'art

Sommaire

2.1 Introduction

La conception d'un moteur de sonification binaurale pour l'aide à la navigation [1] des non-voyants fait intervenir différentes disciplines telles que l'informatique, la perception sonore, la cognition spatiale, l'ergonomie, etc. Dans ce chapitre nous allons présenter un état de l'art des trois principaux domaines qui ont été abordés dans le cadre de cette thèse. Tout d'abord, la section 2.2 introduit la notion de son 3D, son importance dans la perception spatiale, ainsi que le principe de la synthèse binaurale qui a été utilisé pour ce projet. La section 2.3 définit le concept de la sonification et

1. La navigation correspond à l'ensemble des techniques qui permettent : de connaître la position d'un élément mobile par rapport à un système de coordonnées, de calculer le chemin à suivre pour joindre un autre point de coordonnées connues et de calculer les informations relatives au déplacement de ce mobile (distances et durées, vitesse de déplacement, heure estimée d'arivée, etc.). La navigation peut être terrestre, aérienne ou maritime. Il n'est pas question dans ce manuscrit d'aider les non-voyant à faire du bateau.

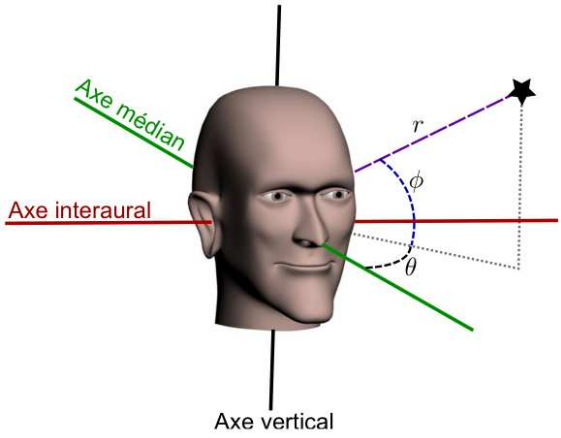


FIGURE 2.1 – Repère sphérique associé à la tête du sujet (figure extraite de [Rébillat, 2011]). La position d'une source sonore par rapport à la tête est définie par l'angle θ correspondant à l'azimut, l'angle ϕ correspondant à l'élévation et par la distance r. Le plan horizontal désigne le plan comprenant l'axe interaural et l'axe médian. Le plan médian désigne le plan comprenant l'axe vertical et l'axe médian.

présente différentes méthodes permettant de transmettre des informations en utilisant la modalité auditive. Enfin, la section 2.4 dresse un inventaire des principaux projets de recherches réalisés pour l'aide à la navigation et au déplacement des non-voyants.

2.2 Son 3D

Le terme "son 3D" est apparu ces dernières années avec l'émergence de nombreuses techniques de synthèse et de reproduction de la dimension spatiale du son. Il fait référence à la capacité de recréer des espaces auditifs virtuels où le son ne provient pas d'une position spécifique de l'espace (un haut-parleur, par exemple), mais de différentes directions possibles grâce à l'utilisation de plusieurs haut-parleurs ou d'un casque stéréophonique. Plusieurs procédés permettent de créer des sons 3D. Parmi les plus aboutis, nous pouvons citer : l'ambisonique ([Gerzon, 1985]), la wave field synthesis ([Berkhout *et al.*, 1993]) et la synthèse binaurale ([Begault, 1994]). La mise en place d'un dispositif de spatialisation du son nécessite une bonne connaissance des mécanismes de perception spatiale du système auditif ainsi que des performances de l'homme à pouvoir localiser des sons dans l'espace. Dans cette section, nous allons commencer par rappeler les indices permettant la localisation auditive. Ce rappel nous permettra dans un deuxième temps de définir le procédé de création de son 3D utilisé pour cette thèse : la synthèse binaurale. Enfin, nous verrons quelles sont les manières de quantifier les performances de localisation auditive et nous rappellerons les résultats généraux de la littérature sur les performances de localisation de sons réels et de sons virtuels (générés à partir de la synthèse binaurale).

Pour cette section, les descriptions spatiales seront effectuées dans le repère de coordonnées sphériques (r, θ, ϕ) ayant pour origine l'intersection de l'axe interaural et de l'axe médian (cf. figure 2.1). Ce repère est considéré comme le plus naturel pour la description de l'espace qui nous

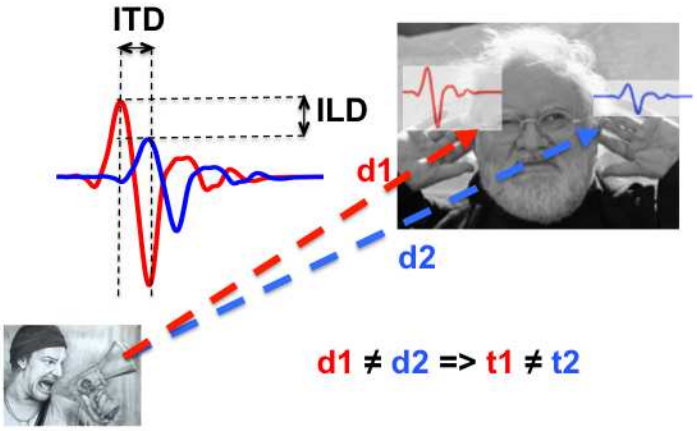


FIGURE 2.2 – Lorsque la source sonore est située en dehors du plan médian de l'auditeur (ici Pierre Henry), l'onde sonore parvient plus tôt et avec une intensité plus forte à l'oreille la plus proche de la source. Ce phénomène est représenté par la différence interaurale de temps et par la différence interaurale d'intensité.

entoure étant donné qu'il est égocentré.

2.2.1 La localisation auditive

Cette partie est consacrée à la présentation des fondements de la localisation auditive, dont la connaissance est essentielle pour étudier les mécanismes de la perception spatiale ainsi que pour mettre en place des systèmes de spatialisation sonore. Contrairement au système visuel qui n'apporte des informations que sur l'hémisphère frontal de notre espace perceptif, le système auditif est capable, grâce à nos deux oreilles, d'analyser une scène sonore selon toutes les directions de l'espace. Nous allons ici rappeler les principaux indices acoustiques de localisation permettant d'expliquer l'extraction d'informations spatiales par le système auditif.

a) Indices interauraux

La première description des mécanismes de perception en azimut a été réalisée par [Rayleigh, 1907] dans la *Duplex Theory*. Cette théorie se base sur la différence de position entre nos deux oreilles pour expliquer la localisation dans le plan horizontal. Ainsi la captation par les deux oreilles d'une même onde sonore donne lieu à des indices dits "interauraux" (figure 2.2) : la différence interaurale de temps (ou ITD, pour *Interaural Time Difference*) et la différence interaurale d'intensité (ou ILD, pour *Interaural Level Difference*) [Blauert, 1997].

i/ La différence interaurale de temps (ITD)

Lorsque la source à localiser est située en dehors du plan médian, il existe une différence de trajet acoustique entre la source et chaque oreille. Il en résulte une différence de temps d'arrivée entre

les deux trajets, résumée par le concept d'ITD. Cet indice recouvre deux mécanismes du système auditif, agissant sur deux plages fréquentielles distinctes :

1. *La différence de phase* entre les signaux sonores arrivant sur chacune des oreilles est perceptible lorsque les longueurs d'onde associées au signal sonore sont inférieures au diamètre de la tête. L'oreille est donc sensible aux différences de phases pour les fréquences inférieures à 1500 Hz. Au-delà, cet indice devient ambigu (la différence de phase peut alors comprendre plusieurs périodes du signal) et les différences de phases ne sont plus informatives. Pour un modèle de tête sphérique (de rayon r), [Kuhn, 1977] a montré que l'ITD dû à la différence de phase peut s'exprimer en fonction de θ l'angle d'incidence de l'onde plane, par :

$$ITD_{BF} = 3\frac{r}{c}\sin\theta$$

2. *La différence de moment d'arrivée de l'enveloppe* entre les signaux captés par les oreilles droite et gauche constitue l'indice temporel le plus plausible à partir de 1500 Hz. L'ITD est alors bien approché par une modélisation sphérique de la tête, en ne tenant compte que de la différence de marche entre les oreilles de l'auditeur (modèle de [Woodworth et Schlosberg, 1954]) :

$$ITD_{HF} = \frac{r}{c}(\sin\theta + \theta)$$

ii/ La différence interaurale d'intensité (ILD)

La tête de l'auditeur agissant comme un obstacle face à l'onde acoustique incidente, la différence d'intensité entre les signaux captés à chaque oreille est dépendante de la position de la source. Cet indice est valable sur toute la gamme de fréquences audibles, néanmoins, aux basses fréquences ($\lambda < r$), l'ILD est très faible car la tête ne diffracte pas l'onde incidente. L'efficacité de cet indice intervient donc à partir de 1500 Hz, lorsque les longueurs d'ondes sont inférieures à la taille de la tête. La surface de la tête réfléchit alors parfaitement l'onde incidente. Il en résulte une différence de niveau d'environ 20 dB pour une source située sur l'axe interaural.

b) Indices spectraux

Les indices interauraux ne dépendent que de l'azimut de la source. La figure 2.3 met en évidence une zone, appelée cône de confusion, où l'ITD ou l'ILD sont constant quelque soit la position de la source sur ce cône. Ainsi, les indices interauraux, ne parviennent pas à expliquer la perception de l'élévation, ni la discrimination entre les sources situées à l'avant et à l'arrière. Cette lacune des indices interauraux a mis en évidence l'existence d'autre indices, dits indices spectraux monauraux.

Les indices spectraux sont caractérisés par les filtrages dus aux pavillons des oreilles ainsi que par les réflexions des ondes incidentes sur les épaules et le torse de l'auditeur. Ces filtrages et ces réflexions sont dépendants de la direction d'incidence des ondes sonores et permettent ainsi l'estimation de l'élévation de la source sonore [Batteau, 1967, Batteau, 1968]. Ces indices sont monauraux au sens où ils apparaissent indépendamment aux tympans de chaque oreille.

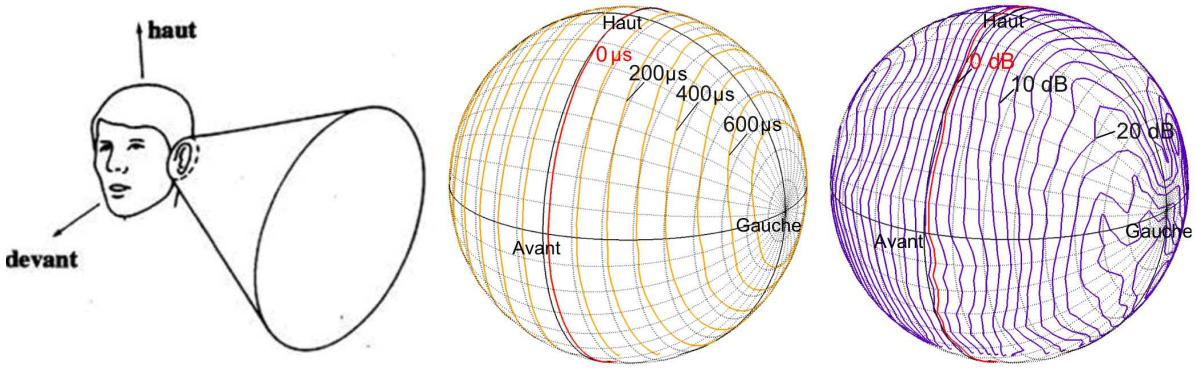


FIGURE 2.3 – Gauche : Illustration de la notion de cône de confusion. L'hyperboloïde correspond aux positions de sources qui génèrent une ITD et une ILD constante pour un modèle de tête sphérique (figure extraite de [Chateau, 1996]). Centre : Évolution de l'ITD aux basses fréquences sur la sphère. Les lignes relient les directions correspondant à une même valeur d'ITD. La ligne rouge représente l'iso-ITD $0\mu s$, les lignes sont espacées de $100\mu s$ (figure extraite de [Guillon, 2009]). Droite : Évolution de l'ILD sur la sphère. Les lignes relient les directions correspondant à une même valeur d'ILD, et sont espacées de 2 dB. La ligne rouge représente l'iso-ILD 0 dB (figure extraite de [Guillon, 2009]).

c) Fonction de transfert acoustique

Notre morphologie est donc à l'origine d'un encodage spatial binaural duquel dépend notre aptitude à localiser des sons dans l'espace. Les indices, interauraux et monauraux, liés au diamètre de la tête ainsi qu'à l'ensemble torse - tête - oreille externe sont à l'origine de ce qu'on appelle les HRTF, pour *Head Related Transfert Function* (fonction de transfert relative à la tête) [Batteau, 1968, Begault, 1994, Blauert, 1997].

D'un point de vue signal, les phénomènes acoustiques observés entre la source et l'entrée des conduits auditifs d'un auditeur peuvent être modélisés comme deux systèmes linéaires invariants, caractérisables par leurs réponses impulsionnelles $h_L(t)$ et $h_R(t)$, tels que :

$$x_{L,R}(t) = h_{L,R} * x(t)$$

où $x(t)$ correspond au signal acoustique émis par la source et $x_L(t)$ et $x_R(t)$ sont les signaux reçus aux oreilles gauche et droite.
Sous sa forme fréquentielle, le problème est défini par les relations :

$$X_{L,R}(j\omega) = H_{L,R}(j\omega) \cdot X(j\omega)$$

où H_L et H_R sont les fonctions de transfert traduisant les phénomènes acoustiques subis par le signal $x(t)$ entre la source et l'entrée des deux oreilles. Les signaux $x_L(t)$ et $x_R(t)$ étant dépendant de la direction, il existe un couple de fonction de transfert pour chaque direction (θ, ϕ). L'ensemble de ces fonctions de transfert pour toutes les directions est appelé "jeu d'HRTF". Leurs versions temporelles h_L et h_R sont appelées *Head Related Impulses Responses* ou HRIR (pour réponses

impulsionnelles liées à la tête).

Les HRTF offrent une approche globale de la perception auditive spatiale, étant donné qu'elles contiennent toutes les informations acoustiques dont le système auditif a besoin pour localiser une source fixe dans une position donnée de l'espace. Néanmoins, à cause des variations morphologiques observables entre les individus, les HRTF peuvent être très différentes d'un individu à un autre.

d) Indices de perception de la distance

La variation de la distance d'une source sonore affecte de multiples façons les propriétés acoustiques du son atteignant les oreilles d'un auditeur [Zahorik *et al.*, 2005]. Il existe donc plusieurs indices permettant de percevoir la distance r d'une source sonore, l'influence de ces indices peut varier en fonction du milieu d'écoute (intérieur ou extérieur), de la proximité de la source (proche ou lointaine) et de la familiarité de l'auditeur avec celle-ci :

1. **L'intensité** : Lorsque la distance augmente, le niveau sonore de la source acoustique décroît. La nature précise des variations d'intensité dépend des conditions environnementales ainsi que des propriétés acoustiques de la source. Pour une source en champ libre, la perte d'intensité est inversement proportionnelle au carré de la distance. Le niveau de perte en décibel lorsque la source passe d'une distance r_1 à une distance r_2 peut s'exprimer ainsi [Coleman, 1963] :

   $$\text{“perte en dB”} = 20\log_{10}(\frac{r_2}{r_1})$$

   Par conséquent, lorsque la distance source/observateur est doublée, le niveau d'intensité de la source subit une atténuation de 6 dB.
   Cette loi n'est applicable que pour des distances supérieures à 1 m. Pour les sources situées en champ proche, la présence de la tête de l'auditeur influe sur le niveau d'intensité arrivant aux deux oreilles et rend difficile la perception de la distance sur la seule base de l'intensité.
   L'intensité n'apporte, de plus, qu'une information relative sur la distance. Si la source est fixe, cet indice peut être confondu au niveau d'intensité sonore de la source. Ainsi, il est difficile d'estimer la distance égocentrique d'une source peu familière en champ libre [Mershon et King, 1975].

2. **Le rapport “champ direct” sur “champ réverbéré”** : dans un environnement réverbérant, la décroissance de l'intensité avec la distance est plus faible qu'en champ libre, mais un autre indice intervient : le rapport d'énergie entre le champ direct et le champ réverbéré. À proximité de la source, le champ direct est prépondérant dans le signal perçu. Au fur et à mesure que la distance augmente, l'intensité du son direct diminue et la part relative au champ réverbéré augmente. La pertinence perceptive de cet indice a été démontrée par [von Békésy, 1960] en mixant des signaux sonores enregistrés en chambre anéchoïque et en chambre réverbérante. Une étude de [Mershon et King, 1975] réalisée sur 160 sujets a permis de montrer que le jugement de la distance est plus précis dans un environnement réverbéré

que dans un environnement anéchoïque. Cette étude montre aussi que contrairement à l'intensité, le rapport champ direct/champ réverbéré est un indice de jugement absolu et que peu d'écoutes suffisent pour évaluer, au moins grossièrement, la distance. La perception de la distance dépend cependant du niveau de réverbération de la salle [Nielsen, 1992], mais [Shinn-Cunningham, 2000] a montré que les auditeurs sont capables d'adapter leur perception en fonction de la salle d'écoute.

3. **Contenu spectral** : Pour des distances supérieures à 15 m, les propriétés d'absorption de l'air modifient considérablement le spectre de la source sonore. Cette absorption étant dépendante de la longueur d'onde, les hautes fréquences sont plus vite atténuées que les basses fréquences. À ces modifications, dues à l'absorption, peuvent s'ajouter des filtrages entraînés par les éventuelles réflexions sur des surfaces non idéales [Blauert, 1997]. Comme pour l'intensité, le spectre de la source doit être connu pour que cet indice soit informatif. Il s'agit donc d'un indice de localisation relatif.
4. **Indices binauraux** : Pour des sources en champ proche (<1.5 m), il a été montré dans plusieurs études que les indices binauraux (ITD et ILD vus précédemment) permettent également de percevoir la distance [Brungart *et al.*, 1999, Coleman, 1968]. En effet, lorsque la source est proche et en dehors de l'axe médian les indices binauraux varient significativement en fonction de la distance et constituent un indice de perception absolue de la distance d'une source sonore. Une étude de [Shinn-Cunningham *et al.*, 2000] fournie une analyse détaillée de la variation de ces indices binauraux en fonction de la position des sources et met en évidence la dépendance de ces indices en fonction de la distance.
5. **Indices dynamiques** : Dans la vie quotidienne, les auditeurs et les sources sonores sont rarement stationnaires. Les mouvements de translations et de rotations de l'auditeur (ou de la source) entraînent des variations des indices acoustiques vus précédemment, augmentant ainsi la quantité d'information disponible.

   La parallaxe de mouvement, qui induit un changement de direction de la source sonore, entraîne un déplacement relatif plus important pour les sources proches que pour les sources lointaines ; c'est un indice permettant d'estimer la distance absolue d'une source sonore. L'apport de cet indice a notamment été étudié par [Kim *et al.*, 2001].

   Enfin, l'effet *Doppler*, qui introduit une variation continue du spectre de la source sonore lors des déplacements rapides de celle-ci peut aussi permettre un jugement relatif de la distance. Cependant cet effet n'a qu'une faible influence sur la perception de la distance [Rosenblum *et al.*, 1987].

e) Autre indices

De nombreux indices, non acoustiques, peuvent aussi influer sur la localisation des sources sonores. La vue est notamment un indice dont l'influence est prépondérante sur la modalité auditive. Le phénomène le plus flagrant qui permet ce constat est "l'effet ventriloque" [Recanzone, 1998]. Il consiste à entendre un son comme provenant de sa source sonore présumée sur le plan visuel, même

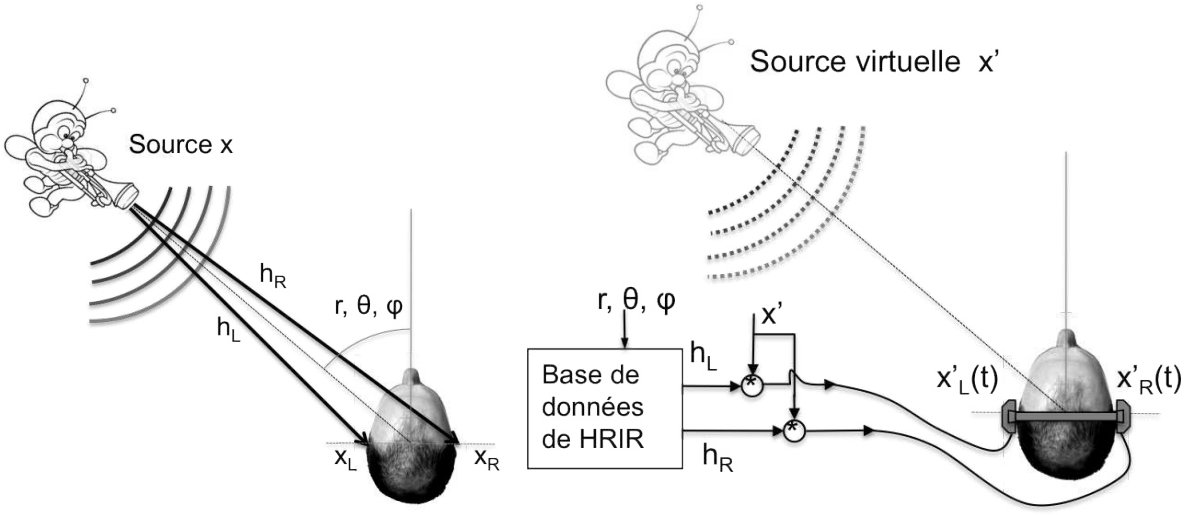


FIGURE 2.4 – Principe de la synthèse binaurale. Captation du signal à gauche, restitution à droite.

si celle-ci n'en est pas la source émettrice. Pour en faire l'expérience, il suffit d'aller au cinéma où le son provient généralement de haut-parleurs situés sur les bords ou derrière l'écran à des positions fixes, alors que la voix des acteurs nous semble toujours sortir de leur bouche [Chion, 1994].

La localisation est aussi influencée par les mouvements de l'auditeur. Ces mouvements sont perçus par les sens vestibulaires (liés à l'équilibre) et proprioceptifs (liés aux mouvements du corps et de ses différentes parties) [Majdak *et al.*, 2008].

Enfin, la familiarité avec la source est aussi un facteur influant sur l'audition spatiale de la même façon que les présupposés que l'auditeur peut avoir sur cette source. Ainsi nous sommes habitués à localiser les sons d'oiseaux ou d'avion en l'air alors que nous nous attendons à percevoir les voix autour du plan horizontal. La modification de ces attentes peut entraîner des erreurs de localisation, il est donc nécessaire pour les tests de localisation d'utiliser des sons "neutres" (que les sujets ne connaissent pas).

2.2.2 La synthèse binaurale

La synthèse binaurale est une technique de spatialisation consistant à reproduire, à l'aide d'un casque stéréophonique, le champ acoustique correspondant à une source provenant d'une position (θ, ϕ) à l'entrée des deux oreilles. Cette technique repose sur l'utilisation des HRTF. Elle permet de reproduire au mieux les indices de localisation nécessaires pour procurer à l'auditeur l'illusion que le son provient de la position (θ, ϕ). Ce principe est illustré sur la figure 2.4.
Dans cette section, nous allons aborder le principe, ainsi que quelques aspects techniques, de la synthèse binaurale.

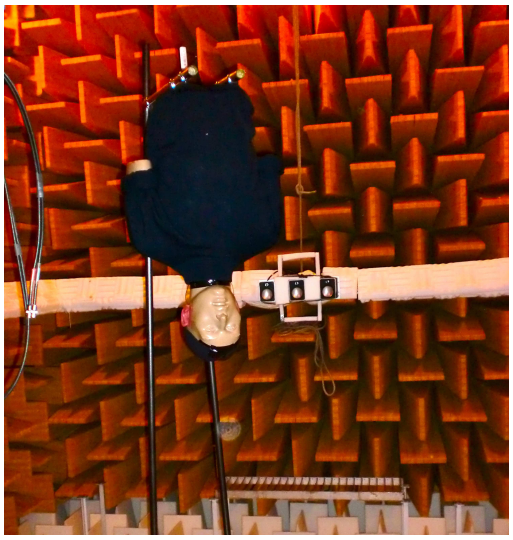


FIGURE 2.5 – Photo du système de mesure d'HRTF de la chambre anéchoïque de l'IRCAM avec le mannequin KEMAR (positionné à l'envers) attendant patiemment que l'on mesure les HRTF de son hémisphère bas.

a) Principe

Nous avons vu dans la section 2.2.1.c) que pour une position de source (θ, ϕ) donnée, il est possible de déterminer une fonction de transfert relative à la tête (HRTF) du trajet acoustique de la source vers chaque oreille. Le principe de la synthèse binaurale est d'utiliser directement ce couple d'HRTF comme filtres appliqués à un signal monophonique et de délivrer sans diaphonie les deux signaux résultants à chaque oreille à l'aide d'un casque stéréophonique [Begault, 1994, Blauert, 1997, Nicol, 2010]. Le cerveau décode naturellement les indices acoustiques contenus dans les filtres. Il externalise alors le son pour le replacer dans l'espace à la position où la source aurait été située pour fournir de tels indices temporels et spectraux à l'entrée des conduits auditifs de l'auditeur. On parle de *source sonore virtuelle*.

b) Mesure, égalisation et interpolation des HRTF

Afin de pouvoir effectuer l'opération de filtrage de la synthèse binaurale, il est nécessaire de disposer d'une base de donnée d'HRTF. Une base ou un jeu d'HRTF est constituée d'un certain nombre de paires d'HRTF (pour les oreilles gauche et droite) correspondant chacune à des positions (θ, ϕ). Ce jeu d'HRTF peut être obtenu soit par un modèle, soit par la mesure. Dans les deux cas, il faut avoir recours à un échantillonnage discret de l'espace. L'obtention d'un jeu d'HRTF par la mesure s'effectue en réalisant une série de mesures de réponses impulsionnelles pour un ensemble de directions de l'espace avec des microphones omnidirectionnels miniatures placés à l'entrée des conduits auditifs du sujet après les avoir bouchés [Wightman et Kistler, 1989a, Djelani *et al.*, 2000, Pernaux, 2003]. La mesure des HRTF nécessite une chambre anéchoïque, avec une installation mécanique lourde et complexe, dont la

précision est fondamentale pour obtenir de bons résultats (comme le montrent les travaux de [Bronkhorst, 1995]). Plusieurs bases de données d'HRTF mesurées sur des humains ou sur des têtes artificielles existent. Parmi les plus connues, nous pouvons citer la base CIPIC [Algazi *et al.*, 2001b], la base [LISTEN, 2003] de l'IRCAM (dont le système est représenté figure 2.5) ou la base [Tohoku, 2001].

Afin de compenser les réponses en fréquence des différents éléments de la chaîne de mesure (haut-parleur, microphones), il est nécessaire de procéder à une égalisation des HRTF à partir d'une mesure de référence du système d'acquisition. Plusieurs méthodes, telles que l'égalisation en champ libre ou l'égalisation en champ diffus sont exposées dans la thèse de [Larcher, 2001]. Une autre méthode consiste à compenser chaque mesure indépendamment en réalisant une mesure des fonctions de transfert haut-parleur/microphones de toutes les positions avec les microphones placés aux mêmes positions que pour la mesure du sujet [Dobrucki *et al.*, 2010].

La synthèse de sons spatialisés dans toutes les directions de l'espace nécessite une grille de mesure très fine et donc un temps d'acquisition du jeu d'HRTF très long. Les mesures sont généralement espacées de 5 à 15° en azimut et en élévation. Pour placer des sources virtuelles entre ces points, les mesures doivent être interpolées. Un grand nombre de manières d'effectuer l'interpolation spatiale des filtres HRTF existe. Pour plus d'information, le lecteur pourra se référer à [Larcher et Jot, 1997, Carlile *et al.*, 2000].

c) Implémentation

Il existe plusieurs méthodes d'implémentation de la synthèse binaurale. Les plus fréquentes, sont : la convolution directe et l'utilisation de filtres à phase minimale et de retards purs. Quelque soit la technique d'implémentation choisie, cette opération requiert des ressources non négligeables, d'autant plus qu'il est en général nécessaire de pouvoir la réaliser en temps réel afin de pouvoir effectuer des manipulations dynamiques des sources sonores virtuelles. Le moteur de synthèse binaurale doit donc être capable de réaliser ce filtrage pour des positions de sources (fixes ou mouvantes, équipées d'un capteur de position) en fonction de la position de la tête (équipée d'un capteur de position et d'orientation). Afin d'obtenir un rendu correct, il est nécessaire que la latence du système soit minimale. Cette latence est définie par le temps qui s'écoule entre l'instant où l'auditeur effectue un mouvement de la tête et celui où les filtres correspondants aux nouvelles positions sont mis à jour. [Sandvad, 1996] a montré que le premier facteur de dégradation de la perception d'une source virtuelle binaurale correspond à la latence. Une valeur de latence acceptable pour la synthèse de tout type de son semble être de 75 ms selon [Brungart *et al.*, 2004] (d'autre études, telles que [Wenzel, 1999] ont montré que pour des sons relativement long, une latence de 250 ms peut être acceptable). Afin d'obtenir des mouvements lisses et réalistes, il est aussi nécessaire de recalculer la position de la source en fonction de la position de la tête avec une fréquence minimum de 50 Hz.

Comme nous l'avons mentionné section 2.2.1.c), les HRTF dépendent de la morphologie des

sujets. Elles sont donc individuelles. Pour obtenir des sources virtuelles les plus réalistes possibles, il est nécessaire de disposer des HRTF individuelles de l'auditeur. Les systèmes de mesures étant peu nombreux et l'acquisition d'un jeu d'HRTF étant relativement longue, il est compliqué d'effectuer la synthèse binaurale avec des HRTF individuelles dans le cadre d'une utilisation commerciale. Il en résulte des dégradations dans la perception des sources virtuelles qui seront détaillées dans la section 2.2.3. Plusieurs méthodes d'individualisation des HRTF existent, celles-ci seront détaillées dans le chapitre 3 avec la méthode proposée dans cette thèse.

2.2.3 Performances de localisation auditive

L'estimation des performances de localisation du système auditif a fait l'objet de nombreuses études perceptives, tant au niveau de la localisation de sons réels (avec des haut-parleurs) en champ libre (ou environnement anéchoïque) ou en environnement clos (avec le champ réverbéré), qu'au niveau de la localisation avec des sons virtuels (issus de la synthèse binaurale). Après avoir évoqué quelques considérations sur les tests perceptifs permettant de quantifier les performances de localisation, nous détaillerons, dans cette section, les performances moyennes du système auditif en écoute naturelle, puis nous les comparerons aux résultats obtenus en écoute avec de la synthèse binaurale. Nous nous restreindrons pour cet exposé à la localisation de sources sonores en champ libre et au cas où une seule source sonore est présente.

a) Tests perceptifs

Afin d'estimer les limites du système auditif, il est nécessaire de mesurer les performances de localisation pour un certain nombre de sujet. Deux catégories de tests perceptifs sont utilisées en localisation sonore : les tests de localisation relative et les tests de localisation absolue.
La localisation relative consiste à évaluer l'angle minimum audible entre deux sources sonores identiques situées à la même distance de l'auditeur [Mills, 1958, Hartmann, 1989]. Ce protocole, basé sur une tâche de discrimination, a l'avantage de réduire la composante motrice de la réponse (celle-ci pouvant introduire un biais dans les expériences de localisation absolue). Cependant, les indices utilisés par le sujet pour discriminer les deux sources ne sont pas forcement des indices de localisation et cette tâche peut être réalisée sans que le sujet ne localise correctement les deux sources [Makous, 1990].
Les tests de localisation absolue visent à évaluer la capacité du sujet à désigner la position d'une source sonore dans l'espace. Ils consistent à faire écouter un certain nombre de stimuli spatialisés pour différentes directions de l'espace et à demander au sujet de reporter à chaque fois la direction perçue du stimulus. L'erreur angulaire moyenne de localisation commise par les sujets testés est ensuite estimée à partir de tests statistiques. Selon l'objet de l'étude, il est aussi nécessaire de s'intéresser au temps de réponse des sujets, ou à des critères de facilité de localisation, de crédibilité ou de bonne externalisation pour les sources virtuelles.

Pour les travaux de cette thèse, nous nous sommes intéressé à la localisation auditive absolue. Nous nous focaliserons donc sur les résultats de ce type de test dans la suite du document.

Plusieurs techniques de report du jugement de la position perçue ont été utilisées dans les différentes études de la littérature sur les performances de localisation auditive absolue. La technique de report doit permettre au sujet d'exprimer le plus fidèlement possible la position perçue tout en étant intuitive et rapide. Du simple report verbal (en degrés ou en heure), utilisé par [Wightman et Kistler, 1989a, Wenzel *et al.*, 1993], aux techniques de pointages égocentrées utilisant la tête, le torse ou le bras [Makous, 1990, Brungart *et al.*, 1999], en passant par l'utilisation d'une interface physique [Djelani *et al.*, 2000] ou graphique [Larcher, 2001, Pernaux, 2003], ces techniques ont toutes des avantages et des inconvénients. Plusieurs études ont cependant mis en évidence de meilleures performances pour les techniques de reports égocentrés faisant intervenir une partie du corps (pointage avec la main ou la tête) [Haber *et al.*, 1993, Pernaux, 2003, Majdak *et al.*, 2010]. Afin de réduire l'incertitude de la mesure liée à la méthode de recueil des réponses et à l'adaptation au protocole expérimental, plusieurs auteurs ont fait répéter la tâche de localisation pendant plusieurs heures aux auditeurs avant que les réponses ne soient enregistrées [Makous, 1990, Wightman et Kistler, 1989b], parfois avec un retour sur la qualité de la réponse [Carlile *et al.*, 1997, Martin *et al.*, 2001, Brungart et Simpson, 2009].

b) Localisation de sons réels

La majeure partie des expériences décrites dans cette partie ont été réalisées dans des conditions "idéales" de laboratoire (avec la tête fixe, dans le silence, en condition anéchoïque, avec une seule source statique généralement et avec un spectre large bande).

i/ Performances de localisation en azimut

Dans son ouvrage de référence sur la perception spatiale du son, [Blauert, 1997] fait un rapport exhaustif des études sur la localisation auditive. Il y introduit la notion de flou de localisation (ou *localization blur*, en anglais), comme étant l'erreur de localisation perçue dans une zone de l'espace. La figure 2.6 reporte les flous de localisation dans le plan horizontal calculés à partir des résultats des expériences de [Preibisch-Effenberger, 1966] (réalisée sur 600 sujets) et de [Haustein et Schirmer, 1970] (réalisée sur 900 sujets). Sur cette figure sont représentées les positions moyennes et les écarts types des réponses des sujets à des stimuli auditifs (bruit blanc de 100 ms) provenant de quatre directions différentes (0°, 90°, 180° et 270°). On constate que la précision de localisation est maximale dans la direction frontale (azimut 0°) où le flou de localisation est de ±4°, plus faible à l'arrière (flou de localisation de ±6°) et minimale pour les positions latérales (flou de localisation de ±10°). [Blauert, 1997] montre que suivant le stimulus utilisé le flou de localisation pour une source frontale peut varier de 0.75° (pour des impulsions) à 12° (pour des fréquences pures).

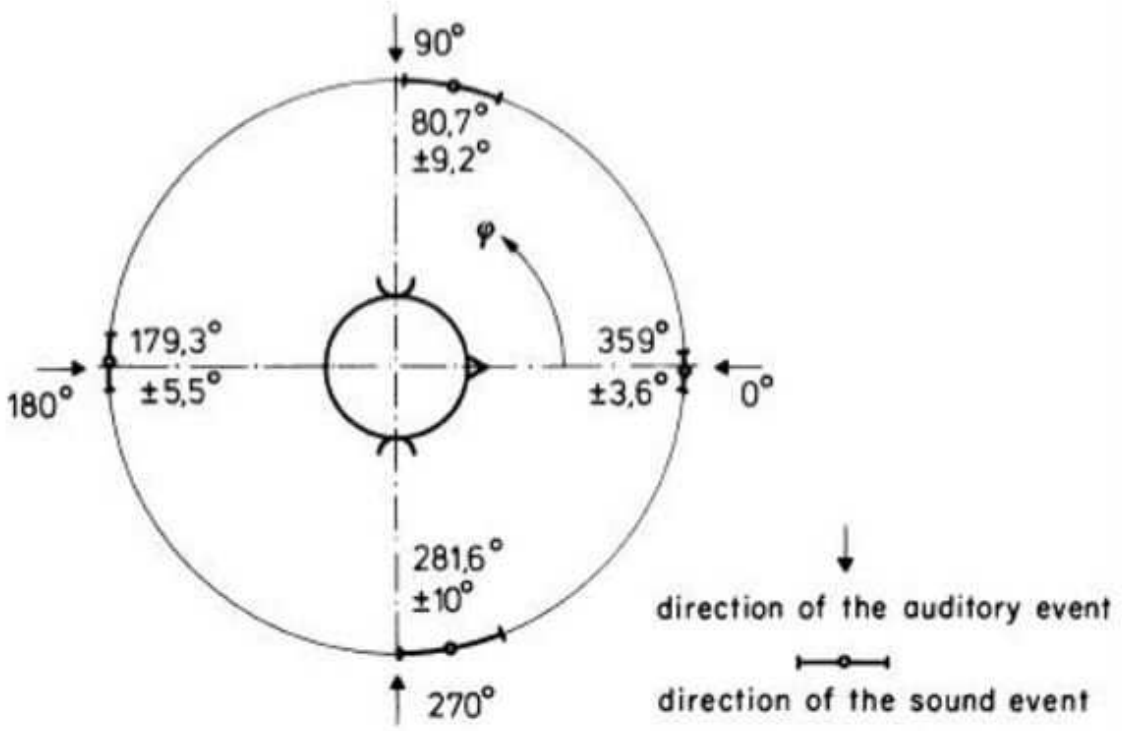


FIGURE 2.6 – Récapitulatif du flou de localisation auditif dans le plan horizontal. Image tirée de [Blauert, 1997] d'après des études de [Preibisch-Effenberger, 1966] et [Haustein et Schirmer, 1970] - sur 600 et 900 sujets, bruit blanc de 100 ms. Les flèches indiquent la provenance du son, les cercles représentent les positions moyennes des réponses des sujets et les portions d'arc de cercles les écarts types.

ii/ Performances de localisation en élévation

Au niveau de l'élévation, les performances de localisation du système auditif sont plus floues que pour la localisation en azimut. La figure 2.7, tirée de [Blauert, 1997], présente les résultats d'une campagne de tests de localisation de [Damaske et Wagener, 1969] réalisée sur sept sujets pour des sources sur le plan médian avec un signal de parole. De nouveau, l'erreur est minimale pour les cibles situées devant et à faible élévation (flou de localisation de ±9°), elle augmente en fonction de l'élévation (±10° à 36°, ±13° à 90°) et est maximale dans l'hémisphère arrière (±15° à 36° pour un azimut de 180°).

Une expérience réalisée par [Oldfield et Parker, 1984] sur huit sujets avec un bruit blanc pour des élévations allant de -40 à 40° a permis de dresser une cartographie détaillée de l'acuité de localisation dans une grande partie de la sphère auditive. Leurs résultats vont dans le même sens que les études résumées par [Blauert, 1997] mais ajoutent plus de détails sur le flou de localisation en azimut en dehors du plan horizontal ainsi que sur les erreurs en élévation en dehors du plan médian.

iii/ Confusions avant/arrière

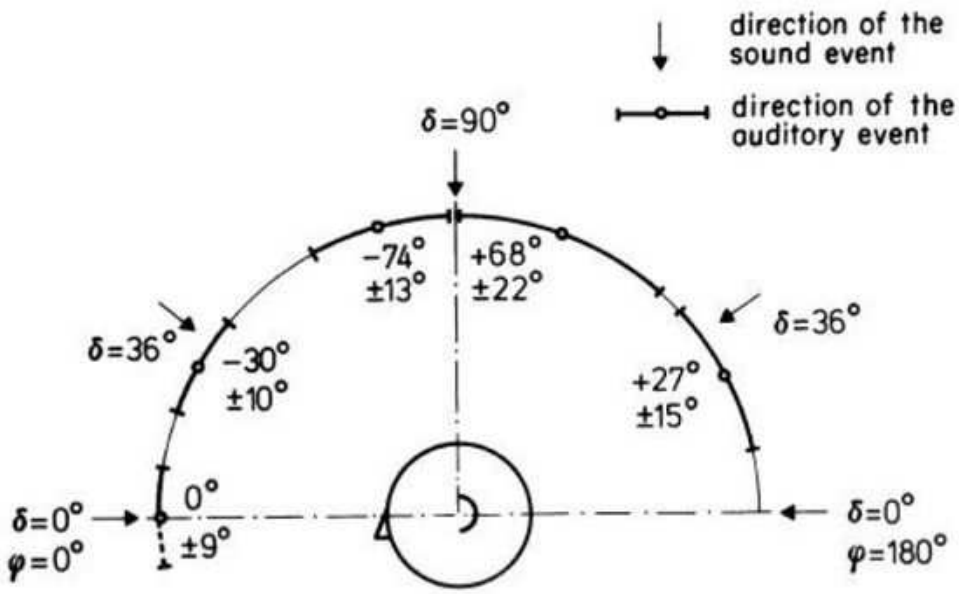


FIGURE 2.7 – Récapitulatif du flou de localisation auditif dans le plan médian. Image tirée de [Blauert, 1997] d'après une étude de [Damaske et Wagener, 1969] réalisée sur 7 sujets avec un signal de parole. Les flèches indiquent la provenance du son, les cercles représentent les positions moyennes des réponses des sujets et les portions d'arc de cercles les écarts types.

Nous avons vu dans la section 2.2.1.b), qu'il existe des zones, appelées cônes de confusions, qui correspondent à des iso-valeurs de l'ITD ou de l'ILD. Malgré la présence des indices spectraux, il est relativement courant de voir apparaître des confusions de localisation dans ces zones. Ces confusions peuvent se traduire en inversions avant/arrière et haut/bas. Pour les confusions avant/arrière, par exemple, une source réelle placée devant l'auditeur à 30° sera perçue à l'arrière à 150°. Dans la plupart des études avec des sons réels, le taux moyen d'inversions avant/arrière est autour de 5% [Carlile *et al.*, 1997, Makous, 1990, Wightman et Kistler, 1989b]. Le taux moyen d'inversions haut/bas a quant à lui, été moins étudié. L'étude de [Wenzel *et al.*, 1993] sur 16 sujets reporte environs 6% de confusions haut/bas.

iv/ Performances de localisation en distance

L'aptitude des humains à percevoir la distance d'une source sonore est relativement médiocre. En général, les auditeurs ont tendance à sous-estimer les distances supérieures à 1 mètre et à surestimer les distances inférieures [Zahorik, 2002]. La dispersion des reports de distance perçue est toujours assez élevé (de l'ordre de 25% de la distance réelle de la source) et ce indépendamment du type de stimuli ou de la méthode de d'estimation de la distance (i.e. réponses reportées sur une échelle explicite (en mètre ou en pied) [Zahorik, 2002], ou basées sur une action physique [Loomis *et al.*, 1998b]). La figure 2.8, tirées de [Blauert, 1997] d'après des travaux de [Haustein, 1969], récapitule les performances d'estimation de la distance pour des sources de 1 à 8 mètres.

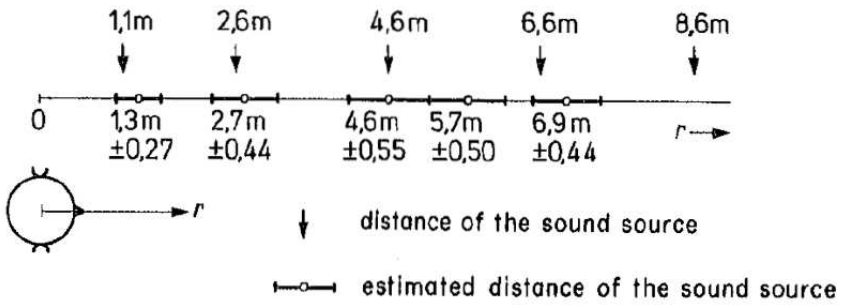


FIGURE 2.8 – Récapitulatif des performances de localisation auditive de la distance. Image tirée de [Blauert, 1997] d'après une étude de [Haustein, 1969] réalisée sur 20 sujets avec des impulsions de bruit blanc. Les flèches indiquent la distance des sources, les cercles représentent les moyennes des réponses des sujets et les portions de ligne plus épaisses les écarts types.

c) Localisation de sons virtuels

Bien que reproduisant théoriquement tous les indices acoustiques de localisation, les performances de localisation de sons virtuels générés avec de la synthèse binaurale sont en général beaucoup plus mauvaises que les performances de localisation de sons réels. Dans cette partie, nous allons voir les dégradations dues à la synthèse binaurale avec des HRTF individuelles et non-individuelles.

i/ Performances de localisation en direction

La comparaison de la localisation de sons réels (diffusés par des haut-parleurs en chambre anéchoïque) à la localisation de sons virtuels (avec des HRTF individuelles) a été réalisée pour des positions similaires par [Wightman et Kistler, 1989b]. Dans cette étude de référence, les auteurs ont observé les artefacts les plus significatifs de la synthèse binaurale : un pourcentage plus important de confusions avant/arrière (11% pour les sons virtuels contre 5% pour les sons réels) et une erreur angulaire plus grande en élévation. La localisation en azimut n'est par contre quasiment pas modifiée. Ces dégradations de performances peuvent être expliquées par la précision des mesures réalisées, la quantification de la grille de mesure (les HRTF ne peuvent pas être mesurées de façon continue dans toutes les directions), par l'influence de la fonction de transfert du casque ou par d'autre facteurs tels que l'absence d'indices visuels corrélés aux indices auditifs.

L'utilisation d'HRTF non-individuelles introduit de grandes distorsions dans la perception des sources virtuelles. [Middlebrooks, 1999b] a comparé les performances de localisation avec HRTF personnalisées et non personnalisées sur un grand nombre de position en azimut et en élévation. Ses résultats mettent en évidence une grande augmentation du taux de confusion (les auteurs ne distinguent pas les inversions avant/arrière des inversions haut/bas) qui passe de 5% pour les HRTF individuelles à 20% pour les HRTF non-individuelles ; une dégradation des performances en élévation (avec une erreur d'environs 25° pour les HRTF individuelles contre 40° pour les non-individuelles) ; et une légère dégradation des performances en azimut (l'erreur augmente d'environs 4°). D'autre études réalisées uniquement avec des HRTF non-individuelles [Zahorik *et al.*, 2006] ou comparant la localisation de sources virtuelles non-individuelles à des sources réelles [Wenzel *et al.*, 1993]

confirment ces résultats.

Les études de localisation avec des sons virtuels reportent aussi d'autre problèmes perceptifs, tels que l'altération du timbre, le manque d'externalisation de la source (la source paraît être située dans la tête, entre les deux oreilles) ou encore une grande taille apparente de la source [Begault, 1994, Larcher, 2001].

ii/ Performances de localisation en distance

Étant donné que les HRTF sont mesurées à distance fixe et en chambre anéchoïque, les performances de localisation en distance avec des sons virtuels ont été peu étudiées et sont très faibles. Il est néanmoins possible par l'ajout de l'indice d'intensité (loi en $1/r^2$ citée dans le paragraphe 2.2.1.d)) et de l'effet d'absorption par l'air, d'approcher la perception de distance avec des sons réels en champ libre [Brungart, 1993, Zahorik, 2002]. La variabilité des résultats reste néanmoins très grande à cause du manque d'externalisation des sons pour les sources frontales. [Begault, 1992] a montré que l'ajout d'un effet de salle dans la synthèse binaurale pouvait améliorer la perception de la distance ainsi que la sensation d'externalisation du son. Ses résultats ont été confirmés par plusieurs études de [Kopčo *et al.*, 2008, Kopčo et Shinn-Cunningham, 2011] utilisant des réponses impulsionnelles binaurales enregistrées dans des salles réverbérantes. Il apparait nécessaire pour simuler correctement la distance en environnement virtuel binaural d'ajouter un effet de salle, bien que cet ajout entraîne une dégradation des performances de localisation en azimut.

iii/ L'effet du casque

La synthèse binaurale étant diffusée sur un casque, la fonction de transfert de celui-ci peut avoir une grande influence sur le spectre des HRTF. Il peut donc être nécessaire de la compenser pour pouvoir contrôler finement la pression acoustique aux tympans de l'auditeur. De nombreuses études se sont penchées sur la question de la compensation du casque pour la synthèse binaurale [Møller *et al.*, 1995, Pralong et Carlile, 1996, Kulkarni et Colburn, 2000, McAnally et Martin, 2002, Schonstein *et al.*, 2008, Schärer et Lindau, 2009, Paquier *et al.*, 2011]. La réponse fréquentielle du casque n'est en général pas plate. De plus, le couplage entre le casque et les pavillons peut entrainer des résonances et des antirésonances prononcées qui peuvent ressembler fortement aux caractéristiques spectrales des HRTF. Il semble être établi que la non compensation du casque peut entraîner une dégradation de la localisation des sources virtuelles, cependant les études de la littératures ne sont pas forcement en accord sur les indices à compenser (seulement la fonction de transfert du casque ou l'ensemble casque et couplage casque/oreilles). Selon [Kulkarni et Colburn, 2000], la difficulté à évaluer de manière fiable la fonction de transfert de l'ensemble casque et couplage reste un problème majeur notamment à cause de son caractère individuel et des différentes manières de positionner le casque (voir [Paquier *et al.*, 2011], pour l'effet de la position du casque). Selon [McAnally et Martin, 2002], les variations fréquentielles engendrées par le casque sont généralement moindres que les colorations des HRTF et ne posent donc pas de problèmes insurmontables. [Pralong et Carlile, 1996] montrent, quant à eux, qu'une calibration

non-individuelle du casque peut engendrer une dégradation de qualité de la localisation équivalente à celle provoquée par l'utilisation d'HRTF non-individuelles. [Wightman et Kistler, 2005] montrent que sans calibration du casque, les performances de localisation en élévation se dégradent et le taux de confusions avant/arrière augmente. Les variations fréquentielles entraînées par le casque étant individuelles et fortement liées à la position du casque, une compensation systématique de cette fonction de transfert semble difficile à mettre en place. De plus, dans le cadre d'un projet commercial, les utilisateurs n'utiliseront pas les mêmes casques et la calibration sera donc impossible. Nous avons donc, pour la suite du document, choisi de ne pas effectuer de calibration du casque, considérant que de toute manière, les effets du casque et du couplage casque/auditeur ne dépendent pas de la position de la source.

2.3 La sonification

Composante essentielle du domaine de l'*Auditory Display (AD)*[2], la sonification utilise des messages sonores non vocaux pour transmettre des informations. Elle est définie par [Kramer *et al.*, 1999] comme "la transformation de relations entre des données visuelles (ou autre) en relations perçues dans un signal acoustique afin d'en faciliter la communication ou l'interprétation". En d'autres termes, la sonification vise à exploiter la modalité auditive en traduisant des informations visuelles, cartographiques ou autre, sous forme sonore ; ceci afin de limiter la surcharge d'informations fournies par les interfaces graphiques.

Les motivations et justifications de l'utilisation de signaux sonores (plutôt que visuels ou autre) pour présenter des informations ont été discutées en détail dans plusieurs études [Sanders et McCormick, 1993, Kramer, 1993, Kramer *et al.*, 1999, Hermann *et al.*, 2011].
Premièrement, la sonification exploite les capacités du système auditif à analyser et reconnaitre avec une grande précision des changements temporels ou fréquentiels [Bregman, 1994, McAdams et Bigand, 1993, Moore, 1997, Kramer *et al.*, 1999]. La modalité auditive est donc bien appropriée pour représenter des informations dépendantes du temps ainsi que des motifs complexes.
Deuxièmement, l'utilisation du son permet de soulager la modalité visuelle lorsque celle-ci est déjà encombrée par d'autre tâche [Fitch et Kramer, 1992], lorsque beaucoup d'informations doivent être affichées [Brewster, 1997, Brown *et al.*, 1989] ou permet de la suppléer lorsqu'elle n'est pas accessible (pour les non-voyants ou les pompiers aveuglés par la fumée d'un incendie) [Fitch et Kramer, 1992].
Troisièmement, la perception auditive permet d'écouter, de surveiller et de traiter plusieurs flux sonores différents en même temps [Moore, 1997].
Enfin, avec l'apparition des smartphones, des tablettes électroniques et autre gadgets technologiques mobiles de taille réduite, ces dernières années, la sonification permet d'afficher des informa-

2. L'*Auditory Display* ou Affichage Auditif est un terme qui englobe tous les aspects d'un système de rendu sonore. Il comprend : le système de diffusion du son (haut-parleurs, casques), le type de rendu utilisé (stéréophonie, synthèse binaurale, spatialisation ambisonique, ...), ainsi que toute solution technique pour la collecte, le traitement et les calculs nécessaire pour obtenir un son en réponse à des données [Kramer, 1993]

tions sans que l'utilisateur n'ait besoin de regarder l'écran et permet de compléter la vision lorsque l'écran est trop petit pour afficher toutes les informations [Brewster et Murray, 2000].

Dans cette partie, nous allons décrire les différentes fonctions de la sonification puis nous aborderons les principales techniques les plus couramment utilisées pour transmettre des informations sous forme sonore.

2.3.1 Les fonctions de la sonification

Étant donné que la modalité auditive a des propriétés inhérentes qui peuvent se révéler bénéfique pour l'affichage d'informations, nous allons examiner ici quelques types de fonctions que l'affichage auditif et la sonification peuvent effectuer. [Buxton, 1989] puis [Edworthy, 1998, Walker et Kramer, 2004] ont décrit les fonctions de la sonification en terme de trois grandes catégories : (1) alarmes, alertes et avertissements, (2) messages d'état, de processus et de suivi d'une tâche, et (3) exploration de données ; auxquels ont été rajoutées plus tard [Walker et Nees, 2011] : (4) le divertissement, le sport et l'exercice.

a) Fonction de notification

La première catégorie de fonction de la sonification, correspond aux notifications. Du simple avertissement, à l'alarme ou l'alerte, les notifications sonores permettent d'indiquer qu'un événement vient de se produire ou va se produire. Le message véhiculé par les alertes est généralement relativement pauvre et a pour but d'indiquer une information simple à l'auditeur [Buxton, 1989, Sorkin, 1987]. Le plus commun des exemples d'alerte sonore est le "ding-dong" de la sonnette de porte qui indique la présence d'une personne derrière la porte. Ce type de signal est aussi utilisé par les micro-onde pour indiquer la fin du temps de cuisson ou par les téléphones portables pour indiquer l'arrivée d'un message texte. Les alarmes et avertissements sont des notifications sonores destinées à véhiculer l'apparition d'une classe restreinte d'événements, le plus souvent urgent et défavorable, qui exigent une réponse immédiate ou au moins une grande attention [Haas et Edworthy, 2006]. Pour ce type de notification, [Spence et Driver, 1997] ont montré que la modalité auditive capte plus facilement l'attention du sujet que la modalité visuelle et qu'elle permet d'éviter les problèmes dus à la limitation du champ de vision. Les alarmes doivent pouvoir véhiculer plus d'informations que les alertes, tel que le niveau d'urgence ou le type de problème. De nombreuses études (telles que [Edworthy *et al.*, 1991, Suied *et al.*, 2008]), ont exploré la relation entre les sons utilisés et le degré d'urgence ressenti par les auditeurs.

b) Fonction d'indication de statut ou de progrès

Généralement la notification seule ne permet pas de donner suffisamment de détails sur l'information qui doit être transmise via la modalité audio. Dans de nombreux cas, il est nécessaire que

la sonification puisse donner plusieurs informations sur l'état d'un système, d'un processus ou sur une série d'événements. Dans ces cas, la modalité audio tire profit de la capacité de l'auditeur à détecter de petits changements dans les événements sonores et permet à l'utilisateur d'avoir les yeux libres pour d'autre tâches [Kramer *et al.*, 1999]. Les systèmes d'exploitation informatiques tels que *Windows* ou *MacOs* utilisent depuis de nombreuses années des "grammaires" de signaux auditifs pour donner des indications de statuts ou de tâches. Ces "grammaires" sonores constituent un ensemble de notifications audio portant chacune un sens particulier faisant référence à une tâche ou un état. Les différentes manières de créer un ensemble de notifications audio ont été étudiées par [Gaver, 1986, Blattner *et al.*, 1989, Dingler *et al.*, 2008] et sont détaillées dans la section 2.3.2.d). L'affichage auditif a aussi été étudié pour présenter des menus [Helle *et al.*, 2001, Walker et Kogan, 2009, Langlois *et al.*, 2010], des barres de progrès [Peres *et al.*, 2007], pour surveiller l'état de différents processus dans une usine [Walker et Kramer, 1996] ou pour indiquer l'horizon artificiel en aviation [Brungart et Simpson, 2008].

c) Fonction d'exploration de données

La troisième fonction de la sonification consiste à permettre l'exploration de données complexes et multidimensionnelles. Alors que l'exploration de données par la modalité visuelle est limitée à deux ou trois dimensions, la modalité audio permet d'explorer et d'étudier des données évoluant temporellement selon de multiples dimensions. Cette fonction se distingue des précédentes étant donnée qu'elle est utilisée pour afficher une vue global des données plutôt qu'un résumé ou un état momentané du système. De l'exploration de données scientifiques [Flowers et Hauer, 1993, Brown *et al.*, 2003, Stockman *et al.*, 2005] à l'écoute de données médicales [Hermann *et al.*, 2002, Baier *et al.*, 2007, Pauletto et Hunt, 2009], ce type de sonification permet de suppléer ou même de remplacer la vision lorsque celle-ci se révèle insuffisante pour analyser des données.

d) Fonction de divertissement

Plus récemment, la sonification a été appliquée au domaine des loisirs, des activités sportives et même de l'art. Les interfaces auditives, d'abord utilisées pour générer des sons environnementaux et accompagner l'image dans les jeux vidéos [Röber *et al.*, 2006, Verron *et al.*, 2010] ont été utilisées avec la sonification pour créer des jeux auditifs (ou *audio games*) [Friberg et Gärdenfors, 2004, Gaudy *et al.*, 2006]. Ces jeux électroniques basés uniquement sur le son permettent de faciliter l'accès des jeux aux non-voyants et parfois même faciliter l'interaction directe entre des joueurs voyants et non-voyants [Stockman *et al.*, 2007].

Au niveau des activités sportives, les études de [Schaffert *et al.*, 2009] sur l'aviron ou de [Godbout et Boyd, 2010] sur le patinage de vitesse, ont montré le potentiel de la sonification à fournir un retour sonore permettant aux sportifs d'améliorer leurs mouvements. La sonification du mouvement peut aussi être utilisée pour informer du bon déroulement d'une tâche comme,

par exemple, de la justesse du geste d'écriture chez les enfants atteints de troubles dysgraphiques [Thoret *et al.*, 2012].

2.3.2 Techniques et approches de sonification

Différentes techniques de sonifications peuvent être utilisées en fonction du type d'information à afficher (i.e. message de notification ou variation d'une donnée continue). Nous allons dans cette partie décrire les principales approches utilisées en sonification.

a) Audification

L'*audification* [Dombois et Eckel, 2011] est la méthode de sonification la plus directe. Elle consiste à afficher des données physiques évoluant au cours du temps en les transposant dans le domaine auditif. Ces données peuvent être issues de capteurs ou de simulations. Le fait de les "écouter" peut permettre d'y détecter des anomalies, de faciliter leur catégorisation ou d'explorer l'effet d'un changement de condition sur ces données. Cette approche peut nécessiter de transformer le signal à sonifier avec des dilatations temporelles ou fréquentielles afin de le transposer dans le domaine des ondes audibles (de 20 à 20000 Hz). L'*audification* est utilisée dans divers domaines, tels que : la médecine : avec le stéthoscope, toujours très utilisé, qui peut-être vu comme un des premiers exemple d'audification ; l'audification des signaux provenant des Électro-Encéphalogramme (EEG) [Olivan *et al.*, 2004] ou l'étude des battements du cœur [Ballora *et al.*, 2000] ; la sismologie [Dombois, 2002, Meier et Saranti, 2008] ; la physique [Pereverzev *et al.*, 1997, Vézien *et al.*, 2009] ou les statistiques [Frauenberger *et al.*, 2007].

b) Sonification par mapping de paramètres

La sonification par *mapping de paramètres* [Grond et Berger, 2011] est la plus commune des méthodes de sonification utilisée pour l'analyse de données scientifiques. Elle consiste à représenter une ou plusieurs dimensions des données par des changements de un ou plusieurs paramètres acoustiques. Les paramètres sonores utilisés peuvent être d'origine fréquentielle (faire varier la fréquence, le timbre, l'enveloppe spectrale, etc.), temporelle (tempo, motif rythmique, etc.) ou d'amplitude (volume, temps d'attaque, modulation d'amplitude, etc.).

Étant donné qu'il est possible de faire varier un grand nombre de paramètres sonores et qu'il existe de nombreuses stratégies de mapping, cette méthode permet de nombreuses représentations différentes d'un même phénomène tout en échappant aux limitations bi- ou tridimensionnelles des représentations graphiques. Le choix des paramètres acoustiques et le mapping de ceux-ci doit être le résultat de plusieurs considérations telles que le nombre de données à sonifier, leur échelle de variation ainsi que les tâches à effectuer avec ces données (surveillance, interprétation, analyse, etc.). Il est cependant nécessaire de prendre en compte les limites psychophysiques du système auditif

représentation	**température**	**pression**	**vitesse**	**taille**
intuitive	fréquence	attaque	tempo	intensité
correcte	intensité	fréquence	attaque	tempo
mauvaise	attaque	tempo	intensité	fréquence
aléatoire	tempo	intensité	fréquence	attaque

TABLE 2.1 – Tableau des différents types de mapping utilisés pour contrôler la fabrication du verre dans une usine. Tiré de [Walker et Kramer, 1996]

telles que : la plus petite différence notable (*just-noticeable differences* ou JND), le masquage et les seuils de perception en terme d'amplitude, de fréquence et de variations temporelles.

L'étude de [Walker et Kramer, 1996] a montré que le choix des paramètres acoustiques ainsi que la polarité et l'étendue de leur variations pouvaient avoir une grande influence sur les performances. Ils ont expérimenté dans leur étude, des mappings de différentes données (température, pression, taille et vitesse) à différents paramètres acoustiques (intensité, fréquence, tempo et attaque) en simulant une opération de contrôle de la fabrication de verre dans une usine. Les sujets effectuaient une tâche de surveillance avec un mapping supposé intuitif ainsi qu'avec trois autres types de mapping : "correct", "mauvais" et "aléatoire" (cf tableau 2.1). Étonnamment, les sujets ont obtenu de meilleures performances (en pourcentage de bonne réponse et en temps de réponse moyen) avec les mapping "mauvais" et "aléatoire". Les sujets ont ensuite répété l'expérience avec une polarité de mapping inversée (la fréquence diminue lorsque la température augmente, par exemple). Les résultats ont mis en évidence une grande influence de la polarité du mapping. Cette expérience montre la nécessité pour le designer sonore de ne pas se fier uniquement à son intuition pour effectuer la sonification et de tester plusieurs mapping différents avant de déterminer celui qui représentera au mieux les données à représenter.

La sonification par *mapping de paramètres* est généralement appliquée à des sons MIDI ou à des sons de synthèses à cause de la facilitée d'accès aux différents paramètres acoustiques de ces sons.

c) Sonification basée sur des modèles

Dans l'approche de *sonification basée sur des modèles* [Hermann, 2011], plutôt que de relier les paramètres des données aux paramètres du son, [Hermann et Ritter, 1999] proposent de construire un modèle virtuel dont la réponse sonore à une entrée de l'utilisateur dépendrait des données à sonifier. Ainsi, l'idée est de : (i) construire un scénario virtuel à partir des données à sonifier, (ii) définir un "modèle physique virtuel" qui régirait la réaction vibratoire des données à des excitations externes, (iii) permettre à l'utilisateur d'exciter et d'écouter le système de manière interactive.

Ce type de sonification est généralement basée sur la synthèse par modèle physique [Smith, 1992] ce qui la rend un peu plus agréable que la sonification par *mapping de paramètres*. Plusieurs applications pourront être trouvées, par exemple, dans [Bovermann *et al.*, 2005, Hermann *et al.*, 2001, Tünnermann et Hermann, 2009].

d) Signaux auditifs

Nous appelons *signaux auditifs*, les événements sonores généralement de courte durée utilisés dans les interfaces informatiques pour donner une information sur un état, une action réalisée, le contenu d'un dossier ou encore le résultat d'une action de l'utilisateur sur le système. Il existe plusieurs types de *signaux auditifs*. Dans cette partie nous allons définir les trois types les plus couramment étudiés dans le domaine de l'*Auditory Display* : les *auditory icons*, les *earcons* et les *spearcons*.

i/ Auditory Icons

Les *auditory icons* ont été introduit par [Gaver, 1986] comme "des sons du quotidien reliés à des événements informatiques par analogie avec les événements produisant ces sons". L'idée basique de ces *icônes sonores* est d'utiliser les connaissances de l'utilisateur basées sur l'écoute de tous les jours pour représenter des actions au sein d'une interface. Ils sont constitués de sons brefs qui peuvent être considérés comme l'équivalent auditif des icônes visuels utilisés dans l'ordinateur. La relation entre le son et l'objet représenté doit être naturelle et intuitive : le son est relié sémantiquement à ce qu'il doit représenter (par exemple, dans un ordinateur, l'action de jeter un fichier à la poubelle va être représentée par le son d'un papier que l'on froisse). L'avantage des *auditory icons* est leur temps d'apprentissage très court. Leur inconvénient est la difficulté à trouver un son représentatif (une représentation iconique claire ayant une relation directe avec l'information à sonifier) pour chaque objet, fonction ou action. Ils sont, de plus, peu utile lorsque l'on doit représenter des concepts abstraits.

ii/ Earcons

Les *earcons* sont des motifs sonores abstraits, synthétiques et souvent musicaux qui peuvent être combinés pour créer des grammaires sonores. Ils sont définis par [Blattner *et al.*, 1989] comme des "messages sonores non verbaux utilisés dans les interfaces d'ordinateur pour transmettre des informations à l'utilisateur sur des objets, des opérations ou des interactions". Ce sont des messages symboliques courts, n'utilisant pas la parole et construits à partir de blocs simples appelés "motifs". Selon [Blattner *et al.*, 1989], un motif est une succession de hauteurs arrangées pour produire un motif rythmique et tonal suffisamment distinct pour fonctionner comme une entité individuelle et reconnaissable. À partir de ces motifs, il est possible de créer une syntaxe hiérarchique d'earcons permettant de représenter des données avec plusieurs niveaux d'informations (comme un arbre de données ou la combinaison de plusieurs actions). Selon [Brewster *et al.*, 1993], cet arbre est limité à cinq niveaux car il n'est possible de faire varier que cinq paramètres : le rythme, la hauteur, le timbre, le registre et la dynamique. Contrairement aux *Auditory Icons*, ils ont l'avantage de permettre de représenter tout type d'information ou de concept, par contre, ils nécessitent un apprentissage non négligeable.

iii/ Spearcons

Introduits par [Walker *et al.*, 2006], les *spearcons* sont constitués de phrases synthétisées suffisamment accélérées pour qu'elles ne soient plus reconnues comme de la parole. Basés sur le texte décrivant l'information qu'ils représentent, les *spearcons* peuvent être facilement créés en utilisant un logiciel de synthèse vocale et un algorithme pour accélérer la phrase. Étant donné que la relation entre le *spearcon* et l'objet qu'il représente n'est pas arbitraire, leur utilisation ne nécessite qu'un petit entraînement. De plus, chaque élément à sonifier étant différent, la description textuelle est différente et chaque *spearcon* résultant est unique et distinct des autres. Comme les *earcons*, ils peuvent être combinés pour créer une sonification hiérarchique.

2.4 Les systèmes d'aide aux non-voyants

Les systèmes d'aide aux non-voyants ont pour objectif de restituer certaines fonctions assurées par le système visuel et dont l'absence peut engendrer un besoin chez les déficients visuel. Ces aides techniques sont des moyens destinés à permettre à la personne de compenser l'absence de vision. Elles doivent fournir une information équivalente à celle fournie par la vision par l'intermédiaire d'une autre modalité sensorielle telle que l'audition ou la somesthésie.

Des montres braille ou parlantes, aux logiciels de lecture d'écrans et claviers braille, en passant par les aides au déplacement, il existe de nombreux dispositifs commercialisés et de nombreux projets de recherche visant à augmenter l'accessibilité des informations visuelles pour les personnes non-voyantes. Ces systèmes peuvent être basés sur la somesthésie (en convertissant une image ou un texte en stimuli tactiles, comme par exemple le braille), ou sur l'audition (moteurs de synthèse transformant une chaîne de caractères en information vocale). Ils sont en général basés sur la même architecture : une chaîne d'acquisition, une chaîne de transformation de l'information et un module de restitution de l'information traitée.

Au cours des dernières décennies, le développement des systèmes de guidage par satellite (Global Positioning Systems GPS), des caméras portatives et des systèmes de captation du mouvement, ainsi que la réduction de la taille et l'augmentation de la puissance des ordinateurs personnels ont conduit à l'élaboration d'un grand nombre de projets d'aide à la navigation pour les non-voyants. Une étude récente sur les systèmes d'aide à la mobilité existants pour les non-voyants [Roentgen *et al.*, 2008] identifie plus de 140 dispositifs d'assistance. Ces dispositifs visent à aider les non-voyants dans une des trois phases du modèle de la navigation pédestre proposé par [Adams et Beaton, 2000]. Ce modèle stipule que le déplacement de manière autonome d'un point A vers un point B est une tâche cognitivement complexe qui nécessite plusieurs comportements distincts :

– *la planification d'itinéraire* : l'individu doit prendre en compte son point de départ et son point d'arrivée, comparer les différentes routes permettant de joindre ces deux points et décider, sur la base de critères de distance, de temps de parcours et de sécurité de l'itinéraire, quelle est la meilleure route à emprunter. De cette phase de préparation résulte la mentalisation de

l'itinéraire à suivre, composé d'un certain nombre de points de réorientation reliés entre eux par des segments de route plus ou moins long.
- *la navigation fine* : qui consiste à éviter les imprévus et obstacles ainsi qu'à gérer les difficultés rencontrées aux différents croisements (passages piétons, feux tricolores, obstacles, trafic routier, etc.).
- *la navigation globale* : qui consiste à relier entre eux les différents points de réorientation, se rappeler de la direction à emprunter aux intersections et garder cette direction pour rester sur le chemin.

La recherche sur les systèmes de suppléance visuelle permettant d'aider les non-voyants dans ces trois phases a conduit à deux catégories de systèmes : les systèmes de substitution sensorielle et les systèmes d'augmentation sensorielle [Kaczmarek, 2000]. Apparus dans les années 1970, les systèmes de substitution sensorielle restituent les informations habituellement acquises par une modalité sensorielle en utilisant directement une autre modalité sensorielle. Pour les non-voyants, ils sont en général basés sur les substitutions visuo-tactile et visuo-auditive. Apparus un peu plus tard, les systèmes d'augmentation sensorielle restituent des informations extraites de la modalité visuelle (par exemple), en utilisant une autre modalité. Contrairement aux dispositifs de substitution, ils ne restituent pas l'intégralité du message capté et nécessitent donc une étape de traitement des données visant à extraire l'information pertinente à transmettre.

Nous allons, dans cette section, donner quelques exemples de systèmes de suppléances permettant de fournir une aide au déplacement pour la navigation fine (Electronic Travel Aids - ETA) et une aide à l'orientation pour la navigation globale (Electronic Orientation Aids - EOA), en séparant dans le cas des ETA, les systèmes de substitution et les systèmes d'augmentation sensorielle.

2.4.1 Les aides au déplacement

Les dispositifs d'aide au déplacement ont pour objectif de permettre au non-voyant de se déplacer de manière autonome dans l'espace proche (entre 1 et 20 mètres) et reposent sur la détection des obstacles ou sur la description de l'environnement proche du sujet par une appréciation du relief ou des objets environnants. Ces aides sont, en général, des compléments à la canne blanche ou au chien.

Les premières idées d'expérimentations sur les aides au déplacement sont apparues avec l'émergence de la substitution sensorielle introduite dans les années 60 par Paul Bach-y-Rita. Ces systèmes reposent sur une (ou plusieurs) caméra, considérée comme capteur de substitution de la vision humaine déficiente. Les images, brutes ou peu filtrés, sont directement reproduites sous la forme d'un signal tactile ou auditif. Apparus un peu plus tard, les systèmes d'augmentation sensorielle nécessitent une transformation de l'image ou des données issues de capteurs spécifiques afin de transmettre uniquement les informations pertinentes.

a) Les systèmes de substitution sensorielle

Les premières tentatives de substitution sensorielle avaient pour but d'utiliser la plasticité cérébrale des non-voyants afin de restaurer leurs capacités visuelles sur la base de stimulations tactiles. Premier système, le TVSS, réalisé par [Bach-y Rita *et al.*, 1969] était constitué d'un fauteuil de dentiste équipé d'actuateurs tactiles permettant de convertir les informations visuelles capturées par une caméra en des sensations tactiles à la surface du corps (sur le dos). Dans sa première version, la caméra était fixe et les sujets avaient de grandes difficultés à percevoir des formes. [Bach-y Rita, 1983] puis plus tard [Auvray *et al.*, 2007] ont montré que la manipulation du capteur d'image ou d'information par le sujet est essentielle pour la perception des formes provenant de la scène visuelle. Ce dispositif a ensuite été adapté à l'abdomen [Bach-y Rita, 1983], puis à la langue [Bach-y Rita *et al.*, 1998, Kupers et Ptito, 2004] et au palais [Tang et Beebe, 2003]. Étant donné la faible résolution tactile du dos ou de l'abdomen, [Bach-y Rita *et al.*, 1998] proposent d'utiliser la langue (car c'est un des organes avec la plus forte densité de récepteurs tactiles, permettant ainsi une plus grande résolution de stimulation) et montrent qu'il est possible de reconnaitre des formes simples après un certain temps d'apprentissage avec le dispositif.

Apparus un peu plus tard, les systèmes de substitution visuo-auditive convertissent l'image capturée en information sonore en préservant le maximum d'information spatiale et lumineuse de l'image. Ces systèmes utilisent, en général, la fréquence, l'intensité, le temps et la stéréophonie pour restituer la position des pixels et leur intensité lumineuse.

Le plus connu de ces systèmes, "The vOICe" [Meijer, 1992], est développé depuis 1992 par Peter Meijer. Avec ce système, l'image est convertie en matrice de niveau de gris de 64x64 pixels. La position verticale est codée sur 64 fréquences différentes (plus le motif visuel est haut, plus le son est aigu) et la position horizontale est codée de manière temporelle (le balayage d'une image est réalisé en une seconde). Le niveau de gris de chaque pixel est restitué avec l'intensité du son. Ainsi, plus le niveau de gris d'un pixel est clair, plus le niveau sonore de la fréquence correspondant à ce pixel sera élevé. [Auvray, 2004] a montré qu'il est possible avec ce dispositif de localiser (avec un temps moyen de 100 ± 70 secondes et une erreur de 7 ± 5 cm) et de reconnaitre un objet parmi une dizaine en manipulant la caméra avec la main. Un système similaire, "The Vibe" a été développé par [Hanneton *et al.*, 2010] en utilisant une restitution stéréophonique pour coder la position horizontale des pixels. L'efficacité de ce système sur l'évitement d'obstacles a été validée en condition réelle dans une tâche de navigation réalisée avec 20 non-voyants [Durette *et al.*, 2008]. Le système EAV [Gonzalez-Mora *et al.*, 2006], quant à lui, constitué de caméras portées sur des lunettes, utilise la stéréoscopie pour localiser la surface des objets présents dans la scène visuelle et du son binaural pour synthétiser des sons comme s'ils provenaient de petites enceintes placées à la surface des objets.

Ces différents systèmes permettent de localiser et de reconnaitre des objets à partir d'images avec des règles de conversion très simples ne nécessitant que très peu de ressources et un temps de calcul réduit. Ils ne permettent cependant que de reconnaitre des motifs simples et sont donc

difficilement utilisable en environnements naturels. La différence entre la résolution nécessaire pour percevoir un objet dans une image et la résolution de la modalité sensorielle cible est aussi un frein important. De plus, la somme d'information présente à l'image dans un environnement complexe est trop importante pour pouvoir être interprétée aisément. Il est donc nécessaire d'utiliser des traitements pour sélectionner les informations pertinentes avant de les présenter.

b) Les systèmes d'augmentation sensorielle

Plutôt que de chercher à transcrire l'intégralité de la modalité visuelle à travers une autre modalité, les systèmes d'augmentation sensorielle visent à restituer certaines fonctions du système visuel parmi les plus utiles aux non-voyants. En général, ces systèmes suivent une approche basée sur les télémètres ou une approche basée sur les systèmes de vision artificielle.

L'approche basée sur les télémètres vise à restituer la fonction de détection des obstacles de la vision en calculant la distance aux objets présents dans une zone frontale à l'utilisateur. Il existe principalement deux technologies permettant d'estimer la distance à des objets environnants avec précision : les télémètres à ultrasons et les télémètres laser. Dans chacun des cas, le principe du télémètre est le même : un signal sonore ou lumineux envoyé par un émetteur se réfléchi sur les objets environnant le sujet et sont captés par un récepteur placé à côté de l'émetteur. Le déphasage entre le signal émis et le signal réceptionné permet d'estimer la distance des objets environnants. Les télémètres à ultrasons permettent d'estimer des distances inférieures à 20 mètres avec une précision fortement dépendante des facteurs environnementaux (température, humidité) ; les télémètres laser, qui ont une portée de quelques centaines de mètres, sont plus directionnel et plus précis mais plus chers et inopérant pour détecter les surfaces transparentes (comme les vitres).

De nombreux systèmes basés sur l'approche du télémètre ont été développés dans le cadre de projets de recherche. Certains comme l'Ultracane ont même abouti à une commercialisation (par la société Foresight [3]). L'Ultracane est une canne blanche augmentée d'un télémètre à ultrason restituant la distance avec un retour tactile via des boutons vibrants placés sur la poigné de la canne. Utilisant le principe de triangulation par profilométrie laser, le Télétact [Farcy et Damaschini, 1997] développé par René Farcy (laboratoire Aimé Cotton) a permis d'explorer plusieurs types de modalités de restitution de la distance. Déclinée sous plusieurs versions (Mini Tact, Tom Pouce et Télétact, [Farcy *et al.*, 2006]) permettant un apprentissage graduel du système, cette canne donne des informations sur la distance en utilisant, soit des stimuli tactiles (sur quatre doigts correspondant à quatre plages de distances), soit des stimuli audio (32 notes permettant de représenter des distances allant de 0 à 15 mètres) [Jacquet *et al.*, 2006]. [Farcy *et al.*, 2003] ont montré que l'expertise acquise sur ce système permettrait de reconnaitre certaines formes par exploration de celles-ci en les balayant avec le faisceau. D'autre projets tels que "The GuideCane" de [Borenstein et Ulrich, 1997] ou "The Navbelt" de [Shoval *et al.*, 1998] ou [Bensaoula *et al.*, 2006] utilisent une combinaison de capteurs à ultrasons et de télémètres laser afin de couvrir une plus grande zone. Étant donné la

3. http ://www.ultracane.com/

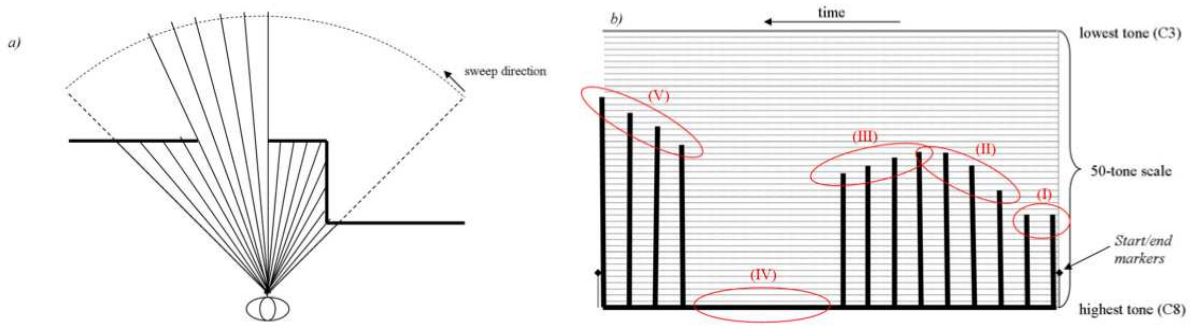


FIGURE 2.9 – Extrait d'une période d'un balayage horizontal extrait de la thèse de [Bujacz, 2010]. a) Valeur de la distance prise tout les 5°. b) Séquence de note correspondant aux distances de la figure a). Les balayages sont effectués de droite à gauche. Sur la figure b), l'axe temporel en abscisse va de droite à gauche et l'ordonnée correspond à la hauteur du son.

taille et surtout le nombre des capteurs, leurs dispositifs souvent volumineux doivent être portés en ceinture ou sur un chariot à roulette ce qui représente une grosse contrainte pour les utilisateurs.

Ces systèmes, bien que permettant de détecter des obstacles, ne remplacent pas la canne blanche ou le chien et restent très limités dans la mesure où ils ne permettent pas de lire les noms des rues ou de s'orienter dans les environnements inconnus.

L'approche basée sur les systèmes de vision artificielle consiste à traiter les signaux provenant d'une caméra afin d'en extraire une information qualitative sur les éléments composant l'image. Elle est plus complexe et demande souvent beaucoup de ressources au système ; de plus, elle requiert deux caméras afin d'utiliser une carte de disparité pour déterminer la distance des objets détectés. Les premiers systèmes de ce type ont cherché à déterminer la position de l'objet le plus proche afin de donner sa position avec du son 3D [Kawai *et al.*, 2000, Fontana *et al.*, 2002a], des sons stéréophoniques [Balakrishnan *et al.*, 2004] ou une interface tactile [Costa *et al.*, 2008]. [Alba *et al.*, 2008] utilisent cette technique pour localiser trois à cinq objets et leur associer des sons différents (en utilisant des sons issus de synthèse FM spatialisés en binaural). Au niveau de la restitution des informations, [Bujacz, 2010] a testé, dans le cadre de sa thèse, différentes méthodes de sonifications 3D des informations extraites de la scène visuelle [Bujacz *et al.*, 2011]. La première méthode, présentée dans [Pelczynski *et al.*, 2006] consiste à utiliser de la synthèse vocale générée par formants et présentée en binaural avec des HRTF individuelles. Les voyelles permettent de différencier les différentes plages de distances et le type de voix (masculine ou féminine) permet de donner une information sur la taille des objets. Dans [Bujacz et Strumillo, 2006], les auteurs présentent une autre méthode de sonification utilisant des sons MIDI. La scène visuelle est balayée périodiquement (de la même façon que pour le système "The vOICe") de droite à gauche et les obstacles rencontrés sont sonifiés séquentiellement. La distance vers les objets est codée par la note du son (plus l'objet est proche, plus la note est aigue) et la direction est codée en binaural. Plusieurs scans de la scène peuvent être effectués en même temps (avec différentes élévations), différents instruments MIDI sont alors attribués à chacun de ces balayages. Les figures 2.9 a et b représentent le balayage d'un environnement (en a) et la représentation sonore en fonction du

temps (en b).

Une autre approche, basée sur des algorithmes de "pattern matching" (confrontation d'un motif avec l'image dans laquelle il est cherché), consiste à rechercher des objets spécifiques dans la scène visuelle. Un des premiers dispositifs à utiliser cette technique de reconnaissance d'objets à été développé par [Hub *et al.*, 2006]. Ce projet est dédié à l'analyse de scènes en intérieur et à la détection d'objets mobiles (ex. chaises), semi mobiles (ex. porte ouverte ou fermée) ou fixes (ex. bureau), il ne traite cependant pas de la manière de restituer l'information.

2.4.2 Les aides à l'orientation

Il existe différents moyens de guider une personne d'un point A vers un point B. Les systèmes RFID (Radio Frequency IDentification) et RIAS (Remote Infrared Audible Signage) permettent grâce à des balises placées aux feux rouges, aux abris bus et aux points de réorientation de guider un utilisateur tenant à la main un dispositif recevant les signaux émis par les balises. De façon non exhaustive, on peut citer les travaux de [Blenkhorn et Evans, 1997, Amemiya *et al.*, 2004, Willis et Helal, 2005, Gifford *et al.*, 2006, Simonov *et al.*, 2008]. Toutefois ces systèmes nécessitent un pré-équipement relativement lourd de l'environnement urbain (pose de balises RFID ou RIAS) ce qui est difficilement envisageable à l'échelle d'une ville. La grande partie des progrès effectués en matière de navigation globale repose plutôt sur le développement de la technologie de géolocalisation par satellite.

L'idée d'utiliser le GPS (pour Global Positionning System) pour aider à la navigation des personnes non-voyantes a été suggérée par [Collins, 1985] et par [Loomis, 1985]. Depuis, de nombreux projets de recherche [Loomis *et al.*, 1994, Strothotte *et al.*, 1995, Helal *et al.*, 2001, Holland et Morse, 2001, Ran *et al.*, 2004, Wilson *et al.*, 2007] et systèmes commerciaux (Kapten, Loadstone GPS, BrailleNote GPS, Trekker, Blind navigator, Mobile Geo) basés sur le principe de géolocalisation ont été développés.

a) Les systèmes commerciaux

Les systèmes commerciaux permettent à l'utilisateur de connaître sa position, les lieux environnants et de planifier un trajet. Ils peuvent être implémentés sur des systèmes dédiés (Kapten, Trekker) ou sur des logiciels intégrés dans un téléphone portable (Mobile Geo). La plupart ont un coût très élevé (1000 à 2000 euros) et ne reposent souvent que sur une simple adaptation des dispositifs conçus pour les automobilistes. Ils ne sont donc pas fiable pour une utilisation par des piétons, qui ont besoin d'informations d'orientation très précise et sont peu adaptés aux besoins des non-voyants. Ces systèmes ne permettent pas de planifier un itinéraire pour piéton en prenant en compte les spécificités des non-voyants pour la traversée des rues et des carrefours. Ces lacunes sont dues à deux problèmes majeur : il n'existe pas de base de données cartographiques (ou SIG pour Système d'Information Géographique) pour piéton et la précision des systèmes de géolocalisation actuels (donnée comme étant de l'ordre de 5 m 95% du temps mais en générale de l'ordre

de 10 à 20 m en ville) n'est pas suffisante pour guider un piéton de façon précise. Si on ajoute à ces problèmes, le manque d'ergonomie de ces systèmes, dont la transmission des informations est réalisée en synthèse vocale et gêne les utilisateurs, il en résulte qu'ils sont peu diffusés et ne peuvent être utilisés que par des non-voyants ayant déjà un bon niveau de mobilité et n'hésitant pas à se déplacer.

b) Les projets de recherche

Cherchant à diminuer les problèmes des GPS commerciaux, de nombreux projets de recherche fondamentale ont été initiés afin de proposer des solutions de guidage spécifiques aux non-voyants. Pour améliorer la précision de la géolocalisation ces projets proposent en général d'utiliser, en plus du GPS, des gyroscopes, des accéléromètres et des podomètres permettant d'affiner le calcul de la position de l'utilisateur.

De nombreuses études ont été menées sur les désidératas des non-voyants quant aux fonctionnalités d'un système d'aide à la navigation. L'étude de [Strothotte *et al.*, 1995] montre qu'en plus des directions à suivre et de leur position, les utilisateurs aimeraient avoir des informations sur les noms des rues, la position des passages piétons, connaître les magasins et les bâtiments administratif à proximité desquels ils passent et avoir des points de repère tels que les changements de revêtements de sol, les escaliers, etc. Une étude de [Golledge *et al.*, 2004] montre que pour la préparation d'itinéraire, les non-voyants interrogés plébiscitent les informations vocales, contenant des instructions sur la longueur et le nombre des segments composant le trajet, le nombre de points de réorientation qu'il comprend, ainsi que la taille des angles formés par les intersections. Les informations plus détaillées comme les points de repères, obstacles, indices sonores et podo-tactiles présents sur le parcours sont jugées comme utiles mais pas indispensables. Elles doivent pouvoir être en option dans ce type de système. Au niveau de l'interface en entrée, les utilisateurs désirent utiliser des commandes vocales qu'ils jugent plus pratiques que les claviers de téléphones, d'ordinateurs ou de braille, bien que les systèmes de reconnaissance vocale soient parfois peu efficace en environnement bruyant. Selon les utilisateurs interrogés, les instructions doivent être fournies sous forme vocale sur (par ordre de préférence) : un petit haut-parleur monté dans un collier ou fixé sur l'épaule, un casque osseux ou une oreillette. Ils sont, par contre, opposés à l'utilisation des dispositifs synthétiseur de braille (qui sont trop compliqués à utiliser en déplacement) et aux casques stéréophoniques (qui masquent les indices acoustiques naturels).

Au niveau du développement d'interfaces de présentation des informations adaptées aux non-voyants, les projets les plus actifs sont le PGS (Personal Guidance System) développé par l'équipe de Loomis (voir [Loomis *et al.*, 2000] pour une revue de leurs travaux) et le SWAN (System for Wearable Audio Navigation) développé par l'équipe de Walker (voir [Wilson *et al.*, 2007] pour une revue du projet). Ces deux projets ont évalué le potentiel de plusieurs types d'interfaces de guidage basées sur la technologie GPS. Les tests menés consistent généralement pour les participants non-voyants, à suivre un itinéraire constitué de segments et de points de réorientation plus ou moins nombreux pendant que les expérimentateurs mesurent le temps de parcours, la vitesse de

FIGURE 2.10 – Photographie du système PGS porté par un utilisateur non-voyant.

marche et la distance parcourue. À ces mesures sont associés des questionnaires post-expérimentaux concernant la facilité d'utilisation, le confort et le sentiment de sécurité apportés par ces dispositifs.

i/ Le Personal Guidance System

Un certain nombre d'études ont été menées par Loomis et ses collaborateurs pour évaluer l'efficacité d'un système de guidage virtuel sonore. Leur système, le PGS, restitue les informations de guidage en utilisant un casque stéréophonique diffusant des instructions sonores spatialisées en binaural de manière à ce que les utilisateurs perçoivent les sons comme provenant d'un endroit précis de l'environnement. Ce type de signal est traité de façon directe par le cerveau et ne nécessite pas de traitement cognitif supplémentaire contrairement aux signaux de parole. Le système PGS détermine la position de l'utilisateur avec un GPS différentiel [4]. En plus de la balise GPS, l'utilisateur porte un ordinateur (dans un sac sur son dos) contenant un système d'information géographique permettant de le situer sur le trajet et de déterminer la route à suivre, ainsi qu'un capteur d'orientation (boussole) permettant de connaître à tout moment la direction dans laquelle il se dirige. La photographie de la figure 2.10, montre le système PGS porté par un non-voyant.

[Loomis *et al.*, 1998a] comparent le mode de guidage sonore "parole virtuelle" à trois autres modes de guidages plus classiques basés sur la parole (modes "gauche/droite", "degré" et "sans boussole"), dans des itinéraires planifiés sur des zones dégagées du campus de l'université de Santa Barbara (Californie). Dans le mode de guidage sonore "virtuel", la boussole est fixée sur le casque et se trouve donc sur la tête de l'utilisateur. Celui-ci entend une voix (énoncant le numéro correspondant au prochain point) provenant du point de réorientation vers lequel il doit se diriger. L'intensité du son augmente au fur et à mesure qu'il s'en approche. Quand l'utilisateur rentre dans un cercle virtuel d'un rayon de 1.5 m autour du point de réorientation, il entend un son virtuel provenant du point de réorientation suivant. Dans le mode "gauche/droite", la boussole est fixée sur le torse de l'utilisateur et permet de corriger la trajectoire empruntée en envoyant des signaux verbaux par l'intermédiaire du casque ("gauche", "droite" ou "tout droit"). Le mode "degré" est le même

4. Pour ces systèmes, une station au sol, dont la position absolue est connue, calcule en permanence les corrections relatives à appliquer au signal GPS pour faire correspondre la position réelle avec la position calculée.

que le précédent avec une information verbale supplémentaire, concernant l'angle de rotation à effectuer par l'utilisateur pour faire face au prochain point de réorientation ("gauche 80°"). Enfin, le mode "sans boussole" est le même que le mode "degré" mais l'information concernant la direction empruntée par l'utilisateur n'est plus issue des données de la boussole mais de l'extrapolation de deux points de relevé de position successifs par le GPS (ce qui implique que si l'utilisateur arrête de bouger, le système n'est plus capable d'extraire les informations d'orientation).
Les résultats sur les temps de parcours indiquent que le meilleur mode de guidage parmi ceux testés est le mode "virtuel", qui donne un temps de parcours inférieur aux trois autres. De plus, il est celui que les utilisateurs plébiscitent comme celui qu'ils préféreraient utiliser. Le mode de guidage donnant lieu aux temps de parcours les plus longs et aux moins bons jugements subjectifs est le mode "sans boussole". Cette expérimentation souligne la nécessité d'utiliser une boussole pour garder l'information concernant la direction empruntée par l'utilisateur ainsi que l'intérêt que peuvent avoir les sons virtuels pour le guidage piéton.

Une deuxième expérience, effectuée par [Loomis *et al.*, 2005] introduit une nouvelle interface de guidage nommée "interface de pointage haptique" (HPI pour Haptic Pointer Interface), s'inspirant du système RIAS. L'utilisateur a dans la main une boîte reliée à une boussole électronique. Quand la main pointe dans un angle de moins de 10° autour du prochain point de réorientation, l'ordinateur émet un signal auditif, via un haut-parleur fixé sur le torse. Cette expérimentation compare cinq systèmes de guidage : deux basés sur les sons virtuels (utilisation d'un casque avec rendu binaural) et trois basés sur l'HPI (utilisation de haut-parleurs). Le système "parole virtuelle" est le même que dans l'expérimentation précédente à la différence que le système énonce la distance qui sépare l'utilisateur du prochain point de réorientation plutôt que le numéro lui correspondant. Le système "son virtuel" guide l'utilisateur par l'émission d'un bip sonore spatialisé. Le système "son HPI" émet des bips sonores quand la main pointe dans la bonne direction (à moins de 10° du point d'orientation). Le système "parole HPI" donne des indications verbales ("tout droit" quand la main pointe dans la bonne direction, "gauche" ou "droite" quand la main s'écarte de plus de 10° du point d'orientation). Enfin, le mode "pointage corporel" est similaire au mode "son HPI" mais c'est le corps et non la main qui est pris en compte pour le calcul de la direction. La distance au prochain point de réorientation est annoncée verbalement toutes les 8 secondes pour tous les systèmes, sauf la "parole virtuelle" qui donne cette information continuellement.
L'efficacité du guidage sonore virtuel est confirmée par cette expérimentation, puisque le système "parole virtuelle" donne les meilleurs temps de parcours, suivi du système "son virtuel", puis dans l'ordre, "pointage corporel", "parole HPI" et "son HPI". En revanche, les distances parcourues sont comparables, ce qui indique que les systèmes se valent quand il s'agit de suivre une trajectoire. Le gain de temps réalisé avec les systèmes virtuels se situe donc aux points de réorientation. En effet, alors qu'avec un système de pointage (manuel ou corporel) l'utilisateur a besoin d'un certain temps pour trouver la direction du prochain point, celle-ci est immédiatement perceptible avec un système de guidage sonore virtuel. Les jugements subjectifs émis par les participants suggèrent que chacun des systèmes possède ses avantages et inconvénients. Le pointage manuel est jugé plus facile que le pointage corporel mais il a le désavantage d'occuper une main (sachant que la plupart

du temps l'autre est déjà occupée par une canne blanche). Les sons virtuels ont l'avantage d'être rapidement informatifs et de laisser les mains libres, mais l'obligation d'utiliser un casque pour le rendu binaural (alors que les autres systèmes utilisent des haut-parleurs) perturbe la perception des sons en provenance de l'environnement, ce qui est un problème souvent soulevé par les non-voyants. Enfin la présence d'informations verbales est appréciée car elle est informative, bien qu'elle nécessite plus d'attention que la simple émission de bips sonores.

Une autre étude, de [Marston *et al.*, 2006], a pour but de passer d'un test en environnement contrôlé (campus) à des environnements plus quotidiens et structurés (pâté de maison et parc en ville) et de comparer le système "son HPI" au système "son virtuel" amélioré par l'utilisation d'un casque à tubes d'air (qui a la particularité de ne pas bloquer les sons environnementaux). Dans ce test, les expérimentateurs laissent la possibilité aux participants de choisir quand activer l'apparition des informations sonores. Les deux systèmes donnent des résultats satisfaisants dans les deux types d'environnement, avec toutefois un avantage non significatif en faveur du système de sons virtuels. Ils constatent que dans l'environnement du pâté de maison, qui est structuré par les routes et les trottoirs, les utilisateurs font moins appel aux informations de guidage qu'à l'intérieur du parc, qui constitue un environnement plus dénudé. Les participants évaluent très favorablement le rendu sonore des casques à tube d'air, ils les jugent comme "ne bloquant pas les sons extérieurs" (moyenne de 4,5 sur une échelle de likert (de 1 à 5)) et considèrent qu'un système de guidage commercialisé devrait offrir ce genre de rendu sonore (4,6 sur 5).

Dans un souci d'étudier l'allocation de ressources attentionnelles, [Klatzky *et al.*, 2006] ont mis en lumière ce qui semble être un avantage considérable pour les sons virtuels par rapport au langage. Alors que dans des conditions de navigation "normales" (la seule tâche étant de se rendre d'un point A à un point B), un système d'informations verbales ("gauche, droite, tout droit") révèle des performances égales à celles produites avec un système de sons spatialisés ; ce dernier s'avère supérieur au verbal quand la navigation est parasitée par une tâche distractrice (ici, une tâche N-back : signaler dans une liste défilante, l'apparition à N intervalles, d'un même item, les valeurs de N pouvant varier pour augmenter ou diminuer la charge cognitive). En effet, les temps et les distances de parcours sont, dans ce cas, plus courts avec les sons spatialisés, et les résultats à la tâche de comptage sont meilleurs. Il apparaît donc que les sons spatialisés sont traités à un niveau perceptif alors que le langage nécessite une médiation cognitive supplémentaire. L'utilisation de sons non verbaux permet donc de libérer une certaine charge de travail qui va pouvoir être allouée à une autre tâche, comme par exemple parler avec une autre personne pendant le trajet, garder en tête la liste des courses ou préparer un rendez-vous.

ii/ Le SWAN

Le SWAN est un projet développé par le département de psychologie du Sonification Lab du Georgia Institute of Technologie et porté par l'équipe de Bruce Walker. Basé sur la combinaison d'un GPS avec des capteurs inertiels, boussoles, podomètres et autres capteurs (l'architecture de ce système est présentée figure 2.11), ce système vise à guider les non-voyants en utilisant de la

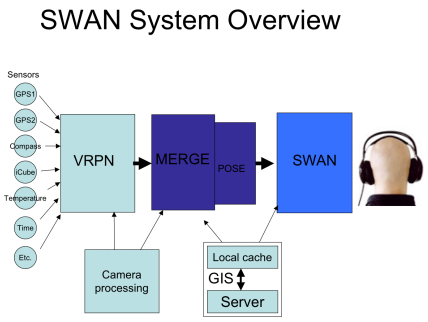


FIGURE 2.11 – Architecture du system SWAN, [Wilson *et al.*, 2007].

sonification spatialisée transmise via un casque stéréophonique à conduction osseuse. Ce système (présenté dans [Wilson *et al.*, 2007]), doit permettre à l'utilisateur de connaître sa position et son orientation, de trouver et suivre un chemin sécurisé pour piéton jusqu'à sa destination tout en étant conscient des principales caractéristiques de son environnement.

Le SWAN utilise une interface de présentation des messages uniquement basée sur l'audio pour guider l'utilisateur sur son trajet (en utilisant des sons appelés "balises sonores"), tout en indiquant la position d'objets ou de bâtiments caractéristiques de l'environnement urbain. L'utilisateur est guidé par un son virtuel répétitif positionné en direction du prochain point de réorientation à atteindre mais à une distance fixe de l'utilisateur (afin de ne pas introduire d'atténuation du son). Lorsque l'utilisateur s'approche de la cible, la fréquence de répétition de la balise sonore s'accélère. Lorsque le point est atteint, un son indique le succès de la tâche puis la balise sonore est placée dans la direction du nouveau point à atteindre. Si l'utilisateur rate un point de passage, le son de celui-ci provient alors de l'arrière avec un changement de timbre de la balise sonore permettant de lever les ambiguïtés avant/arrières. La présentation des informations additionnelles (points d'intérêt, surfaces de transitions, arrêts de bus et autre caractéristiques de l'environnement) stockées dans le SIG est réalisée avec une combinaison d'*auditory icons*, d'*earcons* et de *spearcons* spatialisés. Ceux-ci sont présentés en même temps que la sonification de la trajectoire à suivre.

Une étude expérimentale a été réalisée avec une version virtuelle de ce dispositif pour évaluer les effets dus aux types de sons utilisés pour le guidage (impulsion de type sonar, son pur ou bruit rose), à la taille du cercle virtuel entourant les points de réorientation (0.5, 1.5 ou 15 mètres) et à l'apprentissage (analyse de la progression suite à la réalisation du guidage sur trois chemins différents) [Walker et Lindsay, 2003, Walker *et al.*, 2006]. Pour cette étude, les sujets assis sur une chaise tournante se déplaçaient dans l'environnement virtuel à l'aide d'un joystick et se réorientaient avec la chaise. Les relevés de temps et de distance parcourue permettent de tirer un certain nombre de conclusions. Tout d'abord, cette étude confirme bien le fait qu'un guidage sonore virtuel est efficace pour la navigation, puisque tous les participants ont réussi à effectuer la totalité des trois chemins. Elle montre en plus une amélioration significative des performances avec l'expérience puisque les participants (voyants et sans expérience préalable) vont plus vite et parcourent moins

de distance au fur et à mesure des tests. Les auteurs constatent une amélioration nette entre le premier et le deuxième parcours et une amélioration, moins grande mais néanmoins significative du deuxième au troisième, ce qui sous-entend que le seuil d'apprentissage maximum n'est pas atteint en trois parcours et que les performances peuvent encore s'améliorer. Comme prévu (sur la base du spectre du son), c'est le bruit rose qui donne lieu aux meilleures performances, mais cet effet n'atteint pas le seuil de significativité. Enfin la taille du cercle virtuel autour du point d'orientation a une influence sur le temps et sur la distance parcourue. Les conclusions des auteurs sont qu'un cercle de taille moyenne (environ la taille d'une foulée humaine) est la solution la plus sûre (même si elle n'est pas celle qui mène aux meilleurs temps de parcours). En effet, il est facile de passer à côté du cercle s'il est trop petit, ce qui peut amener l'utilisateur à faire demi-tour et ainsi perdre du temps ; et même si un grand cercle constitue un gain de temps car il s'avère plus facile à trouver, le fait de prendre les virages trop tôt peut amener dans un contexte urbain, à marcher droit vers un mur ou à traverser la rue au mauvais endroit. En conclusion, il semble que les effets d'apprentissage et les effets de la taille du cercle utilisé pour le suivi d'itinéraire ont une plus grande influence sur les performances que le type de son utilisé pour le guidage.

iii/ Une autre modalité de guidage ?

La modalité auditive n'est pas la seule à avoir été testée comme modalité de guidage : [Marston *et al.*, 2007] ont testé l'efficacité de systèmes de guidage binaires, l'un sonore, l'autre tactile (par l'intermédiaire d'un moteur vibrant fixé à la taille) qui envoie une stimulation quand l'utilisateur se dirige dans la bonne direction ("on-course cue") ou au contraire quand il s'en éloigne ("off-course cue"). Ce système de guidage binaire s'avère efficace puisqu'il donne des vitesses de déplacement comparables à celles des systèmes testés précédemment, avec des performances égales pour le système tactile et le système sonore. Il est, de plus, bien accepté par les utilisateurs qui jugent à 4,9 sur 5 qu'un tel type de guidage binaire devrait être disponible en option sur un dispositif commercialisé. Ce guidage binaire ne peut pas être envisagé comme le seul moyen de communication entre le système et l'individu car l'ajout d'informations verbales est parfois incontournable (nom des rues, etc.) mais il a l'avantage d'être très simple et de ne pas recruter de ressources attentionnelles de manière trop importante.

Les stimulations tactiles comme moyen de guidage ont aussi été étudiées par [Henze *et al.*, 2006]. Après avoir présenté leur système d'un point de vue théorique, [Heuten *et al.*, 2008] présentent les résultats d'un test pratique de ce dispositif tactile. Ce dernier se compose d'un module de localisation et de navigation par GPS et d'une ceinture équipée d'une boussole électronique et de six moteurs vibrants qui peuvent être activés un à un ou simultanément avec des intensités de vibration variables pour une plus grande précision. La localisation des utilisateurs et le calcul de l'itinéraire sont pris en charge par le système, ce qui allège la charge cognitive liée au traitement de l'information cartographique. L'utilisation d'un dispositif tactile a le multiple avantage de donner une stimulation directement interprétable (comme les sons virtuels), sans obstruer les canaux auditifs, tout en restant relativement discret.

Les mesures effectuées avec le système concernent la précision de la perception d'une direction ainsi que l'aide apportée en termes de guidage en situation réelle. Les tests indiquent que les participants perçoivent la direction indiquée par le système avec 15° d'erreur en moyenne. La précision est fortement dépendante de la différence d'angle entre l'orientation à percevoir et la position du moteur vibrant. En effet, les six moteurs sont placés autour de la taille à 60° d'écart l'un de l'autre, ce qui mène à une bonne performance quand l'orientation à percevoir est proche de l'un des six moteurs et à une précision moindre quand elle tombe entre deux moteurs (dans ce cas l'intensité de la vibration des deux moteurs adjacents doit être prise en compte, ce qui complique la perception). Les capacités de perception tactile de l'être humain entrent aussi en ligne de compte. Il se trouve que les tests montrent une plus grande précision quand les stimulations à percevoir se situent à l'avant du corps plutôt qu'à l'arrière, ce qui dans une tâche de navigation est une coïncidence heureuse, puisqu'il est très rare de se déplacer à reculons.

L'efficacité de ce dispositif est évaluée avec des individus voyants dans une tâche de suivi d'itinéraire prédéfini, dans un milieu ouvert (sans trottoirs ou murs à suivre ni points de repère particuliers), dans laquelle sont enregistrés la déviation moyenne par rapport à l'itinéraire initial ainsi que le temps total de parcours. Deux parcours différents sont testés, l'un avec des virages angulaires, l'autre présentant de longues courbes. Les vitesses de marche sont respectivement de 3 et 3,3 km/h dans les deux parcours et les déviations moyennes par rapport au chemin prédéfini sont de 6,57 et 7,21 m. Le fait que les stimulations tactiles correspondant au prochain point de réorientation sont données quand l'utilisateur entre dans un cercle de 15 m autour d'un point et le fait qu'il n'y ait aucun repère de type mur ou trottoir dans les parcours testés laissent à penser que cette précision pourrait être améliorée par la diminution du diamètre du cercle et la présence de repères, de manière à être plus satisfaisante pour le guidage d'une personne aveugle. Ces tests montrent toutefois qu'une stimulation uniquement tactile est suffisante pour transmettre les informations du GPS à un piéton de manière intelligible.

Chapitre 3

Individualisation des HRTF : Adaptation auditive

Sommaire

3.1 Introduction

L'objectif de ce chapitre est d'étudier la possibilité d'adaptation du système auditif à la localisation spatiale, avec des HRTF non-individuelles en utilisant un processus audio-kinesthésique dans un environnement auditif virtuel (VAE pour *Virtual Auditory Environment*).

Comme nous l'avons vu dans la section 2.2, l'utilisation d'HRTF non-individuelles pour la synthèse binaurale induit de nombreux artefacts qui nuisent à la localisation spatiale et donc à l'utilisation de sons 3D virtuels dans des applications telles que celles développées dans le cadre du projet NAVIG. Afin de palier à ce problème, nous proposons une nouvelle méthode d'individualisation qui consiste à forcer l'adaptation de l'auditeur à des indices de localisation qui ne sont pas les siens. Dans la section 3.2, nous rappellerons les artefacts induits par l'utilisation d'HRTF non-individuelles ainsi que les méthodes d'individualisation les plus connues que l'on peut rencontrer dans la littérature. Le concept ainsi que la mise en place d'un dispositif permettant l'adaptation du système auditif seront décrits dans la section 3.3. Enfin l'expérience perceptive mise en place pour évaluer les possibilités d'adaptation du système auditif à la synthèse binaurale ainsi que les résultats de cette expérience seront présentés dans les sections 3.4 et 3.5.

3.2 L'individualisation des HRTF

Nous avons vu dans la section 2.2 que le système auditif humain décode la position de sources sonores en se basant sur des indices acoustiques contenus dans les HRTF. Ces indices de localisation peuvent être séparés en indices de disparité binaurale (principalement reliés à la latéralisation des sources sonores) et en indices spectraux (permettant la perception de l'élévation des sources sonores). La procédure permettant de créer des sons virtuels spatialisés à partir de la synthèse binaurale consiste à convoluer le signal sonore à spatialiser avec les HRTF correspondants à la position à simuler puis de présenter les deux signaux résultants sur un casque stéréophonique.

3.2.1 Imperfections de la spatialisation avec des HRTF non-individuelles

Les indices spectraux des HRTF varient fortement en fonction des facteurs morphologiques (tel que la forme de l'oreille externe, du torse, les dimensions de la tête, ...), ils sont donc différents pour chaque personne. L'impact de l'utilisation d'indices d'HRTF non-individuels (mesurés sur une personne différente de celle pour laquelle est destinée la synthèse) pour la création d'un VAE avec la synthèse binaurale a été l'objet de plusieurs études [Wenzel *et al.*, 1993, Wightman et Kistler, 1993]. Ces études évoquent trois types de dégradations liés à l'utilisation d'HRTF non-individualisées : la non externalisation du son, une augmentation des confusions avant/arrière et haut/bas ainsi que des distorsions angulaires dans le plan médian.

Au niveau de l'externalisation du son, l'apparition de dégradations liées à l'utilisation d'HRTF non-individuelles a été mise en évidence par [Kim et Choi, 2005] et [Völk *et al.*, 2008]. Ces dégradations apparaissent principalement sur le plan médian et sont moindres sur les positions latéralisées. En général, la perception de sources générées avec des HRTF non-individuelles est souvent intracrânienne pour des sources positionnées à l'avant ; elle est légèrement externalisée à l'arrière. Ces dégradations dépendent fortement des individus ainsi que de la correspondance entre les HRTF des individus et les HRTF utilisées pour la spatialisation binaurale. La figure 3.1 représente plusieurs

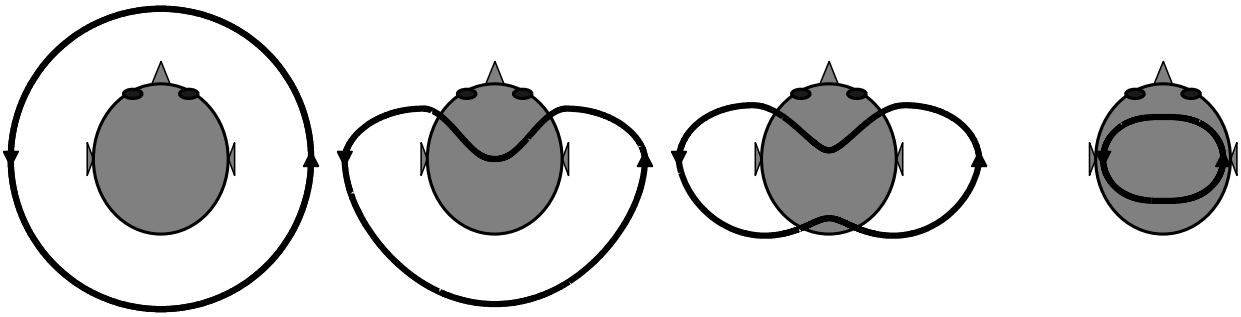


FIGURE 3.1 – Exemples de trajectoires (au centre et à droite) perçues par des sujets à l'écoute d'un son spatialisé effectuant un cercle parfait autour de la tête (à gauche) synthétisé avec des HRTF non-individuelles. Ces trajectoires sont inspirées de discussions avec les utilisateurs de la synthèse binaurale ainsi que des résultats présentés dans [Begault et Wenzel, 1993, Kim et Choi, 2005].

types de trajectoires décrites par des sujets à l'écoute d'une trajectoire circulaire. On remarque que dans certains cas, le son n'est pas perçu à l'avant mais qu'il l'est à l'arrière, alors que pour d'autres, le son n'est externalisé que sur les côtés ou même complètement intracrânien. Ce problème est généralement accentué en élévation.

Au niveau des confusions avant/arrière et haut/bas, les études de [Wenzel *et al.*, 1993] et [Wightman et Kistler, 1993] mettent en évidence une grande augmentation des confusions pour les sujets avec des HRTF non-individuelles (environs 20%). Généralement ces confusions sont de l'avant vers l'arrière (les sources positionnées à l'avant sont perçues à l'arrière) et du bas vers le haut (les sources positionnées vers le bas sont perçues vers le haut).

En terme de localisation, on observe une grande variabilité des résultats selon les sujets. Si pour une minorité, il n'y a pas de dégradation des performances, pour la majorité des sujets, elle est effective. Généralement l'azimut des sources reste bien estimé, alors que l'élévation est nettement moins bien perçue. Les sujets reportent également une perception très diffuse des sources pourtant synthétisées comme ponctuelles [Wenzel *et al.*, 1993].

3.2.2 Individualisation des HRTF : état de l'art

L'impossibilité de mesurer et donc d'utiliser des HRTF individuelles pour chaque utilisateur potentiel dans une situation "grand publique" a engendré de nombreux travaux sur le problème de l'individualisation afin de réduire les artefacts entrainés par la synthèse binaurale. La majeure partie de ces études se focalise sur l'adaptation des HRTF à un utilisateur donné ou sur la fabrication d'HRTF individuelle par modélisation numérique. Bien que nous ayons dans cette étude envisagé le problème de l'individualisation de façon différente, il paraît important de mentionner les différentes méthodes envisagées pour l'individualisation en mettant en avant leurs atouts et leurs faiblesses.

a) Acquisition d'HRTF par modélisation numérique

Cette méthode consiste à résoudre numériquement le problème acoustique de la propagation entre une source et des microphones placés à l'entrée des conduits auditifs. Elle se base sur l'acquisition de la morphologie de la tête de l'auditeur en 3D par un scan laser, une image obtenue par IRM ou l'analyse de clichés photographiques. Les travaux de [Katz, 2001] et [Kahana et Nelson, 2007] ont permis de montrer la faisabilité du calcul précis d'HRTF par des méthodes telles que les éléments finis de frontière (BEM pour Boundary Element Method en anglais). Des maillages réalisés à partir d'un scan 3D permettent de calculer numériquement le champ sonore au niveau des oreilles. Cette méthode a l'avantage de permettre un contrôle fin des paramètres du modèle (morphologie, positionnement des micros, impédances de surface) et d'évaluer directement leur influence sur les HRTF. Son principal inconvénient est la difficulté d'obtenir des maillages individuels. De plus, la puissance et le temps de calcul nécessaires pour la résolution de cette simulation sont limitatifs étant donné que la limite fréquentielle de validité du modèle est directement liée à la finesse du maillage.

b) HRTF non-individuelles issues d'une base de données

Une grande disparité inter-sujet dans les performances de localisation a été mise en évidence par les nombreux tests de localisation menés dans le cadre des études sur la synthèse binaurale. Il ressort de ces travaux l'existence de "bons" et de "mauvais" localisateurs. L'analyse fine de leurs HRTF met en évidence une dépendance spatiale particulièrement prononcée chez les "bons localisateurs" [Wenzel *et al.*, 1988]. Une première hypothèse communément admise consiste à dire que l'utilisation des HRTF d'un "bon localisateur" est un bon choix pour réduire les artefacts de la synthèse non-individuelle. Malheureusement, il paraît difficile (voir impossible) de trouver cette HRTF "universelle". Plutôt que de chercher une seule HRTF qui corresponde à tout le monde, plusieurs études proposent de permettre aux individus de sélectionner les HRTF qui leurs conviennent le mieux dans une base de données (constituée par la mesure des HRTF d'un grand nombre de sujets, dont les morphologies sont représentatives d'une grande partie de la population). Ces techniques se basent sur des tests psychoacoustiques consistant à juger perceptivement la trajectoire d'un son tournant sur le plan horizontal et/ou vertical [Seeber et Fasti, 2003, Iwaya, 2006]. Afin de réduire le nombre d'HRTF à juger, Brian FG Katz, a mis en place un test permettant de réduire le nombre d'HRTF d'une base de donnée. Travaillant sur la base [LISTEN, 2003], il a fait écouter à tous les sujets de la base de données des "tours de tête" sonores pour tous les jeux d'HRTF disponibles. Ces "tours de tête" étaient des fichiers-sons présentant des salves de bruit blanc spatialisées avec le jeu de HRTF de la tête en question. La trajectoire était constituée de deux tours sur le plan horizontal et d'un aller retour sur l'arc médian. Chaque sujet a jugé toutes les HRTF de la base par rapport à sa perception de ces trajectoires ("mauvais", "pas mal" et "excellent"). Les résultats de cette évaluation sont représentés figure 3.2. Une analyse itérative a en ensuite été réalisée pour trouver le minimum de jeux d'HRTF permettant de satisfaire les sujets du test. Cette étude a permis de sélectionner sept jeux d'HRTF (dans les 45 de la base [LISTEN, 2003]) permettant de satisfaire

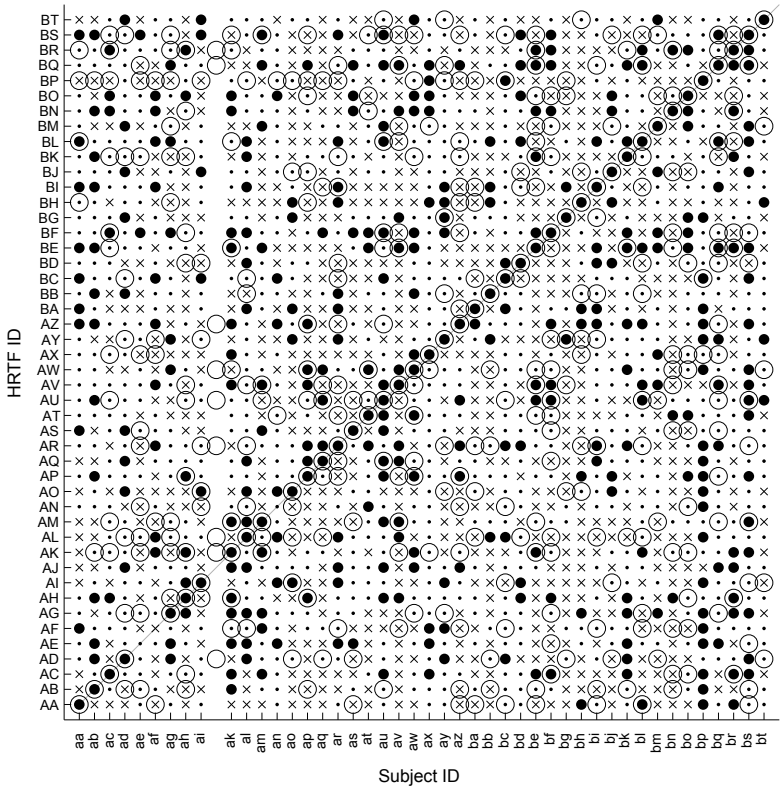


FIGURE 3.2 – Résultats de l'évaluation perceptive de la base Listen par les sujets de la base Listen. (×) mauvais ; (·) pas mal ; (•) excellent ; (◯) transposé d'excellent.

tout le monde. Les performances obtenues par cette méthode ont été évaluées perceptivement lors de l'expérience présentée section 3.4. Le processus de réduction de la base d'HRTF ainsi que la validation perceptive sont présentés dans [Katz et Parseihian, 2012]. Cette technique ne correspond pas vraiment à une individualisation des HRTF, elle permet néanmoins de trouver les HRTF les plus proches des sujets et ainsi de réduire les artefacts liés à la synthèse non-individualisée.

c) Transformation de HRTF non-individuelles

Cette méthode proposée dans [Middlebrooks, 1999a, Middlebrooks, 1999b] et [Middlebrooks *et al.*, 2000] consiste à transformer de façon contrôlée les HRTF par scaling fréquentiel. Elle repose sur une idée simple : les cavités du pavillon sont vues comme des résonateurs en parallèle, si les dimensions des cavités sont modifiées d'un certain pourcentage, les caractéristiques des résonances seront modifiées d'un même facteur. Cela se traduit par une homothétie du profil spectral, donc un décalage des creux et pics spectraux. Middlebrooks fait l'hypothèse que pour tout couple de sujets, il est possible de réduire les différences entre leur HRTF pour les deux oreilles et toutes les directions, avec un seul et unique facteur d'homothétie, appelé facteur de scaling. Si par des tests de localisation, l'auteur montre une amélioration appréciable

des performances de la majorité des sujets, de nombreux artefacts subsistent principalement à cause du fait que les différences inter individuelles ne sont pas seulement liées à des différences au niveau de la taille des pavillons. Le principal inconvénient de cette méthode est que le calcul du facteur d'homothétie nécessite la connaissance des HRTF de l'individu. Elle ne permet donc pas d'éviter la session de mesure. Elle peut par contre permettre de réduire le nombre de mesures (le facteur d'homothétie étant unique à tout le jeu d'HRTF).

d) *Tuning* du spectre des HRTF

Le *tuning* d'HRTF consiste à amplifier ou atténuer le spectre d'amplitude par bande de fréquences afin d'adapter les HRTF à un sujet. Partant de l'idée que la qualité des indices spectraux réside dans leurs fortes variations spatiales, ainsi que dans l'existence de résonances et antirésonances marquées, plusieurs études ([Zhang *et al.*, 1998], [Lee *et al.*, 2004] et [Park *et al.*, 2005]) ont exploré la possibilité d'exagérer ces caractéristiques afin que les HRTF contiennent davantage d'indices permettant la discrimination spatiale. Cette technique semble être potentiellement efficace, malheureusement le *tuning* est individuel et le choix des paramètres à ajuster n'est pas évident.

e) Modélisation des HRTF par apprentissage statistique

La modélisation par apprentissage statistique consiste à relier les HRTF aux paramètres morphologiques décrivant l'individu sur lequel elles ont été mesurées. La première phase consiste à réaliser l'apprentissage sur une base de données contenant les mesures des HRTF d'un grand nombre d'individus ainsi que les données anthropométriques d'intérêt. Le modèle ainsi généré, il suffit de lui fournir en entrée les données anthropométriques nécessaires d'un nouvel auditeur, pour obtenir de nouvelles HRTF en sortie. Cette méthode, a été étudiée notamment par [Jin *et al.*, 2000], [Busson, 2006], [Xu *et al.*, 2009] et [Schönstein et Katz, 2010].

3.3 Adaptation rapide aux HRTF en utilisant un environnement virtuel

Nous avons vu dans le chapitre précédent les différents artefacts induits par l'utilisation d'HRTF non-individuelles pour la synthèse binaurale, ainsi que les différentes méthodes envisagées pour individualiser ces HRTF.

La difficulté à trouver une méthode simple, efficace et peu coûteuse en calcul nous pousse à envisager le problème d'un point de vue différent. En effet, bien que la synthèse binaurale effectuée avec des HRTF individuelles soit souvent considérée comme un cas idéal permettant une grande fidélité pour le rendu des scènes sonores, le résultat, même dans les meilleures conditions, produit de nombreux artefacts dus au manque de naturel du VAE. Afin de minimiser ce type d'anomalies et d'habituer l'utilisateur au VAE, les études employant la spatialisation binaurale font souvent mention

de l'utilisation de sessions d'adaptation à l'environnement de test ([Wightman et Kistler, 1989b], [Wightman et Kistler, 1999] et [Algazi *et al.*, 2001a]). Cet effet, appelé "effet d'apprentissage" correspond à une adaptation procédurale qui regroupe les améliorations de performances dues à la familiarisation avec la tâche, le stimulus, la méthode de report de jugement ainsi qu'avec l'environnement expérimental. Il permet après quelques séances d'entraînement d'améliorer les performances de localisation de façon conséquente. Ce constat amène à considérer le problème de l'individualisation dans le sens inverse. Plutôt que de chercher à adapter des HRTF à un individu, est-il possible de forcer le système auditif d'un individu à s'adapter à des HRTF non-individuelles ?

3.3.1 Apprentissage en localisation sonore

a) Construction de la carte audio-spatiale

De la naissance à l'âge adulte, l'anatomie des mammifères évolue, entraînant des modifications morphologiques qui induisent une modification des indices de localisation. Malgré ces modifications, les mammifères restent capables de localiser des sources sonores avec une bonne précision tout au long de leur vie sans jamais percevoir de changement majeur dans leur façon d'entendre les sons. Il existerait donc des mécanismes permettant au système auditif de s'adapter aux modifications des indices acoustiques en calibrant en permanence la carte audio-spatiale.

Plusieurs études ont démontré la capacité du système auditif animal à se calibrer et à s'adapter à des changements majeurs tant du point de vue de la perception des fréquences, que des niveaux d'intensités ou de la localisation sonore [Keuroghlian et Knudsen, 2007]. Le système auditif des mammifères est capable d'utiliser l'expérience sensorielle pour se calibrer durant le plus jeune âge et pour faire des ajustements chez les adultes, ceci afin de maintenir une précision de localisation acceptable [Knudsen, 1984]. La construction de la carte audio-spatiale passerait donc par une collaboration entre le système auditif et les autres systèmes sensoriels.

Si de nombreuses études ont fait l'hypothèse que le système visuel (qui offre des informations particulièrement fiables et précises sur la position des objets) est le sens prépondérant dans la construction de la carte audio-spatiale [King, 2009], certaines études sur les performances de localisation chez les non-voyants montrent que les indices visuels ne sont pas forcément indispensables. [Aytekin *et al.*, 2008] résument les travaux effectués dans ce domaine et concluent que l'intégration proprioceptive de la modification des indices de localisation suite à des mouvements volontaires pourrait suffire à construire une carte audio-spatiale.

b) Études sur la plasticité du système auditif

Le terme de plasticité est très largement utilisé en neurobiologie depuis les années 1970. Il désigne la modification d'une propriété ou d'un état face à une modification de l'environnement (stimulus externe). Le cerveau est ainsi qualifié de "plastique" ou de "malléable", c'est à dire qu'il est capable de se modifier par l'expérience.

Dans ce qui pourrait être une des premières expériences sur la plasticité cérébrale, [Stratton, 1896] a expérimenté l'effet du port d'un dispositif visuel permettant d'inverser le sens des images arrivant à la rétine. Portant ce dispositif pendant huit jours d'affilé, il montra qu'au bout de 5 jours certains éléments reprenaient un sens normal et qu'à la fin de l'expérience, son cerveau avait réussi à s'adapter partiellement aux changements de sa vision.

De la même façon que pour la vision, la localisation sonore a fait l'objet de plusieurs études d'apprentissage. La majorité de ces études a évalué la capacité du système auditif à s'adapter à des indices de localisation altérés (par la pose de moulage, l'exposition à un milieu subaquatique, ou l'utilisation d'HRTF non-individuelles, ...). Dans ces études la calibration se fait soit de manière passive (donc naturelle) en laissant le sujet interagir avec son environnement ([Young, 1928], [Javer et Schwarz, 1995], [Hofman *et al.*, 1998], [Van Wanrooij et Van Opstal, 2005], [Carlile *et al.*, 2007] et [Savel *et al.*, 2009]), soit de manière active en provoquant l'adaptation par l'utilisation d'un protocole expérimental dédié avec l'usage d'une rétroaction particulière ([Shinn-Cunningham *et al.*, 1998a], [Zahorik *et al.*, 2006] et [Honda *et al.*, 2007]).

i/ Calibration par la vue

La première étude mettant en évidence le phénomène de plasticité du système auditif chez l'homme adulte a été menée par [Young, 1928]. Pour cette étude, qui se présente sous la forme d'un "journal de bord", l'auteur a porté un "pseudo-phone" pendant une période de 18 jours consécutifs. Il rapporte dans son journal l'évolution de ses sensations ainsi que les changements ressentis au niveau de sa perception spatiale des sons. Le "pseudo-phone" (représenté figure 3.3) est un instrument permettant de produire des illusions auditives au niveau de la localisation en inversant les indices acoustiques arrivant aux oreilles. Il est constitué de petits pavillons positionnés aux niveaux des oreilles et reliés à l'oreille opposée par un tube acoustique. Ce "casque" particulier entraine une inversion des indices ITD ainsi qu'une simplification des indices spectraux des HRTF (due à la forme des pavillons acoustiques qui se substituent aux pavillons naturels de l'oreille). Young porta ce dispositif une heure par jour pendant les neuf premiers jours, deux heures pour les six suivants puis continuellement pendant les trois derniers jours. Pendant les premiers jours, il note une dissociation des indices de localisation visuels et auditifs, des inversions gauche-droite, des phénomènes de localisations vagues et parfois doubles ainsi que certains sons non-localisables. Il note un retour à une localisation sonore correcte à partir du moment où la position des sources devient connue (essentiellement lorsque celles-ci entrent dans le champ de vision). Après 18 jours, aucune habituation au dispositif n'est apparue pour les sources en dehors du champ de vision mais une inversion de la localisation dans le champ visuel s'est opérée (les sources provenant de la droite sont bien localisées à droite). Après le retrait du "pseudo-phone", la perception auditive spatiale de l'auteur est immédiatement revenue à la normale et aucune perturbation ultérieure n'a été notée au niveau de la localisation.

Plus récemment une étude menée par [Hofman *et al.*, 1998] a montré que quatre sujets dont la

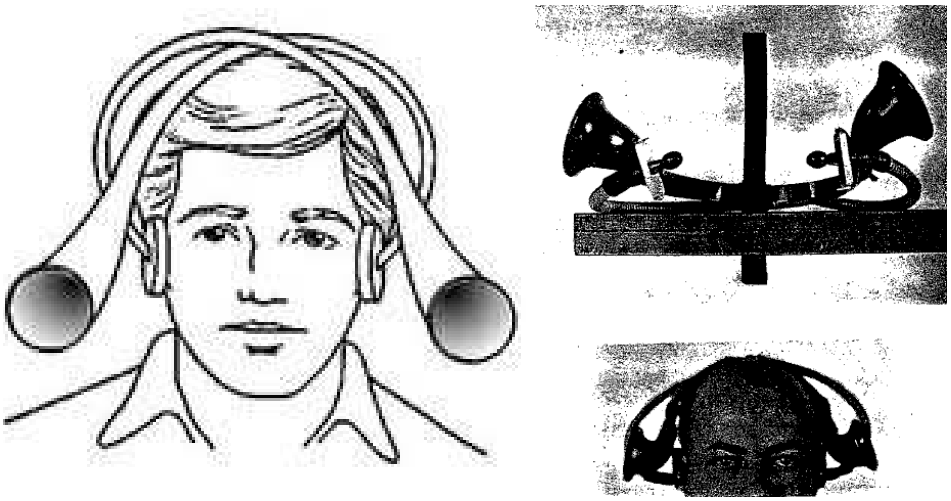


FIGURE 3.3 – Schema (à gauche) et photo (à droite) du pseudophone de [Young, 1928].

forme des pavillons d'oreille avait été modifiée par des moules insérés dans leur pavillons (altérant ainsi les indices spectraux et la localisation en élévation, figure 3.4) ont progressivement acquis de nouveau la capacité à localiser des sons dans le champ visuel après quelques semaines de réadaptation passive. En outre, il est apparu que le système auditif peut conserver la capacité de décoder simultanément plusieurs "sets" d'indices spatiaux (les normaux et ceux modifiés). Ainsi, une fois les inserts permettant de modifier la forme des pavillons enlevés, les sujets ont très rapidement retrouvé leurs repères et récupéré leur capacité habituelle de localisation. Pour conclure cette étude, l'auteur émet l'hypothèse que le système auditif est capable de décoder plusieurs jeux d'indices spatiaux de la même façon que nous sommes capables d'apprendre plusieurs langues et de passer de l'une à l'autre sans difficulté.

Dans une expérience similaire, [Carlile *et al.*, 2007] ont tenté de déterminer si l'adaptation remarquée par [Hofman *et al.*, 1998] pouvait s'étendre à l'extérieur du champ visuel. Dans cette étude, huit sujets ont porté des inserts dans les deux oreilles pendant une période allant de 28 à 62 jours. Au terme de la période d'adaptation, les sujets avaient amélioré leur performance de localisation de 10° en élévation sans toutefois rejoindre leur performance initiale. Comparant cette amélioration entre la zone visuelle et la zone non-visuelle, les auteurs ont montré que l'apprentissage était effectif de la même manière dans toutes les zones. Comme pour l'expérience de [Hofman *et al.*, 1998], le retrait des inserts n'a entraîné qu'un léger post-effet (performances plus faibles comparées aux performances initiales). Un mois après avoir enlevé les inserts, la performance de localisation des sujets a été re-testée avec les inserts. Les résultats ont montré que les sujets avaient gardé la mémoire des indices altérés ; leurs performances étaient égales à celles atteintes au terme de la période d'adaptation, confirmant ainsi l'hypothèse de l'acquisition d'une deuxième carte audio-spatiale.

Dans une étude cross-modale sur neuf sujets, [Zwiers *et al.*, 2003] ont montré qu'une compression spatiale du champ visuel (en utilisant des lentilles de compression d'un facteur 0.5×) entrainait des changements systématiques et adaptatifs au niveau de la localisation sonore permettant de restaurer la fusion des informations visuelles et auditives ; ceci en seulement 2-3 jours d'adaptation.

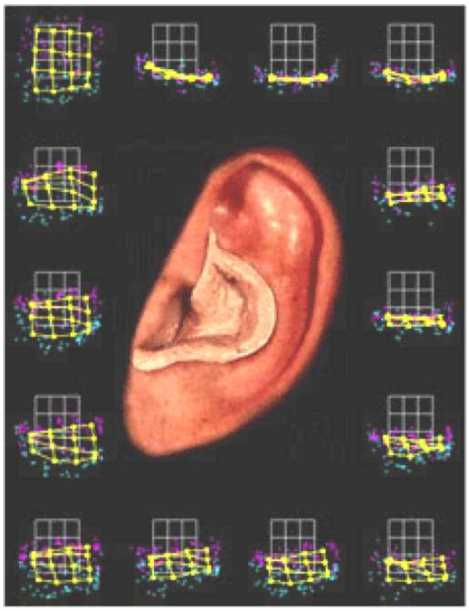


FIGURE 3.4 – Photos des inserts utilisés pour l'expérience de [Hofman *et al.*, 1998].

Les auteurs ont mis en évidence une compression de la localisation en azimut pour la région de l'espace correspondant au nouveau champ visuel et une distorsion de la perception en dehors de celui-ci.

Ces différentes études mettent en évidence un mécanisme de recalibration passive du système auditif chez l'humain adulte principalement dans le champ visuel. Elles montrent qu'il est possible de forcer l'adaptation auditive en modifiant artificiellement les indices de localisation et que cette adaptation ne correspond pas vraiment à une adaptation mais plutôt à l'apprentissage d'une nouvelle carte audio-spatiale. Cependant, la limitation de cet apprentissage à la zone visuelle (hormis pour l'expérience de [Carlile *et al.*, 2007]) ne nous permet pas de poser des hypothèses sur le rôle des modalités sensorielles autres que la vision pour la calibration du système auditif.

ii/ Calibration sans la vue

Les études sur les performances de localisation des non-voyants ont permis aux chercheurs d'étudier plus profondément le rôle de la vision ainsi que celui des autres sens pour la calibration du système auditif. Dans une tentative de mettre en évidence le rôle majeur du retour visuel dans le développement du système auditif humain, [Zwiers *et al.*, 2001a] ont montré un déficit de performance de localisation chez les non-voyants dans la région frontale. Toutefois, d'autres études telles que [Lessard *et al.*, 1998] et [Doucet *et al.*, 2005] ont démontré que la vision n'est pas forcement essentielle et que d'autres modalités sensorielles tel que la somesthésie ou la proprioception peuvent être suffisantes pour le développement de la localisation des sons en azimut et en élévation.

Il est assez difficile de trouver un consensus dans la littérature sur les performances de localisation des non-voyants. Contrairement aux résultats de [Zwiers *et al.*, 2001a], les résultats de [Lessard *et al.*, 1998] montrent des performances équivalentes entre les voyants et les non-voyants

dans le plan horizontal. Ces divergences de résultats pourraient être expliquées par une différence de point de vue, non pas entre les chercheurs, mais entre les voyants et les non-voyants. En effet, dans une étude sur les performances de localisation des non-voyants, [Lewald, 2002] posa l'hypothèse que le déficit de performances des non-voyants pouvait être expliqué par la méthode de report de la position des sources perçues. Ses résultats indiquent que le point de référence des non-voyants pour indiquer la position d'une source n'est pas forcement comme pour les voyants le centre de la tête. Cette hypothèse est confirmée par [Zwiers *et al.*, 2001b] qui en modifiant le point de référence utilisé par les non-voyants pour la tâche de pointage trouvent des performances égales entre les voyants et les non-voyants. Pour indiquer une direction, les non-voyants ne créeraient donc pas un vecteur entre le centre de leur tête et le bout de leur doigt mais plutôt entre leur épaule et leur doigt.

Ces résultats tendent à montrer que la vision n'est pas le seul sens permettant la calibration du système auditif et que d'autres modalités sensorielles peuvent jouer un rôle majeur dans les mécanismes d'adaptation du système auditif spatial. Malheureusement, il n'existe pas ou peu d'études sur la recalibration d'un système auditif modifié chez les non-voyants et il est donc difficile de savoir si l'apprentissage d'une deuxième carte audio-spatiale serait possible sans la vue. Certaines études réalisées dans des environnements virtuels ont cependant permis d'en savoir un peu plus sur le rôle de la proprioception dans la calibration du système auditif.

iii/ Adaptation du système auditif dans un environnement virtuel

Dans le contexte des VAE, l'utilisation de sons 3D synthétisés avec des HRTF non-individuelles peut être considérée comme la situation "d'écouter avec les oreilles de quelqu'un d'autre". En effet, comme nous l'avons vu dans la section 3.2, l'utilisation d'HRTF non-individuelles (avec des indices spectraux distordus et des indices binauraux non adaptés à ceux de l'utilisateur) produit le même effet que les modifications physiques du pavillon de l'oreille effectuées par Hofman dans son étude mais dans une situation virtuelle. La différence majeure entre ces deux situations est que les inserts entrainant des modifications physiques du pavillon peuvent être gardés plusieurs jours en continuité alors qu'il n'est pas possible d'utiliser continuellement un VAE pendant plusieurs jours. Il est donc nécessaire de mettre en place des protocoles expérimentaux permettant de provoquer l'adaptation. Inversement, alors que pour les modifications réelles on ne peut que réduire le conduit acoustique ou agrandir la taille du pavillon, l'utilisation d'un VAE permet de nombreux types de modifications non accessibles autrement. Il est par exemple possible (comme nous le verrons par la suite) de séparer les indices ITD et les indices spectraux des HRTF afin de n'individualiser qu'une partie des HRTF.

Deux études de Shinn-Cunningham ([Shinn-Cunningham *et al.*, 1998a, Shinn-Cunningham *et al.*, 1998b]) utilisant un VAE avec un retour visuel ont montré qu'avec 8 sessions de deux heures, les sujets étaient capables de s'adapter à des indices de localisation binauraux (ITD et ILD) virtuellement distordus. C'est indices "supernormaux" étaient créés en élargissant les indices d'un jeu d'HRTF dans la zone frontale et en réduisant ceux correspondant

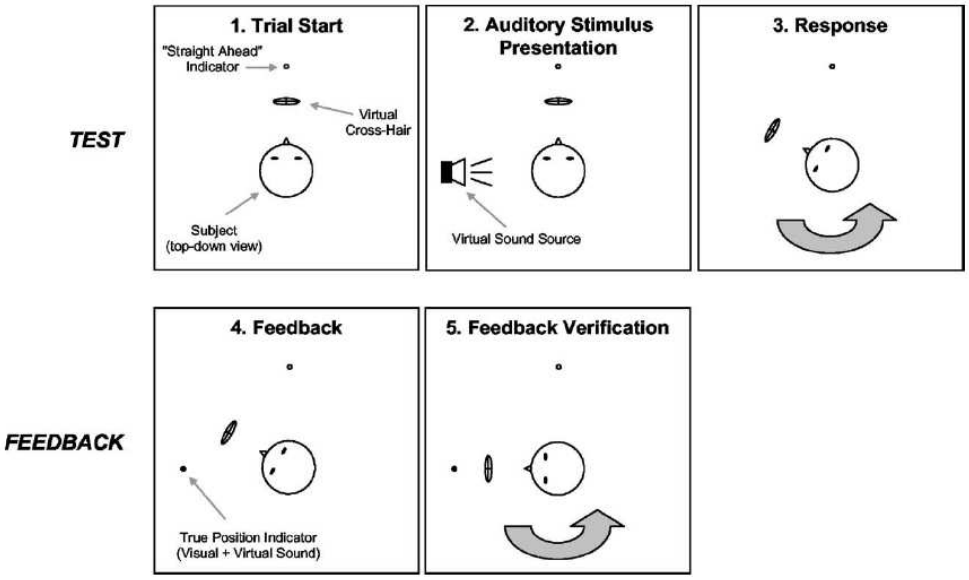


FIGURE 3.5 – Protocole d'adaptation avec retour visuel utilisé dans l'expérience de [Zahorik *et al.*, 2006].

aux zones périphériques. Avec cette transformation, une source sonore positionnée à l'azimut θ était synthétisée en utilisant les HRTF correspondant normalement à la position $f(\theta)$. Cette distorsion ayant pour effet d'augmenter l'angle minimum audible dans la région frontale et de le diminuer sur les côtés. Les auteurs ont comparé deux types d'entrainement à ces HRTF modifiées : une tâche motrice active et une tâche passive. Dans les deux cas, la recalibration de la carte audio-spatiale était favorisée par une information visuelle (point lumineux) indiquant la position réelle de la cible. Les auteurs n'ont pas trouvé de différences statistiques entre les deux types de tâches qui ont chacune permis une adaptation significative aux indices altérés, montrant ainsi que le mécanisme d'adaptation pouvait aussi être effectif dans un environnement virtuel.

Dans une étude avec des HRTF non personnalisées, [Zahorik *et al.*, 2006] met en évidence une adaptation rapide (deux sessions de 30 minutes) aux indices spectraux en utilisant un retour visuel. Les sujets (12 au total) étaient placés dans un dispositif d'adaptation audio-visuel immersif composé d'un casque visuel (Head Mounted Display) et d'un casque audio équipé d'un système de tracking. La tâche d'apprentissage était similaire à un test de localisation où les sujets devaient pointer leur nez en direction de la position perçue de la source qui n'était présentée qu'une seule fois. Un retour visuel était ensuite fourni sous la forme d'un point lumineux virtuel produit en même temps que la répétition de la source sonore à la position de la cible. Il était ensuite demandé à l'utilisateur de pointer son nez en direction du retour visuel avant de passer au stimulus suivant. Ce protocole est illustré figure 3.5. Les participants ont effectué cette tâche deux fois en quatre jours. Les résultats ont mis en évidence une amélioration de la discrimination des sources avant/arrière avec un taux de confusion diminuant de 38% à 23%, ils ne montrent par contre aucune amélioration de la perception en élévation. Un test de localisation avec les HRTF non-individuelles a été effectué quatre mois après l'expérience. Les résultats ont montré que l'amélioration acquise pendant les

tâches d'adaptation a été retenue et que les performances des sujets étaient restées les mêmes qu'après les deux sessions d'apprentissage, ceci permettant une fois de plus d'alimenter l'hypothèse de l'apprentissage d'une deuxième carte audio-spatiale.

Combinant apprentissage avec des sons virtuels et réels, [Honda *et al.*, 2007] ont exploré l'effet d'un jeu avec des sons virtuels sur la performance de localisation de sons réels. Le jeu consistait à chasser des abeilles virtuelles en tapant avec un marteau équipé d'un capteur dans la direction du son de l'abeille. Les participants, équipés d'un casque audio avec un capteur de position, avaient la possibilité d'évoluer dans ce VAE généré en fonction de leur mouvement. L'évaluation des performances de localisation de sons réels a été réalisée avec un système de 36 haut-parleurs positionnés tout les 30° aux élévations 0° et ±30° sur un cercle de rayon 1.2 m. Les réponses des sujets étaient interprétées par un opérateur qui notait le point le plus proche sur une grille. Cette étude a montré qu'avec sept sessions de 30 minutes de cet apprentissage perceptuomoteur utilisant seulement un retour audio, les sujets sont capables d'améliorer significativement leur performance de localisation de sons réels. Ces résultats sont à modérer étant donné que la méthode de report des résultats semble peu précise. Ils montrent cependant qu'en jouant avec la synthèse binaurale, les sujets découvrent leur espace auditif et y deviennent plus sensible.

3.3.2 Construction d'un VAE permettant une adaptation audio-spatiale

Cette étude est la suite d'une étude préliminaire menée par [Blum *et al.*, 2004] sur 10 sujets. Elle vise à mettre en place puis évaluer un système favorisant l'adaptation forcée du système auditif à des HRTF non-individuelles en permettant une exploration active de la carte audio-spatiale. Dans cette partie, en nous basant sur les études présentées dans la section 3.3.1, nous allons détailler les différentes réflexions ayant favorisé la mise en place de la plateforme d'entraînement audio-kinesthésique rapide.

a) Quels indices adapter ?

Nous avons vu dans le paragraphe 3.3.1.b) qu'il est intéressant d'étudier indépendamment la calibration du système auditif en séparant les indices monauraux (composantes spectrales des HRTF) et binauraux (ITD). Étant donné qu'il est assez simple de faire une individualisation grossière des indices ITD en utilisant la circonférence de la tête et des épaules de l'auditeur [Aussal *et al.*, 2012], nous nous focaliserons dans cette étude sur l'adaptation perceptuelle aux composantes spectrales des HRTF. Nous utiliserons pour cela des HRTF hybrides (constituées d'indices ITD individualisés et d'indices spectraux non-individuels) dont la construction est détaillée dans [Katz et Parseihian, 2012].

Dans la section 3.2, nous avons vu que les performances obtenues avec des HRTF non-individuelles dépendent de la similarité entre ces HRTF et les HRTF de l'individu qui les utilise. En effet si ces HRTF sont perceptivement similaires, les dégradations entraînées par la synthèse

non personnalisée sont moindres que si ces HRTF sont éloignées. Afin de tester l'influence de cette similarité, nous allons étudier l'effet d'adaptation à des indices spectraux considérés comme "bons" (les HRTF non-individuelles utilisées pour la synthèse seront perceptivement proches des HRTF du sujet) et nous le comparerons à l'effet d'adaptation à des indices spectraux considérés comme "mauvais" (les HRTF non-individuelles utilisées pour la synthèse seront perceptivement éloignées des HRTF des sujets).

b) Aspect temporel

Les études sur l'apprentissage s'attachent à distinguer deux types d'apprentissage : l'apprentissage perceptif et l'apprentissage procédural.

- L'**apprentissage perceptif** correspond à une amélioration du traitement du stimulus liée à des modifications neurosensorielles. Cet apprentissage conduit à des modifications fonctionnelles, liées à la représentation mentale des stimuli. Il est de bas niveau. Il donne à une entité la capacité intrinsèque d'apprendre, d'exécuter ou de manipuler consciemment certains objets. Il est involontaire et inconscient (acquisition de la faculté de percevoir). Un peintre, par exemple, distingue les nuances des couleurs plus que toute autre personne à force de travailler les couleurs. Les oreilles d'or de la Marine acquièrent, à force d'entraîner leur écoute, la capacité de reconnaitre un sous-marin en se basant seulement sur les variations acoustiques des sons des différents moteurs.
- L'**apprentissage procédural** (ou apprentissage de la tâche) correspond à une amélioration de la performance liée spécifiquement à la pratique de la tâche. Il est dû à une familiarisation à l'environnement expérimental (chambre anéchoïque, chambre sourde, IRM, ...), à la méthode de recueil des réponses (pointage d'une position avec le doigt, la tête ou une interface de report ....), la mémorisation des consignes expérimentales, ainsi qu'à la mise en place par le participant de stratégies permettant d'optimiser le temps de l'expérience. L'apprentissage procédural inclut le phénomène d'automatisation de certaines connaissances ou compétences.

Les apprentissages procéduraux et perceptifs se déroulent sur des évolutions temporelles différentes. Alors que l'apprentissage procédural a lieu sur un temps court, il est établi que l'apprentissage perceptif a lieu sur un temps plus long [Robinson et Summerfield, 1996]. Il existe néanmoins de nombreuses preuves en faveur d'un apprentissage perceptif rapide [Moore *et al.*, 2003] et [Ortiz et Wright, 2009]. Il est donc difficile de distinguer ces deux types d'apprentissage sur la simple base temporelle.

Étant donné que nous utilisons un environnement virtuel, une adaptation passive étalée dans le temps (comme pour l'expérience de [Hofman *et al.*, 1998]) n'est pas envisageable car elle nécessiterait de pouvoir porter un casque et un dispositif permettant de créer un environnement virtuel en continu pendant plusieurs jours. Nous nous baserons donc sur des temporalités similaires à l'expérience de [Zahorik *et al.*, 2006], donc des sessions d'adaptation courtes ne durant que quelques dizaines de minutes.

Afin d'évaluer les effets d'adaptation du système auditif, nous utiliserons un protocole qui consiste à encadrer les tâches d'apprentissage de pre- et de post-tests permettant d'évaluer les performances de localisation des sujets. Les différences de performances entre les tests de localisation permettront de quantifier l'effet d'adaptation. Pour séparer l'apprentissage perceptif de l'apprentissage procédural, nous utiliserons un groupe contrôle qui effectuera les sessions d'adaptation avec ses propres HRTF. Nous considérons pour cela que les sujets contrôles ont des HRTF optimales et que si l'on observe pour ce groupe une amélioration de performances, elle est uniquement due à un apprentissage procédural.

Afin d'explorer le nombre de sessions nécessaires à l'adaptation, nous étalerons notre test sur trois jours. Certains sujets ne feront qu'une seule session d'adaptation le premier jour (les jours suivants servant à tester la pérennité de l'apprentissage), alors que d'autres feront une session par jour (donc trois sessions).

c) Aspect sonore

Dans le chapitre 2, nous avons vu que notre aptitude à localiser des sons pouvait dépendre de la nature de ces sons ainsi que de leur familiarité. Nous possédons plus de repères fréquentiels pour des sons familiers (comme la voix humaine par exemple), ce qui nous permet d'être plus attentifs aux filtrages induits par les indices de localisation et nous mène donc à une meilleure estimation de leur position dans l'espace [Blauert, 1997]. Certains sons peuvent être aussi associés à des positions privilégiées. Par habitude et connaissance, le chant d'un oiseau est plus facilement perçu comme venant d'en haut que des bruits de pas. Afin de ne pas être tributaire d'aspects cognitifs de cet ordre, nous utiliserons des stimuli les plus neutres possibles ; pour permettre une adaptation équivalente dans toutes les composantes du spectre, ceux-ci devront être large bande. Nous choisirons donc des bruits roses et des bruits blancs, qui sont généralement employés dans la plupart des études de localisation.

Afin de contrôler au mieux les paramètres pouvant influencer la perception auditive, nous restreindrons également l'étude à la perception de l'incidence du seul son direct. Nous nous placerons donc dans un contexte de rendu binaural anéchoïque et n'ajouterons aucun effet de salle.

d) Par quels moyens favoriser l'apprentissage ?

Nous avons vu dans la section 3.3.1.b) que le système auditif est capable de s'adapter à des indices de localisation altérés de façon naturelle (comme dans l'expérience de [Hofman *et al.*, 1998]) ou provoquée (comme dans l'expérience de [Zahorik *et al.*, 2006]). Étant donné que l'utilisation d'un VAE empêche la possibilité d'une adaptation naturelle, nous cherchons à développer une plateforme permettant de provoquer l'effet d'adaptation. L'étude de [Zahorik *et al.*, 2006] a montré que l'adaptation est possible à travers un retour visuel par des points lumineux indiquant les positions effectives des stimuli sonores. Cependant, afin de ne pas induire une tâche de mémorisation

consciente du sujet, nous cherchons à éviter la méthode du “feedback” où l'on fournit au sujet un retour, une correction sur sa perception. En effet, avec le “feedback”, il est difficile de savoir si les sujets n'utilisent pas un processus de mémorisation des sons et des positions associées, ce qui mènerait à une amélioration des performances uniquement valable pour les sons proposés.

Nous cherchons donc simplement à stimuler le processus de recalibration auditif en focalisant les sujets sur l'écoute de la sphère auditive des HRTF. Pour cela, nous nous basons sur les retours d'expériences d'utilisateurs de techniques binaurales en VAE : le fait de pouvoir déplacer une source sonore dans l'espace, permet d'avoir une écoute que l'on peut qualifier d'active au cours de laquelle on favorise l'association filtrage-position spatiale car on connait à tout moment la position de la source en même temps que l'on perçoit la modification des indices acoustiques.

Cette manipulation dynamique des sources sonores présente de nombreux avantages. Tout d'abord, elle permet d'explorer rapidement tout l'espace de manière continue, là où les expériences du même type que celles menées par [Zahorik *et al.*, 2005] ne font que le discrétiser. De plus, en termes d'exercice, le pilotage de source sonore a un caractère ludique qui permet d'envisager le protocole de prise en main du système de spatialisation binaurale sous la forme d'un jeu sonore. Enfin, une manipulation interactive de la source sonore en contrôlant sa position avec la main permet de mettre en place une plateforme d'adaptation qui n'est pas dépendante de la vision. Ce dernier point est d'une importance considérable car si les études déjà réalisées ont montré qu'un calibrage du système auditif est possible sans la vue, de nombreuses questions restent en suspend par rapport à la possibilité d'un recalibrage au niveau de l'élévation et de l'arrière. Un calibrage sans la vue aurait de plus l'avantage de pouvoir être utilisée par les non-voyants ce qui présenterait un grand intérêt pour le dispositif du projet NAVIG.

e) Dispositif utilisé

Toutes ces réflexions nous poussent à utiliser un couplage perception/action comme moyen de favoriser l'adaptation rapide du système auditif à des HRTF non-individualisées. Ce couplage se fera par la modalité sensori-motrice étant donné que la modalité visuelle risquerait de perturber notre étude. La phase d'adaptation aux HRTF hybrides se fera les yeux bandés. Ce dispositif doit permettre au sujet de se créer une cartographie audio-kinesthésique de l'espace. Pour cela, il aura la possibilité de manipuler autour de lui une source sonore matérialisée par un objet préhensile. Cela lui permettra d'associer les indices acoustiques de localisation à une information proprioceptive : la position de sa propre main.

3.4 Expérience

3.4.1 Sujets

Le groupe de participants était composé de 24 sujets adultes payés (5 femmes et 19 hommes, âge entre 20 et 60 ans). Aucun d'entre eux n'était au courant de problèmes auditif. Ils ne connaissaient ni le but de l'expérience, ni les positions spatiales proposées. Trois d'entre eux étaient familiers avec les études de localisation et la synthèse binaurale et seulement un était habitué à utiliser ses propres HRTF dans un VAE.

3.4.2 Design et procédure

Le but de cette expérience était d'évaluer l'effet de la procédure d'adaptation à des HRTF non-individuelles sur les performances de localisation sonores. Pour ce faire, nous avons utilisé deux tâches qui ont été répétées plusieurs fois pendant trois jours. La *tâche d'adaptation (A)*, décrite section 3.4.4, conçue comme un jeu sonore où les sujets (yeux bandés) ont le contrôle total d'une source virtuelle spatialisée sur la position de leur main ; cette tâche permet une adaptation rapide aux HRTF non-individuelles en utilisant une interaction naturelle basée sur un couplage perception/action. Un *test de localisation (L)* classique (décrit section 3.4.5), effectué avant et après la *tâche d'adaptation* permet de quantifier l'amélioration de performance.

En plus de quantifier l'amélioration, nous souhaitions évaluer son évolution sur trois jours et son efficacité en fonction du type d'HRTF utilisées (HRTF considérées comme "bonnes" ou "mauvaises" par le sujet). Afin d'évaluer l'effet du type d'HRTF utilisé sur l'adaptation, les sujets ont commencé l'expérience par un *test de classification (C)* consistant à évaluer perceptivement sept jeux d'HRTF et ainsi établir la similarité entre chaque jeu d'HRTF non-individuelles et les HRTF du sujet (voir section 3.4.3). Suite à cela, 20 sujets ont été divisés aléatoirement en deux groupes ; 10 sujets se sont vu attribuer les HRTF non-individuelles qu'ils avaient considérées comme étant les "meilleures" dans le *test de classification* (groupe *good*, G). Pendant toute la suite de l'expérience chacun de ces sujets a donc entendu des sons spatialisés avec le jeu d'HRTF considéré comme étant le plus similaire à ses propres HRTF. Les 10 autres sujets se sont vus attribuer les HRTF les moins similaires à leurs propres HRTF (groupe *bad*, B). Quatre sujets possédant les mesures de leurs propres HRTF ont été inclus dans l'expérience afin de servir de groupe contrôle (C). L'expérience s'est déroulée sur trois jours consécutifs afin d'évaluer l'effet du temps et du nombre de répétitions sur l'adaptation. Pour chacun des deux groupes tests (G et B), la moitié du groupe a effectué une *tâche d'adaptation (A)* seulement le premier jour, alors que l'autre moitié a effectué une *tâche d'adaptation (A)* chaque jour (*i.e.* trois fois). Le groupe contrôle a quant à lui effectué trois sessions d'adaptation (une par jour). Chacun des sujets a effectué un total de quatre *tests de localisation* (L_1, L_2, L_3 et L_4) ; le premier avant la première session d'adaptation afin d'évaluer les performances initiales de chacun des sujets, les trois suivants après chaque session d'adaptation pour les groupes ($C3$, $G3$ et $B3$) ou un par jour pour les groupes ($G1$ et $B1$). Le tableau 3.1 résume la configuration de l'expérience

Type de groupe	Nb	Type d'HRTF	Jour 1	Jour 2	Jour 3
$G1$	5	Bonne non-individuelle	$L1 \rightarrow A \rightarrow L2$	$L3$	$L4$
$G3$	5	Bonne non-individuelle	$L1 \rightarrow A \rightarrow L2$	$A \rightarrow L3$	$A \rightarrow L4$
$B1$	5	Mauvaise non-individuelle	$L1 \rightarrow A \rightarrow L2$	$L3$	$L4$
$B3$	5	Mauvaise non-individuelle	$L1 \rightarrow A \rightarrow L2$	$A \rightarrow L3$	$A \rightarrow L4$
$C3$	4	Individuelle	$L1 \rightarrow A \rightarrow L2$	$A \rightarrow L3$	$A \rightarrow L4$

TABLE 3.1 – Configuration des groupes de participants.

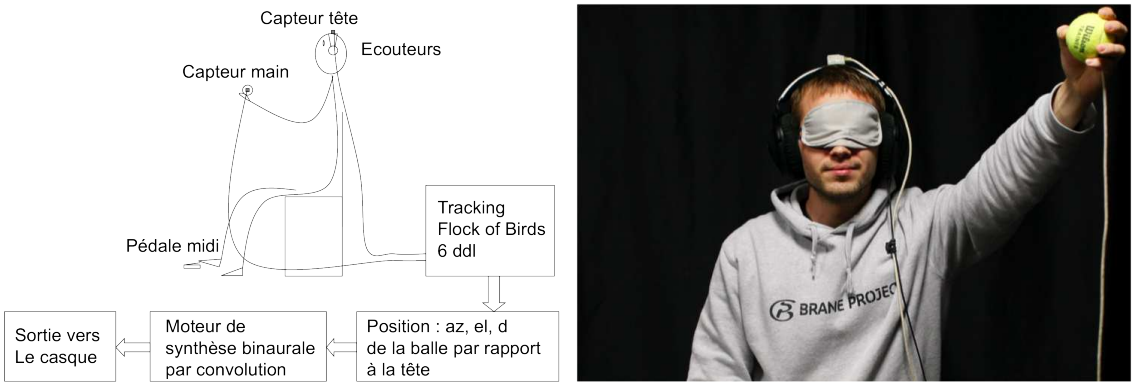


FIGURE 3.6 – Schema de la plateforme audio-kinesthésique d'adaptation aux HRTF non-individuelles, à gauche ; photo du système à droite.

pour chacun des groupes.

Pour la totalité de l'expérience, les sources sonores binaurales ont été synthétisées avec le *LIMSI Spatialisation Engine* [Katz *et al.*, 2011], un moteur de spatialisation temps réel développé avec le logiciel Max/MSP et basé sur la convolution d'HRIR. Les sujets étaient équipés d'un casque stéréophonique ouvert (modèle Sennheiser HD570) suivis avec un capteur magnétique de position et d'orientation 6-DoF (modèle Flock-of-bird) positionné sur le haut du casque. Ils tenaient dans leur main une balle de tennis contenant un capteur de position et interagissaient avec le système en utilisant une pédale midi. Pour le rendu binaural de la *tâche d'adaptation* et du *test de localisation*, la position de la main à chaque instant était calculée par rapport aux coordonnées du capteur positionné sur le casque et décalé au centre de la tête du sujet. Toutes les phases de l'expérience ont eu lieu dans la même chambre calme (niveau de bruit de fond : 35 dBA SPL) avec l'utilisateur assis sur une chaise pivotante, le pied positionné à côté de la pédale midi permettant d'enregistrer les réponses. Afin de présenter une situation comparable au cas réel où les sujets novices utilisent leurs propres écouteurs, aucune égalisation du casque n'a été effectuée.

3.4.3 Classification des HRTF non-individuelles (C)

Les HRTF utilisées comprenaient sept jeux d'HRTF sélectionnés dans les 46 jeux de la base [LISTEN, 2003] mesurés à l'IRCAM. La sélection de ces sept jeux d'HRTF a été faite selon la méthode proposée par [Katz et Parseihian, 2012]. Chaque HRTF est décomposée en composante spectrale (contenant les indices spectraux) et en délai pur (contenant les indices ITD). Ceci permet

de mettre en place des HRTF hybrides dont l'ITD est personnalisé par rapport à la circonférence de la tête du sujet puis combiné à une composante spectrale sélectionnée dans la base des sept HRTF retenues. Les sept jeux d'HRTF ont donc tous les mêmes indices ITD. Pour sélectionner la partie spectrale des HRTF, un test perceptif a été mis en place. Dans ce test, il était demandé au sujet de classer les sept HRTF sur une échelle continue entre "bonne" et "mauvaise" par rapport à leur perception de deux trajectoires sonores (une dans le plan horizontal et une dans le plan médian). La première trajectoire correspondait à un son effectuant deux rotations autour du sujet sur le plan horizontal par pas de 30°. La deuxième suivait un arc de cercle sur le plan médian (azimut = 0°) partant d'une élévation de −45° devant le sujet pour aller à −45° derrière par pas de 15°. Le son partait de devant, allait vers l'arrière en passant par le dessus de la tête puis revenait devant en passant par le même chemin. Le stimulus utilisé était un burst de bruit blanc d'une durée de 0.23 sec fenêtré avec une fonction de Hanning (afin d'éviter les "clics"). La classification résultante permettait pour chaque sujet de déterminer la "meilleure" et la "plus mauvaise" des HRTF non-individuelles c'est à dire celle donnant le meilleur rendu et celle donnant le plus mauvais rendu. Ce classement a permis de sélectionner le jeu d'HRTF à attribuer à chaque sujet.

3.4.4 Tâche d'adaptation (A)

Le concept de la *tâche d'adaptation (A)* était d'immerger le sujet dans un VAE incluant un retour proprioceptif sur les informations sonores spatiales. Afin que le processus d'adaptation s'effectue inconsciemment, cette tâche a été pensée comme un jeu sonore. Elle peut ainsi être effectuée sans que le sujet ne soit au courant du but de l'expérience. Ce jeu peut être vu comme un cache-cache sonore. L'utilisateur doit trouver des sons d'animaux cachés autours de lui en explorant l'espace avec une balle de tennis contenant un capteur de position. La recherche est facilitée par l'ajout d'un retour sonore virtuellement positionné sur la balle tenue par le sujet. Ce retour sonore consiste en une alternance entre un bruit blanc et un bruit rose (avec une largeur spectrale allant de 50 à 20000 Hz) avec un niveau global de 55 dBA mesuré au niveau des oreilles. La fréquence de l'alternance varie en fonction de la distance angulaire entre la main tenant la balle et la position de l'animal à trouver. Ainsi, plus la distance angulaire est petite, plus l'alternance est rapide (suivant la métaphore du compteur Geiger). La durée des sons varie donc entre 3.0 s et 0.2 s, chaque son a un onset/offset de 5 ms permettant de supprimer les "clics" tout en gardant une attaque prononcée (favorisant ainsi une meilleure perception de l'ITD). L'alternance entre bruit blanc et bruit rose a été choisie de façon à avoir un spectre le plus large possible afin de favoriser une adaptation complète à tous les indices spectraux des HRTF. La position virtuelle du son par rapport à la position et à l'orientation de la tête de l'auditeur était maintenue sur la balle grâce à un rendu binaural mis à jour toutes les 50 ms. Ce rafraichissement rapide du rendu audio permet à l'auditeur de bouger sa main ou sa tête tout en ayant l'impression que le son reste positionné sur la balle qu'il tient dans sa main. Il peut ainsi déplacer le son autour de sa tête et explorer sa nouvelle carte audio-spatiale. Une fois la cible trouvée, le son permettant la recherche est remplacé par un son d'animal qui est à son tour "aimanté" sur la position de la balle pendant quelques secondes avant

d'être de nouveau remplacé par le son permettant la recherche. Les sons d'animaux sont tirés de plusieurs bases d'échantillons gratuits ; leur durée moyenne est de 5 ± 2 s. La position des cibles à chercher est sélectionnée de façon aléatoire dans une liste de 50 positions en s'assurant que cela permette à l'auditeur d'explorer toute la sphère. La durée de cette tâche était fixée à 12 min afin que les sujets aient le temps de trouver au moins la moitié des cibles et qu'ils explorent ainsi la totalité de la carte audio-spatiale. Pendant qu'ils effectuaient cette tâche, les sujets avaient les yeux bandés, ils étaient placés sur une chaise pivotante leur permettant ainsi de plus facilement scanner tout l'espace.

3.4.5 Tâche de localisation (L)

Le *test de localisation L* consistait à reporter la position perçue d'un son statique spatialisé en utilisant une technique de report par pointage avec le doigt, validée par une pédale midi. Cette technique de report de positions 3D utilisant la main, a l'avantage d'être écologique, égocentrée et naturelle pour l'utilisateur. Elle est considérée comme étant la meilleure technique de report de position spatiale pour les expériences avec des sujets les yeux bandés [Haber *et al.*, 1993]. Les sujets devaient se tenir dans une position de référence, la tête droite et ne devaient pas bouger pendant la présentation du stimulus.

Le stimulus utilisé était assez court pour exclure les mouvements de la tête pendant sa présentation. Il consistait en un train de trois "burst" gaussiens à bande fréquentielle large (avec une largeur spectrale allant de 50 à 20000 Hz) de 40 ms, fenêtrés par des rampes de Hamming de 2 ms et séparés par des silences de 30 ms. Ce stimulus a été sélectionné sur la base de l'étude de [Dramas *et al.*, 2008a] qui a analysé l'effet de la répétition ainsi que de la durée des burst sur la précision de localisation. Leurs résultats montrent une amélioration de la précision de localisation pour des triples bursts de 40 ms (comparés à un seul burst de 200 ms). Le niveau global mesuré dans l'oreille du stimulus était de 55 dBA.

Après la présentation du stimulus, chaque sujet devait pointer sa main tenant la balle équipée du capteur dans la direction du son perçu et valider sa réponse avec la pédale midi. Pour tenir la balle, aucune main n'était imposée et le sujet avait la possibilité de changer de main en fonction de la position de la source simulée. La position de la source perçue était calculée entre la position et l'orientation de la tête, lors de la présentation du stimulus et la position de la main lors de la validation avec la pédale midi. Aucun retour sonore ou visuel n'était donné au sujet quand à la position de la cible sonore.

Un total de 25 positions régulièrement reparties sur une sphère (voir tableau 3.2) ont été aléatoirement présentées avec pour chacune cinq répétitions. Les sujets ont donc dû localiser un total de 125 cibles dont ils ne connaissaient pas la position. La durée moyenne de ce test était de 10 minutes.

Median		Front		Back	
Latéral	Polaire	Latéral	Polaire	Latéral	Polaire
0	0	30	0	-30	180
0	-30	90	0	30	180
0	30	-90	0	-23	-158
0	90	-30	0	-8	140
0	150	23	22	44	166
0	180	8	-40	11	-140
0	210	-44	-14	-26	106
		-11	41	10	107
		27	74		
		-10	73		

TABLE 3.2 – Angle latéral (θ) et polaire (ϕ) des 25 positions utilisées pour le test de localisation (en coordonnées polaire interaural, voir section d)).

Test	C3	G1	G3	B1	B3
Session d'adaptation 1	28.4 (7.8)	29.2 (5.1)	28.6 (4.6)	24.6 (4.7)	31.0 (7.1)
Session d'adaptation 2	34.6 (5.4)	/	34.8 (4.6)	/	34.6 (5.7)
Session d'adaptation 3	35.2 (5.3)	/	35.4 (7.3)	/	34.6 (7.8)

TABLE 3.3 – Moyenne du nombre d'animaux trouvée par groupe et par session d'adaptation (écart-type entre paranthèses).

3.5 Résultats

3.5.1 Observations générales

a) Tâche d'adaptation

Les résultats de la session d'adaptation ne sont pas des résultats à proprement parler puisque le but de cette tâche n'était pas de localiser précisément des positions dans l'espace mais juste se familiariser avec les indices HRTF dans toute la sphère péripersonnelle. La seule donnée retenue était le nombre d'animaux trouvé pour chaque session. Ce nombre pourrait nous renseigner sur la difficulté à effectuer la tâche en fonction du type d'HRTF ou sur l'évolution de sa difficulté pour les groupes ayant effectué plusieurs sessions d'adaptation. Les résultats du tableau 3.3 ne montrent aucun effet du groupe sur la moyenne d'animaux trouvée ; ils mettent par contre en évidence une évolution pour les groupes G3, B3 et C3 au cours des trois sessions d'adaptations. Cette évolution de l'ordre de 6 animaux trouvés en plus à partir du deuxième test peut-être expliquée par une habituation à la tâche plus que par un effet d'adaptation étant donné qu'elle a lieu de la même façon pour les trois groupes et qu'elle n'apparait pas entre les sessions deux et trois. L'absence de différence entre les groupes peut-être expliquée par les stratégies utilisées par les différents sujets pour effectuer la tâche. En effet, alors que certains ont reporté avoir cherché le plus d'animaux possible d'autres ont reconnu avoir oublié la tâche pour jouer avec le son qui était dans leur main (trouvant en conséquence moins d'animaux que les autres).

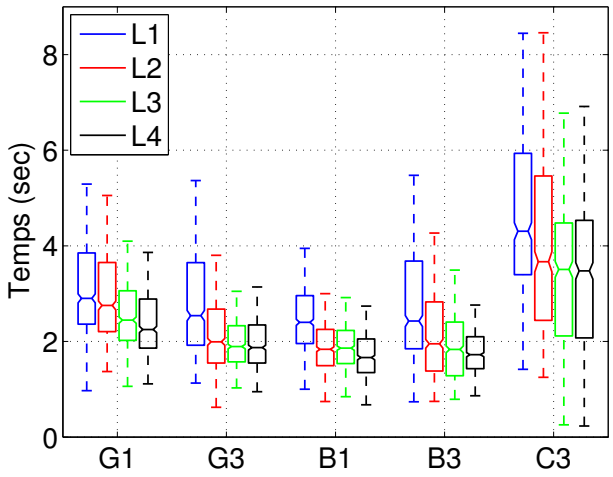


FIGURE 3.7 – Boxplot des temps de réponses par groupes et par test de localisation.

b) Test de localisation : temps de réponse

Le temps de réponse pour localiser une source correspond au temps écoulé entre l'émission du son et la validation de sa position perçue avec la pédale midi. Les boxplots de la figure 3.7 mettent en évidence une grande différence entre le groupe contrôle (dont le temps de réponse est compris entre 0.5 et 8 secondes) et les autres groupes (pour lesquels il est compris entre 1 et 5 secondes). Étonnament, le groupe contrôle a pris en moyenne plus de temps pour effectuer la tâche de localisation que les autres groupes. L'analyse de l'évolution du temps de réponse par groupe en fonction du test met en évidence une diminution du temps de réponse entre L1 et L4 pour chacun des groupes. Cette diminution semble correspondre à un apprentissage procédural. Hormis pour le groupe G1, elle n'apparait quasiment qu'entre les tests L1 et L2. En discutant avec les sujets, il apparait que le temps de réponse n'est pas forcement lié à la faculté d'effectuer le test. Au contraire, le temps plus long passé par le groupe C3 sur la tâche de report, nous laisse penser que les sujets prenaient plus de temps pour répondre lorsque le son était aisement et distinctement localisable, ceci afin de s'assurer de bien reporter cette position précise.

c) Test de localisation : effet de la position

La figure 3.8 met en évidence la disparité des performances en fonction de la position des sources. La représentation utilisée est composée d'un boxplot (à gauche), d'un histogramme (à droite) et de la moyenne de la valeur absolue de l'erreur (point rouge) pour chaque groupe et pour chaque test. Ce type de représentation a l'avantage de combiner les informations statistiques traditionnelles (quartile inférieure (Q1), médiane (Q2) et quartile supérieure (Q3)) contenues dans le boxplot avec la distribution des erreurs représentée par l'histogramme.

Si pour certaines positions, l'erreur est faible aussi bien pour l'azimut que pour l'élévation (comme par exemple les positions latérales gauche et droite), on remarque pour d'autres, de grandes dis-

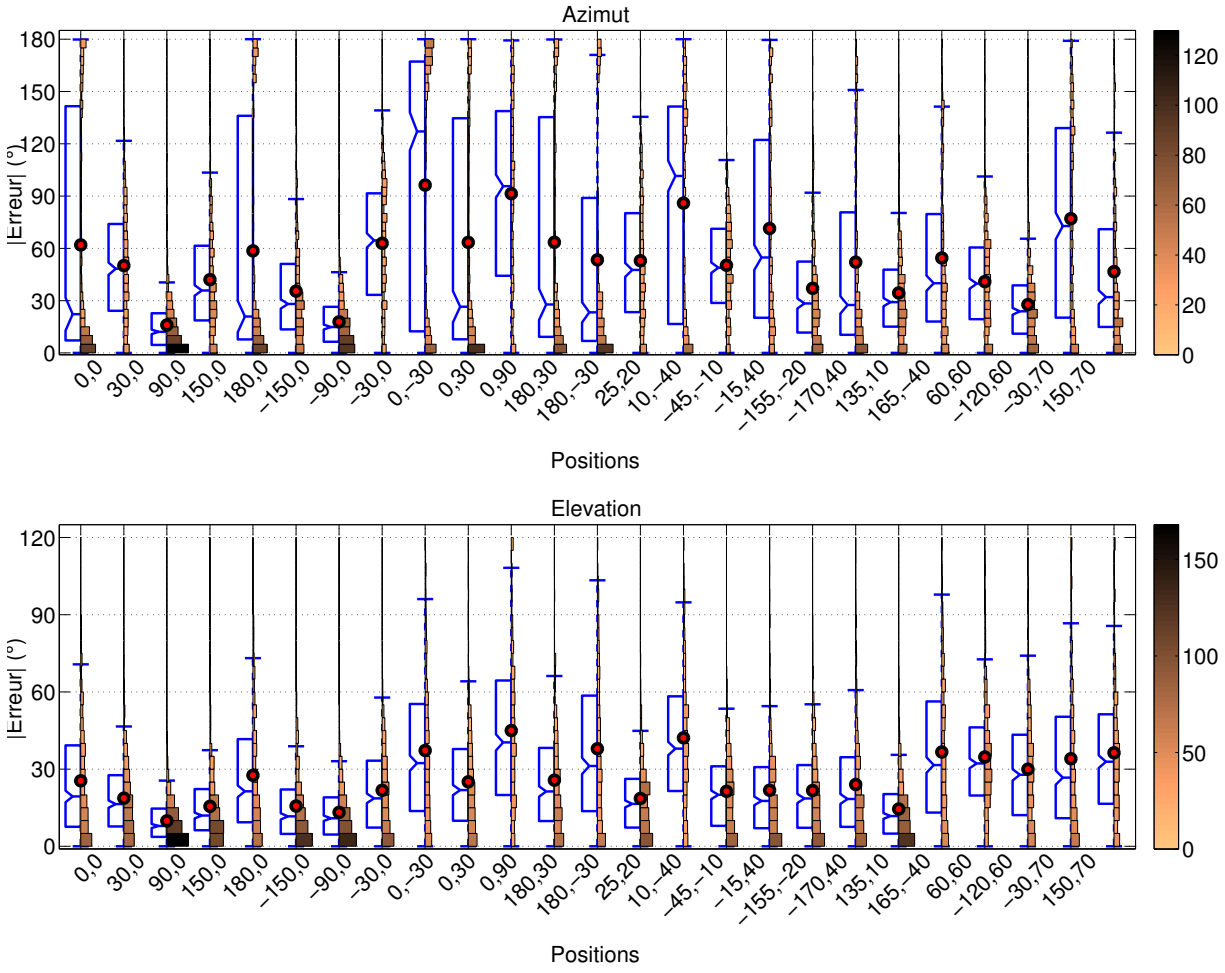


FIGURE 3.8 – Boxplot/Histogramme de la valeur absolue de l'erreur sur l'azimut (en haut) et sur l'élévation (en bas). La valeur moyenne de l'erreur est représentée par un point rouge.

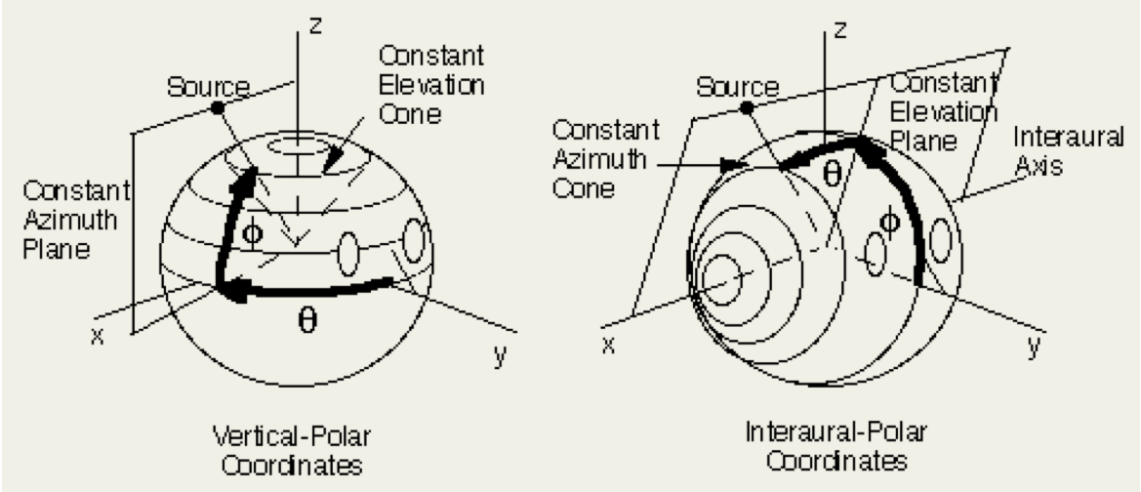


FIGURE 3.9 – Système de coordonnées polaires verticales (classique) à gauche et interaurales à droite (d'après Algazi).

parités mettant en évidence des confusions avant/arrière et des flous de localisation. L'analyse de l'erreur en azimuth montre un grand nombre de confusions avant/arrière pour les sources positionnées sur l'axe médian (azimuth de 0° ou de 180°). Pour ces sources, l'utilisation du boxlot est superflue étant donné que l'erreur est très faible (distribuée autour de 0°) et très élevée en même temps (distribuée autour de 180°). L'analye de l'erreur en élevation met en évidence un flou de localisation pour les sources positionnées en haut (élévation $> 40°$) ou en bas (élevation $< -20°$). En général l'erreur de localisation paraît très élevée pour l'azimut (moyennes comprises entre 15° et 90°) et plus faible pour l'élevation (moyennes comprises entre 15° et 45°).

La répartition bimodale de l'erreur en azimut (due aux confusions avant/arrière) pousse à chercher un autre moyen de représentation des données permettant d'avoir moins de disparités entre les différentes positions.

d) Système de coordonnées et analyse des données

L'analyse des performances de localisation, initialement enregistrées dans le système de coordonnées standard sphérique (azimut et élévation), a été réalisée en utilisant le système de coordonnées polaire interaural introduit par [Morimoto et Aokata, 1984]. Dans ce système (représenté figure 3.9 à droite), les angles d'azimut et d'élévation sont transformés en angle latéral et polaire où l'axe de rotation de l'angle polaire correspond à l'axe interaural. La direction du vecteur entre le centre de la tête et un point sur la sphère s'exprime en fonction de deux angles : l'angle latéral et l'angle polaire. L'angle latéral correspond à l'angle entre le vecteur et le plan médian, il varie entre $-90°$ et 90°. Les sources positionnées sur l'axe médian ont un angle latéral de 0°. L'angle polaire correspond à la rotation autour de l'axe interaural. Il varie entre $-90°$ et 270° et est égal à 0° pour les sources frontales.

Ce système de coordonnées présente un grand avantage pour les études sur la localisation de

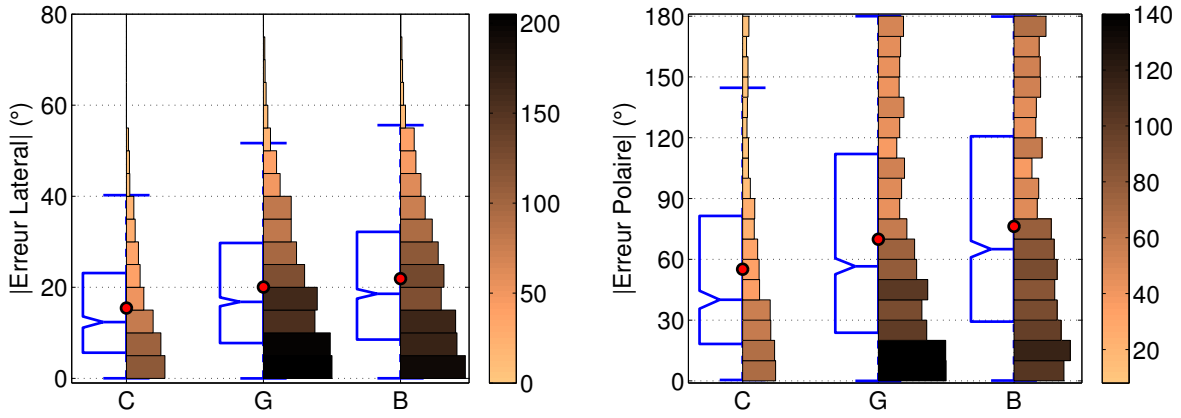


FIGURE 3.10 – Boxplot/Histogramme de la valeur absolue de l'erreur laterale (à gauche) et de l'erreur polaire (à droite).

sources sonores chez les humains étant donné qu'il permet la séparation des indices plutôt temporels, reliés à l'ITD et représentés par l'angle latéral, des indices spectraux, reliés aux HRTF et représentés par l'angle polaire. Avec ce système de coordonnées, toutes les confusions avant/arrière et haut/bas sont contenues dans l'angle polaire. Les erreurs de localisation en angle latéral et polaire ont été analysées en calculant la valeur absolue de la différence entre l'angle de la source perçue et l'angle de la cible.

3.5.2 Différences entre les groupes

Étant donné que pour le premier et le deuxième test de localisation ($L1$ et $L2$) les groupes $G1$ et $G3$ et les groupes $B1$ et $B3$ ne différent pas, il est possible de combiner leurs résultats afin d'analyser les performances initiales de chacun des groupes et d'explorer les différences dues aux types d'HRTF utilisées (non-individuelles "mauvaises", non-individuelles "bonnes" ou "contrôle"). Pour cette section, nous avons donc combiné les résultats des deux groupes avec de bonnes HRTF ($G1$ et $G3$) en un seul groupe G et les résultats des deux groupes avec de mauvaises HRTF ($B1$ et $B3$) en un seul groupe B ($G1 + G3 \rightarrow G$ et $B1 + B3 \rightarrow B$). Nous obtenons deux groupes de 10 sujets et un groupe contrôle de 4 sujets.

a) Erreur sur l'angle latéral

La figure 3.10 à gauche montre la répartition de l'erreur latérale pour chaque groupe pour le premier test de localisation $L1$. Les histogrammes mettent en évidence une distribution normale de l'erreur avec des valeurs moyennes de 15.4° pour le groupe $C3$, de 19.8° pour G et de 21.8° pour B. L'analyse de l'effet du type d'HRTF sur les valeurs moyennes des erreurs latérales de chaque sujet avec un test ANOVA met en évidence une différence significative entre le groupe $C3$ et le groupe G, $[F_{1,12} = 16.01,\ p < 0.005]$; ainsi qu'une différence significative entre les performances

du groupe $C3$ et du groupe B, $[F_{1,12} = 15.75, p < 0.005]$. Aucune différence n'a été trouvée entre les performances des groupes G et B pour le premier test $L1$, $[F_{1,18} = 2.25, p = 0.15]$.

b) Erreur sur l'angle polaire

La figure 3.10 à droite présente la répartition des erreurs polaires obtenues pour le premier test par les groupes $C3$, G et B. Étant donné que l'erreur polaire contient toutes les confusions avant/arrière ainsi que les confusions haut/bas, aucune correction ou suppression de ces erreurs n'a été effectuée afin de pouvoir observer l'effet de leur évolution sur la distribution des réponses. Pour le premier test $L1$, si la distribution de l'erreur sur l'angle polaire est normale pour le groupe contrôle $C3$, elle est multimodale pour les groupes avec des HRTF non-individuelles (G et B) et met en évidence un grand nombre d'erreurs de confusion. Étant donné que les histogrammes présentent des erreurs polaires non uniformément distribuées, les tests statistiques traditionnels tel que l'analyse de la variance (ANOVA) ne peuvent pas être effectués. Pour palier à ce problème, nous avons effectué des tests de Kruskal-Wallis (KW). La statistique du test de Kruskal-Wallis est construite à partir des moyennes des rangs des observations dans les différents échantillons, elle ne travaille pas sur les valeurs des observations, mais sur leurs rangs, il n'est donc pas nécessaire de faire des hypothèses sur la forme des distributions sous-jacentes.

L'erreur polaire moyenne est d'environ 55° pour le groupe contrôle et est comprise entre 70° et 80° pour les autres groupes. Un test KW a été effectué sur la moyenne de l'erreur polaire pour toutes les positions et toutes les répétitions de chaque sujet sur les tests $L1$ et $L2$ afin d'analyser l'effet du type d'HRTF (différences entre les groupes C, G et B). Ce test met en évidence une différence significative entre les groupes $C3$ et G $[\chi^2_{3,24} = 17.98, p < 0.001]$. Une analyse Post-hoc avec un test de Tukey montre que la différence entre les deux groupes est significative pour le test $L1$ ($p < 0.001$) et pour le test $L2$ ($p < 0.005$). Une différence significative est aussi observée entre les groupes $C3$ et B $[\chi^2_{3,24} = 17.02, p < 0.001]$; le test Post-hoc de Tukey montre que la différence entre ces deux groupe est significative pour $L1$ ($p < 0.005$) ainsi que pour $L2$ ($p < 0.005$). Une petite différence est aussi observée lorsqu'on analyse les performances des deux groupes avec des HRTF non-individuelles G et B $[\chi^2_{3,36} = 16.58, p < 0.001]$ avec $p < 0.05$ pour $L1$ et $p < 0.01$ pour $L2$.

Pour résumer, les deux premiers tests permettent de montrer que la distinction entre "bonnes" et "mauvaises" HRTF non-individuelles est bien effective et que le groupe avec des "bonnes" HRTF présente de meilleures performances que le groupes avec des "mauvaises" HRTF. La comparaison avec le groupe contrôle (dont les sujets utilisaient leur propres HRTF) montre que l'utilisation d'HRTF non-individuelles entraîne bien une forte dégradation des performances ainsi qu'une augmentation du nombre de confusions avant/arrière et haut/bas.

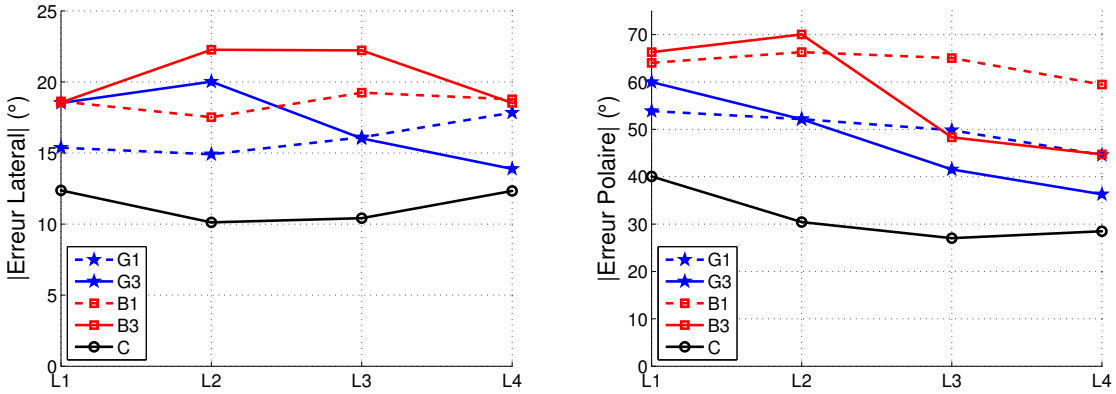


FIGURE 3.11 – Moyenne de la valeur absolue de l'erreur sur l'angle latéral (à gauche) et polaire (à droite) pour chacun des groupes en fonction du test de localisation.

3.5.3 Effets de la tâche d'apprentissage

a) Erreur sur l'angle latéral

La figure 3.11 à gauche présente un résumé des résultats obtenus sur l'erreur latérale. Elle met en évidence l'évolution de la moyenne de l'erreur pour chaque groupe en fonction de chaque test de localisation. Le flou de localisation global est compris entre 13° et 16° pour le groupe *contrôle* et entre 17° et 25° pour les groupes avec des HRTF non-individuelles. Comparé aux autre groupes, le groupe *contrôle* a obtenu de meilleures performances. Ceci peut être expliqué par les approximations du modèle permettant de générer les ITD individualisés (qui a été utilisé pour créer les HRTF non-individuelles hybrides). Étant donné que ce modèle n'est pas parfait, une amélioration des performances latérales après plusieurs sessions d'adaptation est possible. Une légère amélioration a été constatée pour le groupe $G3$, alors qu'aucune amélioration au cours des quatre tests n'a été constatée pour les groupes $G1$, $B1$ et $B3$. Un test ANOVA effectué sur les performances du groupe $G3$ pour les tests $L1$ et $L4$ montre que l'amélioration de 5° observée pour le groupe $G3$ est quasi significative [$F_{1,8} = 4.82$, $p = 0.06$]. Cette amélioration montre une tendance à l'adaptation des sujets aux indices d'ITD non-individuels.

b) Erreur sur l'angle polaire

La figure 3.11 à droite résume les performances obtenues sur l'angle polaire pour chaque groupe et chaque test de localisation. L'analyse de l'évolution de l'erreur pour chacun des groupes au cours des quatre tests de localisation indique une évolution significative des performances pour les groupes ayant effectué trois sessions d'adaptation ($C3$, $G3$ et $B3$) seulement une légère amélioration pour les deux groupes n'ayant effectué qu'une seule session d'adaptation ($G1$ et $B1$).

Pour le groupe *contrôle*, la majeure partie de l'amélioration (10°) a eu lieu après la première session d'adaptation ($L1 \rightarrow L2$). On observe ensuite ($L2 \rightarrow L4$) une amélioration de quelques

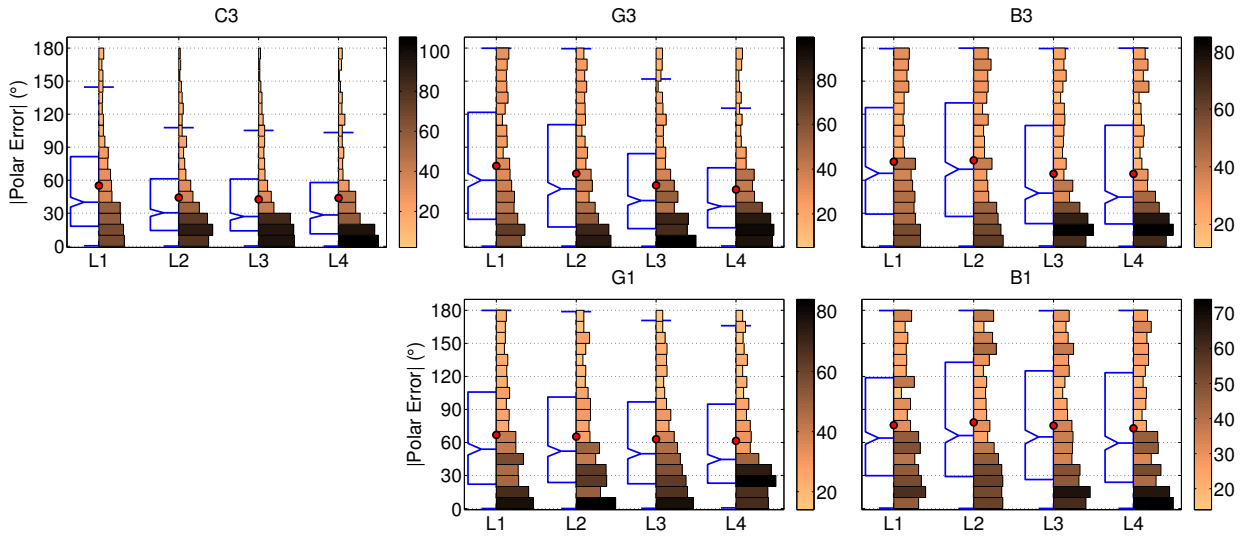


FIGURE 3.12 – Combinaison d'un boxplot et de l'histogramme de la valeur absolue de l'erreur polaire pour chaque test, pour le groupe $C3$ (en haut à gauche), $G3$ (en haut au milieu), $B3$ (en haut à droite), $G1$ (en bas au milieu) and $B1$ (en bas à droite). L'échelle des boxplot est donnée sur l'axe y, le cercle rouge correspond à la valeur moyenne de l'erreur polaire, la légende de l'histogramme (sur la droite) est donnée en nombre d'essais.

degrés. Un test de Kruskal-Wallis montre que cette différence entre $L1$ et $\{L2$ & $L3$ & $L4\}$ est significative $[\chi^2_{3,12} = 8.93,\ p < 0.05]$. L'évolution des performances du groupe $G3$ a été quasi constante pendant toute la durée de l'expérience. L'amélioration des sujets est de 17° pour la moyenne d'erreur (et de 23° pour la médiane). Un test de KW met en évidence une différence significative entre les performances obtenues pour les tests $L1$ et $L4$ $[\chi^2_{3,16} = 11.48,\ p < 0.01]$. Pour le groupe $B3$, l'amélioration est apparue principalement après la deuxième session d'adaptation ($L2 \rightarrow L4$). L'évolution sur l'erreur moyenne est de 15° et l'évolution sur l'erreur médiane de 24°. Un test de KW montre une différence significative entre les tests $\{L1$ & $L2\}$ et le test $L4$ $[\chi^2_{3,16} = 15.25,\ p < 0.005]$.

L'effet des trois sessions d'adaptation est aussi bien visible sur les histogrammes de la distribution des erreurs de la figure 3.12. En effet, si la distribution de l'erreur polaire pour les groupes $G3$ et $B3$ est multimodale pour le test $L1$, elle se transforme petit à petit en distribution quasi normale pour le test $L4$. On peut aussi remarquer que les performances du groupe $G3$ pour le dernier test de localisation $L4$ sont comparables aux performances initiales du groupe contrôle, $C3(L1)$. Un test de KW entre $G3(L4)$ et $C3(L1)$ montre qu'il n'y a quasiment pas de différences entre les résultats de ces deux tests $[\chi^2_{1,7} = 0.96]$. Les performances finales du groupe $B3$ sont quand à elles comparables à celles du groupe $G3(L2)$ après la première session d'adaptation $[\chi^2_{1,8} = 0.32]$.

Pour les groupes n'ayant effectué qu'une seule session d'adaptation ($G1$ et $B1$), une légère amélioration de performances est observée entre le premier et le dernier test de localisation ($L1$ et $L4$). Elle est en moyenne de 8° (9° pour la médiane) pour le groupe $B1$ et de 7° (10° pour la médiane) pour le groupe $G1$. Cette évolution est comparable à l'effet d'apprentissage de la tâche observé sur

	Regression slope				
Test	$C3$	$G1$	$G3$	$B1$	$B3$
L1	0.52 (.22)	0.28 (.23)	0.13 (.10)	0.08 (.03)	0.07 (.13)
L2	0.60 (.24)	0.20 (.22)	0.12 (.17)	0.05 (.04)	0.08 (.11)
L3	0.61 (.24)	0.35 (.21)	0.27 (.22)	0.06 (.06)	0.28 (.20)
L4	0.60 (.29)	0.30 (.18)	0.43 (.12)	0.11 (.13)	0.23 (.18)
	Goodness-of-fit r^2				
L1	0.24 (.14)	0.13 (.16)	0.05 (.06)	0.03 (.02)	0.02 (.03)
L2	0.44 (.35)	0.10 (.13)	0.09 (.14)	0.03 (.05)	0.02 (.02)
L3	0.40 (.27)	0.16 (.18)	0.18 (.17)	0.02 (.03)	0.11 (.17)
L4	0.38 (.30)	0.12 (.11)	0.27 (.14)	0.05 (.09)	0.10 (.11)

TABLE 3.4 – Moyenne des coefficients directeurs des régressions linéaire et goodness-of-fit criteria r^2. Variance donnée entre parenthèse.

le groupe contrôle $C3$. Un test de KW montre une différence significative entre les résultats des tests $L3$ et $L4$ pour le groupe $B1$ [$\chi^2_{3,16} = 11.07$, $p < 0.05$]. Cette différence peut-être expliquée par une concentration des erreurs polaires vers 0° pour le test $L4$. Les performances du groupe $G1$ ne présentent aucune différence significative au court des quatre tests de localisation [$\chi^2_{3,16} = 5.03$, $p = 0.17$].

Une régression linéaire a été effectuée sur les réponses de l'angle polaire. La moyenne et la déviation standard des pentes de régression linéaire et des critères goodness-of-fit r^2 des sujets pour chaque groupe et chaque test sont représentées dans le tableau 3.4. Étant donné qu'aucune correction ou suppression des confusions avant/arrière et haut/bas n'ont été effectuées, les coefficients directeurs des droites de régression sont loin de l'unité normalement attendue pour une localisation idéale. Ces résultats mettent aussi en évidence une grande variabilité entre les performances des sujets avec une déviation standard importante. L'analyse de l'effet des sessions d'adaptation sur les pentes de régression montre une différence d'amélioration entre les deux groupes ayant effectué trois sessions d'adaptation ($G3$ et $B3$) et les autres groupes ($G1$, $B1$ et $C3$). En effet, entre le premier et le dernier test de localisation, la pente de régression augmente d'un facteur 3.27 pour le groupe $G3$ et de 3.13 pour le groupe $B3$ alors qu'elle n'augmente que d'un facteur de 1.15 pour le groupe contrôle $C3$, de 1.05 pour $G1$ et de 1.46 pour $B1$. Les pentes de régressions des sujets du groupe contrôle (comprises entre 0.52 et 0.61) sont plus proches de l'unité que celles des autres groupes (comprises entre 0.07 et 0.43). Ceci met en évidence la différence de perception de l'angle médian entre les sujets disposant d'HRTF individuelles et les sujets avec des HRTF non-individuelles.

Pour résumer, au niveau de l'amélioration des performances de localisation, les deux groupes ayant effectué trois sessions d'adaptation ($G3$ et $B3$) améliorent significativement leurs performances et réduisent leur erreur médiane d'enviros 10° par rapport au groupe contrôle ($C3$). Si l'amélioration du groupe $C3$ doit être attribuée à l'apprentissage de la tâche, le surplus d'amélioration observé chez $G3$ et $B3$ correspond à un apprentissage perceptif. Une petite amélioration de performance a été observée pour les groupes n'ayant effectué qu'une seule session d'adaptation ($G1$ et $B1$). Étant inférieure ou égale à celle observée pour le groupe contrôle, elle doit être considérée

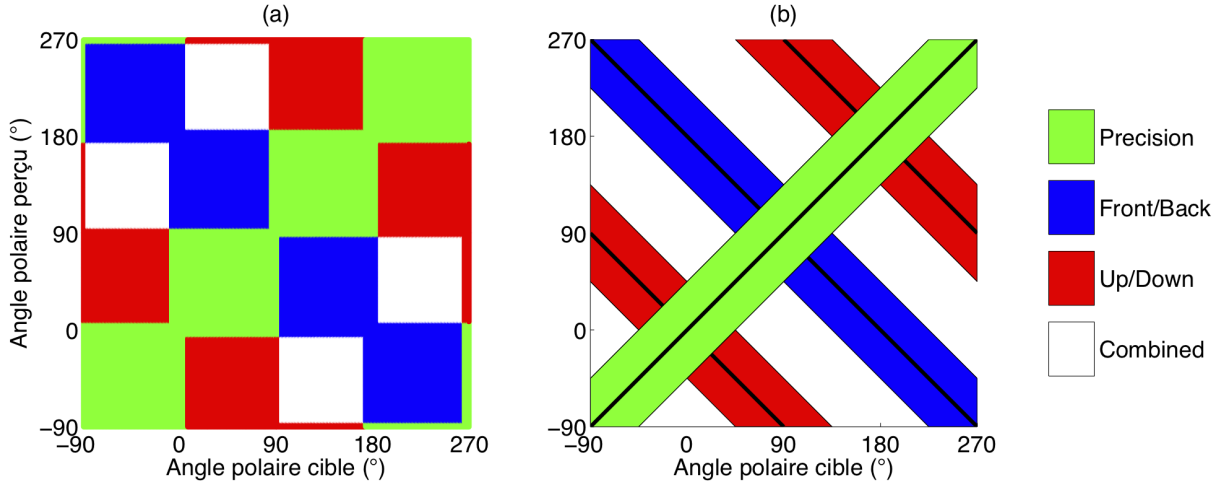


FIGURE 3.13 – (a) Définition des quatre zones de types d'erreurs selon [Martin *et al.*, 2001] ; (b) La définition des zones que nous proposons pour un traitement plus clair des zones limites.

comme une adaptation procédurale au protocole du test. Une seule session d'adaptation n'est donc pas suffisante pour améliorer la localisation avec des HRTF non-individuelles

c) Types d'erreurs de confusions

Certaines études telles que celle de [Zahorik *et al.*, 2006], ont quantifié l'effet d'adaptation en analysant l'évolution des confusions avant/arrière avant et après les sessions d'entrainement. Ce type d'analyse, plutôt que d'observer l'écart à une localisation exacte, permet d'analyser l'évolution des principaux artefacts induits par l'utilisation d'HRTF non-individuelles (i.e. les confusions avant/arrière, les confusions haut/bas ainsi que la non externalisation du son). Étant donné qu'une simple analyse des confusions avant/arrière n'est pas appropriée pour une tâche de localisation sur toute la sphère, une approche de détection des confusions plus détaillée a été mise en place.

Plusieurs méthodes de calcul des confusions de localisation existent dans la littérature. Dans un premiers temps nous allons les rappeler, puis nous introduirons la méthode que nous avons utilisée dans cette étude.

Il n'existe pas vraiment de définition conventionnelle du calcul des confusions en expérience de localisation. Dans la méthode la plus courante (proposée par [Wightman et Kistler, 1989b] et [Wenzel *et al.*, 1991]), si l'angle entre la cible et la position perçue est plus grand que l'angle entre la cible et le symétrique de la position perçue par rapport au plan vertical interaural (ou au plan horizontal), le jugement peut être considéré comme une confusion avant/arrière (ou une confusion haut/bas). Une autre méthode, proposée par [Martin *et al.*, 2001] définit les confusions avant/arrière selon deux conditions. La première est que l'azimut de la cible ou de la position perçue ne tombe pas dans une zone d'exclusion autour de l'intersection entre la sphère et le plan frontal. Cette zone est de 15° (±7.5°) pour une élévation de 0° et égale à 15° divisés par le cosinus de l'élévation ailleurs. La seconde est que la cible et la position perçue ne soient pas dans

le même hémisphère avant/arrière. Cette définition peut être étendue aux confusions haut/bas en prenant comme zone d'exclusion une bande de 15° autour de l'intersection entre la sphère et le plan horizontal et en vérifiant que la cible et la position perçue ne soient pas dans le même hémisphère haut/bas. On peut ainsi définir un espace de réponse divisé en quatre zones :

- une où il n'y a pas de confusions (appelée *precision*),
- une pour les confusions avant/arrière (appelée *front/back*),
- une pour les confusions haut/bas (appelée *up/down*)
- une combinant confusions avant/arrière et haut/bas (appelée *combined*).

Lorsque l'on utilise le système de coordonnées polaires interaurales, toutes ces zones sont contenues dans l'angle polaire. La figure 3.13(a) représente la répartition des zones de ces types d'erreurs en fonction de l'angle de la cible calculée selon la méthode de [Martin *et al.*, 2001]. Avec cette définition, si une cible positionnée à 0° en angle latéral et 8° en angle polaire est perçue à −8° en angle polaire, elle est comptabilisée comme une confusion haut/bas alors qu'elle correspond pleinement au flou de localisation. Pour résoudre ce problème, nous proposons de réorganiser ces zones d'erreurs selon la méthode de [Yamagishi et Ozawa, 2011] qui est étendue ici à tous les types de confusions. La nouvelle répartition est représentée figure 3.13(b). Elle permet d'éviter les mauvais classements de types d'erreurs observés vers les bords des quadrants. Elle est constituée de zones de ±45° autour d'axes d'erreur. La valeur de ±45° a été définie empiriquement, selon la méthode utilisée par [Middlebrooks, 1999b], en examinant l'histogramme de la valeur absolue de l'erreur sur l'angle polaire pour tous les sujets et toutes les configurations. La zone de ±45° autour de l'axe principal ($y = x$ en vert) correspond à l'erreur de *précision*. La zone autour de l'axe $y = -x$ (en bleue) correspond aux erreurs de type *front/back*. Les zones autour des axes $y = 90 - x$ et $y = 450 - x$ (en rouge) correspondent aux erreurs de type *up/down*. En dehors de ces zones se trouvent les erreurs de type *combined* (zone en blanc), c'est à dire celles qui ne peuvent pas être attribuées uniquement à des confusions avant/arrière ou haut/bas. Ce type d'erreurs est très fréquent lorsque la cible n'est pas externalisée. Le sujet entend la cible sonore dans sa tête et pointe plus ou moins au hasard dans une direction. La comparaison des zones des figures 3.13(a) et 3.13(b) met en évidence une sous-estimation des erreurs combinées avec la méthode de [Martin *et al.*, 2001] ainsi qu'une surestimation des autres types d'erreurs. Étant donné que cette étude a pour but de quantifier l'effet des sessions d'adaptation sur l'amélioration des erreurs de précision, la distribution de la figure 3.13(b) semble plus appropriée.

La figure 3.14 représente l'évolution de la réponse en angle polaire pour trois sujets représentatifs des groupes $G3$, $B3$, and $C3$. Les résultats du sujet du groupe $C3$ sont assez précis, avec quelques confusions avant/arrière pour le test $L1$ et quelques erreurs sur le plan médian. Les confusions avant/arrière disparaissent quasi totalement après la première session d'apprentissage alors que les erreurs sur le plan médian subsistent. Les résultats du sujet du groupe $G3$ (avec des "bonnes" HRTF non-individuelles) pour le test $L1$ montrent que ce sujet n'a pas une bonne sensation de l'élévation, les seules sources qu'il semble bien localiser étant les cibles positionnées à 90° en angle polaire (donc au dessus de sa tête). Après les trois sessions d'adaptation, les résultats de ce sujet montrent qu'il a acquis une sensation d'élévation à l'arrière et en bas mais que de nombreuses

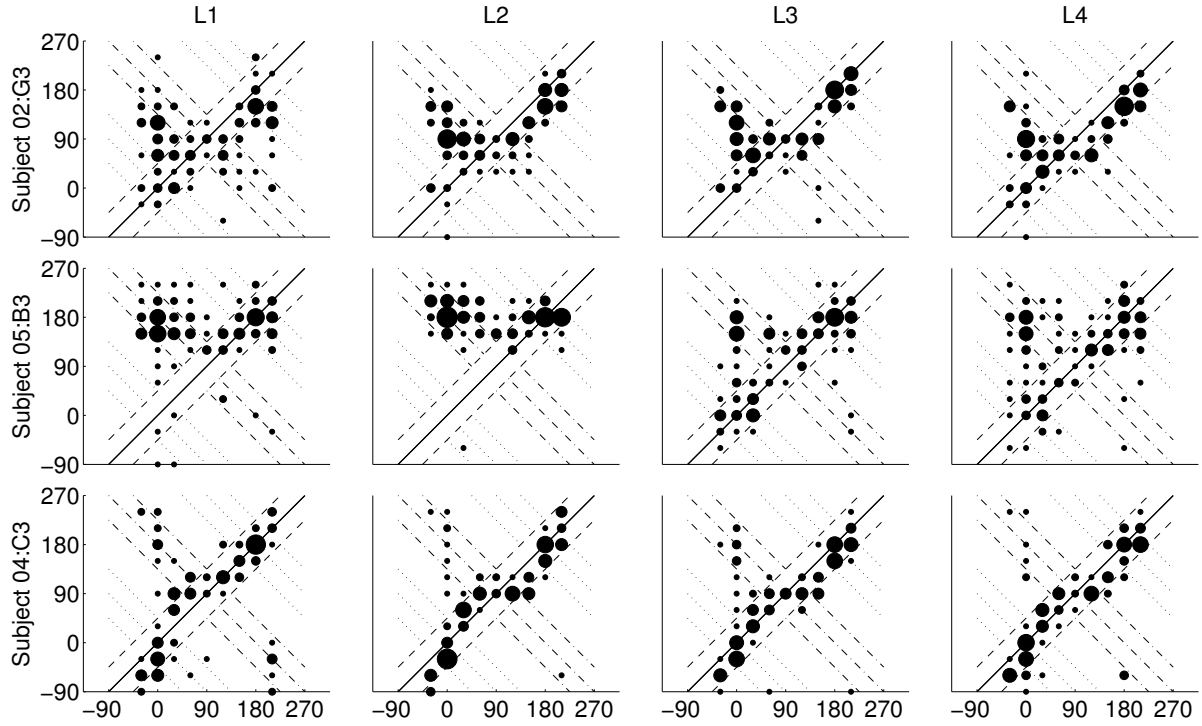


FIGURE 3.14 – Évolution de l'angle polaire perçu pour trois sujets représentatifs des groupes $G3$, $B3$ et $C3$. Les données sont représentées en coordonnées polaire interaural. Le numéro du test de localisation est indiqué en haut de chaque colonne.

erreurs subsistent encore à l'avant. Pour le test $L1$, le sujet du groupe $B3$ (avec de "mauvaises" HRTF non-individuelles) perçoit toutes les cibles dans l'hémisphère arrière (angle polaire compris entre 90° et 270°). Au cours de l'expérience, ce sujet acquiert progressivement la capacité à faire la différence entre l'avant et l'arrière.

Les résultats de l'analyse des types d'erreurs sont donnés en pourcentage dans le tableau 3.5 pour chaque groupe et chaque test de localisation. Idéalement, il ne devrait apparaître aucune confusion et toutes les erreurs devraient être contenues dans la zone de *précision*. Pour le groupe contrôle, toute l'amélioration apparait après la première session d'adaptation. Le pourcentage d'erreur de *précision* passe de 54% pour le test $L1$ à 66% pour le test $L2$ puis reste stable. Les autres types d'erreurs diminuent de 14% à 12% pour les erreurs de type *front/back*, de 7% à 5% pour les erreurs de types *up/down* et de 25% à 17% pour les erreurs *combinées*. L'augmentation du nombre d'erreurs de précision est donc principalement due à la diminution des erreurs combinées. Alors que le groupe $C3$ présente une proportion d'erreurs de précision comprise entre 54% et 67%, les autre groupes (avec des HRTF non-individuelles) obtiennent des proportions d'environs 40% pour le premier test et un maximum de 57% pour le dernier test. L'erreur de précision du groupe $G1$ augmente de 7% entre le premier et le dernier test. Cette amélioration est due à une réduction des erreurs combinées (de 32% à 25%), aucun effet n'est constaté sur les taux de confusions avant/arrière ou haut/bas qui restent stable à respectivement 18% et 6%. Le taux d'erreurs de précision du groupe $G3$ augmente de 41% à 57% ceci est principalement dû à une réduction des confusions avant/arrière (de 25%

error	C3				G3				B3			
type	L1	L2	L3	L4	L1	L2	L3	L4	L1	L2	L3	L4
precision	54	66	67	67	41	46	53	57	36	37	48	50
front/back	14	10	12	12	25	24	16	11	27	24	21	23
up/down	7	7	5	5	8	5	7	7	11	10	9	6
combined	25	17	16	17	27	25	25	25	26	28	23	20
error					G1				B1			
type					L1	L2	L3	L4	L1	L2	L3	L4
precision					43	45	46	50	40	41	39	43
front/back					19	18	18	18	24	29	28	28
up/down					6	6	7	6	11	7	11	8
combined					32	31	29	25	25	23	22	21

TABLE 3.5 – Distribution du type d'erreurs par groupe (en pourcentage).

à 11%). Les taux des autre types d'erreurs restent stables. Pour le groupe $B1$ aucune évolution des différents types d'erreurs n'est observée alors que pour le groupe $B3$ l'erreur de précision augmente de 36% à 50% au détriment de tous les autres types d'erreurs (diminution de 27% à 23% des confusions avant/arrière, de 11% à 6% des confusions haut/bas et de 26% à 20% des erreurs combinées).

d) On continue de jouer ?

Les résultats en angle polaire pour les groupes $C3$ et $B3$ montrent que la fin de la courbe d'apprentissage n'a pas été atteinte. On peut donc supposer que plus de sessions d'adaptation pourraient mener à des performances encore meilleure avec les HRTF individuelles. Afin d'explorer plus en profondeur l'effet du nombre de sessions d'adaptation sur l'amélioration des performances, nous avons demandé à un sujet du groupe $G3$ de poursuivre le test en faisant encore deux sessions d'adaptation. La quatrième session a eu lieu le lendemain de la troisième alors que la cinquième a eu lieu six jours après. L'évolution des résultats de ce sujet est représentée figure 3.15 et dans le tableau 3.6. Ayant été réalisés sur un seul sujet, ces résultats ne sont donnés qu'à titre exploratoire, pour donner des pistes sur les extensions possibles de cette étude. Ils n'ont pas de validité générale.

Ces résultats montrent qu'effectivement l'amélioration des performances continue à augmenter après la quatrième et la cinquième session d'adaptation. Au niveau des performances sur l'angle latéral, les résultats mettent en évidence un palier pour les tests de localisation $L4$ et $L5$ et un diminution de l'erreur après la cinquième session d'adaptation. L'amélioration des performances en angle latéral sur toute la durée du test est de 9.7°. Pour l'angle polaire, l'erreur moyenne diminue de 36.8° entre le premier et le dernier test de localisation ($L1$ et $L6$) et de 11° entre le quatrième et le dernier test de localisation ($L4$ et $L6$). La répartition des types de confusions pour le test $L6$ montre des performances quasiment identiques aux performances du groupe contrôle ($C3$) après la troisième session d'adaptation. Ce constat est confirmé par l'évolution du coefficient directeur de la droite de régression linéaire, dont la valeur après cinq sessions d'adaptation rejoint la valeur

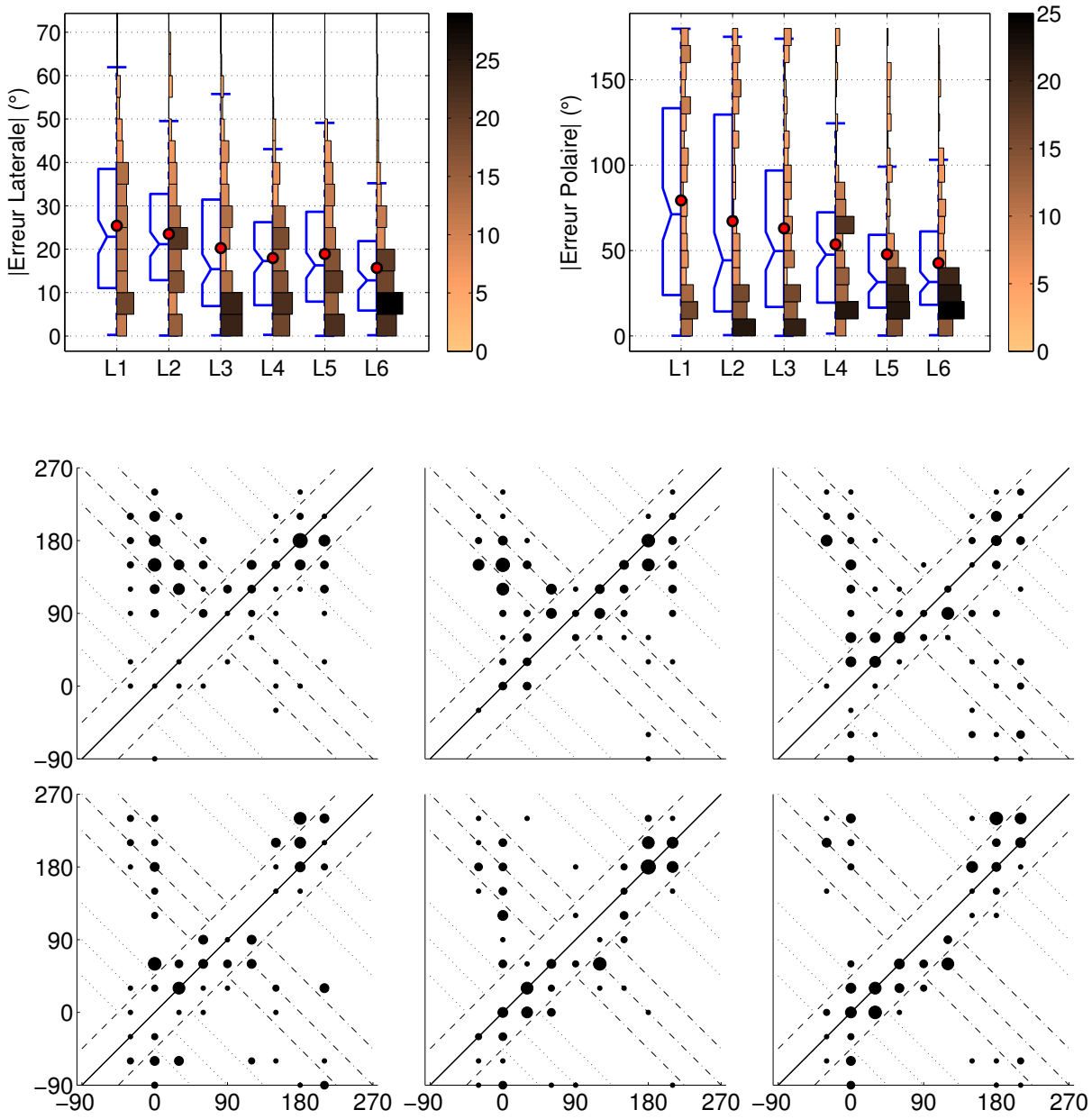


FIGURE 3.15 – Résultats du sujet ayant effectué cinq sessions d'adaptation avec des "bonnes" HRTF non-individuelles. En haut : Évolution de l'erreur latérale (à gauche) et de l'erreur polaire (à droite) avec une combinaison de boxplot et histogram. En bas : Évolution de l'angle polaire perçu en fonction de l'angle cible. Les données sont représentées en coordonnées polaires interaurales

Test	L1	L2	L3	L4	L5	L6
Erreur polaire moyenne	79.4	67.2	63.0	53.6	47.8	42.6
Erreur lateral moyenne	25.4	23.5	20.3	18.0	18.9	15.7
Regression Slope	0.067	0.093	0.043	0.409	0.609	0.613
Goodness-of-fit	0.008	0.017	0.002	0.141	0.270	0.247
precision	40.8	52.8	49.6	52.8	69.6	70.4
front/back	32.0	26.4	22.4	9.6	13.6	8
up/down	7.2	7.2	8.0	15.2	3.2	4.8
combined	20.0	13.6	20.0	22.4	13.6	16.8

TABLE 3.6 – Réccapitulatif des résultats du sujet ayant effectué cinq sessions d'adaptation avec des "bonnes" HRTF non-individuelles.

obtenue par le groupe contrôle au test $L3$ avec néanmoins un facteur de qualité d'ajustement plus faible. Les cinq jours de pause entre la quatrième et la cinquième session d'adaptation ne semblent pas avoir eu d'effet néfaste sur les performances de localisation. Il semble donc que l'adaptation perceptive puisse avoir une certaine pérennité. Cependant, il aurait été intéressant de faire un test de localisation avant la cinquième session d'adaptation pour étudier plus en profondeur l'effet de la pause de cinq jours.

3.6 Discussion

Le but de cette étude était d'explorer l'effet d'un environnement audio-kinesthésique sur la calibration rapide de la carte audio-spatiale des être-humains. L'étude des performances obtenues aux différents tests de localisation entourant les sessions d'adaptation a permis de montrer qu'au moins deux sessions d'adaptation sont nécessaires pour obtenir une amélioration significative des performances de localisation avec des HRTF non-individuelles.

La majorité de l'adaptation aux HRTF non-individuelles s'est faite sur l'angle polaire (lié aux indices spectraux). Une petite amélioration a été observée sur l'erreur latérale pour le groupe avec des "bonnes" HRTF après plusieurs sessions d'adaptation ; cette amélioration de l'interprétation des indices ITD est probablement due à une adaptation aux imprécisions du modèle d'ITD utilisé pour créer les HRTF hybrides. Cette amélioration n'a pas été observée chez les sujets avec des "mauvaises" HRTF non-individuelles, indiquant les éventuelles difficultés d'adaptation à des incohérences de plusieurs indices de localisation dans un délai aussi court. Le processus d'adaptation sur l'angle polaire a été effectif de la même façon pour les deux groupes ayant effectué trois sessions d'adaptation (diminution d'environs 23° de l'erreur polaire), les différences de performances entre les deux groupes se maintenant au cours de l'expérience avec approximativement 8° d'écart entre les deux groupes. Après trois sessions d'adaptation, les performances des sujets du groupe avec des "bonnes" HRTF ont rejoint les performances des sujets du groupe contrôle avec des HRTF individuelles.

Afin de séparer l'effet de la tâche d'adaptation, qui doit entraîner un apprentissage perceptif, de l'apprentissage procédural, ces résultats doivent être modérés par les résultats du groupe contrôle.

L'amélioration du groupe contrôle a été approximativement de 10° sur l'angle polaire. Étant donné qu'elle apparaît entre le premier et le deuxième test de localisation et qu'aucune amélioration n'est observée par la suite, nous considérons que cette amélioration est bien due à une adaptation à la tâche et au système de rendu utilisé. L'amélioration globale pouvant être attribuée de façon certaine à la tâche d'adaptation est donc de $23° - 10° = 13°$. Cet apprentissage perceptif d'une nouvelle carte audio-spatiale a eu lieu en trois sessions de 12 minutes réparties sur trois jours. Pour les deux groupes avec des HRTF non-individuelles, le taux d'erreurs de confusion a été réduit (avec une différence de répartition du type d'erreur entre les "bonnes" et les "mauvaises" HRTF). Les réductions observées sont comparables aux résultats de [Zahorik *et al.*, 2006] qui ont obtenu des réductions de 38% à 23% du taux de confusion avant/arrière pour deux sessions d'adaptation de 30 minutes avec un retour auditif, visuel et proprioceptif.

Comparées aux études antérieures, les performances générales obtenues dans cette étude par les sujets naïfs utilisant des HRTF non-individuelles sont légèrement moins bonnes. Dans leur étude [Wightman et Kistler, 1989b] obtiennent 11% de confusions avant/arrière contre 14% dans cette étude. Les résultats de l'étude de [Wenzel *et al.*, 1993] montrent 31% de confusions avant/arrière (dont 29% sont des erreurs combinées) et 18% de confusions haut/bas (dont 55% sont des erreurs combinées). Si l'on sépare les erreurs combinées des autres confusions, nous obtenons 22% de confusions avant/arrière et 8% de confusions haut/bas ce qui est comparable aux résultats obtenus par le groupe avec de "bonnes" HRTF non-individuelles dans cette étude (22% de confusions avant/arrière et 7% de confusions haut/bas). La méthode de quantification du type d'erreurs utilisée dans cette étude étant différente de celles utilisées dans les études précédentes, une comparaison précise n'est pas possible, mais il est clair que les performances obtenues sont comparables à la littérature, ce qui est suffisant au regard du but de cette analyse.

3.7 Conclusion

Les résultats de cette étude montrent qu'une adaptation perceptive rapide à des HRTF non-individuelles est possible en utilisant un retour auditif, proprioceptif et vestibulaire. De plus, l'adaptation du système auditif ne nécessite pas obligatoirement un retour visuel et elle peut être effectuée par les autres modalités sensorielles dans toute la sphère de perception auditive.

Pour la majorité de cette étude, nous n'avons effectué qu'une ou trois sessions d'adaptation. Au regard des performances obtenues par le sujet ayant fait plus de trois sessions d'adaptation et pour lequel chaque nouvelle session a permis une amélioration, il est possible d'imaginer qu'après trois sessions d'adaptation, la fin de la courbe d'apprentissage n'est pas encore atteinte. Nous ne pouvons rien affirmer quand au nombre de sessions nécessaire à une adaptation complète aux HRTF non-individuelles, mais au regard des résultats obtenus sur ce sujet, nous pouvons imaginer qu'après un certain nombre de sessions, les performances de localisation avec des indices non-individuels pourront coïncider avec les performances de localisation en champ libre. L'effet des sessions étant plus rapide lorsque les sujets utilisent des HRTF non-individuelles perceptivement

proches de leurs propres HRTF, la méthode de sélection proposée par [Katz et Parseihian, 2012] est efficace et nécessaire.

Dans de nombreuses applications telles que les tâches d'orientation et de navigation, l'utilisation d'un environnement virtuel auditif est souvent limité par le problème des confusions et par les artefacts induits par l'utilisation d'HRTF non-individualisées. Les résultats de cette étude sont donc significatifs pour les développeurs et utilisateurs de VAE. Alors que l'adaptation d'HRTF non-individuelles à un utilisateur donné nécessite un grand nombre de mesures et/ou une grande charge de calculs, la méthode présentée dans cette étude permet d'adapter graduellement le sujet en utilisant une sélection de son meilleur jeu d'HRTF comme point de départ. La combinaison d'un ITD individualisé (à partir de la mesure de la circonférence de la tête) avec un jeu de HRTF sélectionné (en utilisant un jugement perceptif) permet la création d'HRTF hybrides optimisées pour le sujet. Ensuite, au moins trois sessions d'adaptation en utilisant un jeu sur une plateforme multimodale permettent à l'utilisateur de rejoindre les performances qu'il aurait obtenu avec des HRTF individuelles et même de les dépasser. L'utilisation d'une procédure d'adaptation ne nécessitant pas la vue permet à cette méthode d'être appliquée aux non-voyants facilitant ainsi leur accès aux nouvelles technologies basées sur les sons 3D.

Chapitre 4

Amélioration des indices de perception de la distance en champ proche par l'utilisation de la sonification

Sommaire

4.1 Introduction

Ce chapitre s'intéresse à la localisation et la saisie d'objets dans l'espace péripersonnel de l'utilisateur. Il présente plusieurs études sur la faisabilité de la mise en place d'un dispositif de substitution sensorielle pour la détection d'objets en champ proche, utilisant la modalité sonore. Dans le cadre

du projet NAVIG, quatre expériences ont été réalisées pour explorer les mouvements de saisie d'objets dans l'espace péripersonnel. Les deux premières, menées à l'IRIT avec des sons réels ont permis d'étudier l'influence du stimulus et du temps de réverbération de la salle sur les performances de localisation. Une autre expérience sur la localisation de sons réels, menée au LIMSI, a permis d'étudier les différences de performances du geste de saisie en fonction de la main utilisée et sur l'étendue des directions accessibles par la main. À partir de ces résultats, une nouvelle expérience a été mise en place avec des sources virtuelles. Étant donné les erreurs observées pour les sons réels, plusieurs métaphores de sonification de la distance ont été conçues et testées pendant l'expérience avec les sons virtuels. Construites de la même façon que dans le chapitre 5, de manière à répondre aux attentes des utilisateurs en matière d'esthétique, ces métaphores de sonification sont basées sur des effets sonores, permettant une sonification indépendante du type de son utilisé.

Dans la section 4.2, nous rappellerons le contexte de la navigation en champ proche ainsi que les résultats importants de la littérature sur la perception de sources sonores dans l'espace péripersonnel. La section 4.3 détaillera l'expérience menée au LIMSI sur les sons réels après avoir rappelé les résultats obtenus dans les expériences préliminaires menées à l'IRIT. Enfin, le concept de sonification fondée sur des effets ainsi que l'expérience réalisée avec les sons virtuels seront détaillés dans la section 4.4.

4.2 Contexte

Les différentes études présentées dans ce chapitre ont eu lieu dans le cadre de la partie champ proche du projet NAVIG. Afin de poser les bases nécessaires à ces études, nous allons exposer dans cette section le concept de la navigation en champ proche ainsi que les principaux résultats obtenus dans les précédentes études sur la localisation en champ proche.

4.2.1 Navigation en champ proche

La saisie d'un objet dans l'espace proche est une tâche que les personnes non-voyantes effectuent très souvent et pour laquelle ils doivent développer une bonne mémoire de la position des objets dans l'espace. Bien que accessible, cette tâche est compliquée et leur demande souvent beaucoup de temps tout en étant moins précise que pour les personnes voyantes. Afin d'augmenter chez eux la capacité à localiser et attraper rapidement un objet dans l'espace proche, nous souhaitons mettre en place un dispositif qui permet de substituer aux positions visuelles, des positions sonores. L'objectif final du dispositif NAVIG en champ proche est de permettre au non-voyant de localiser un objet, d'avoir des informations sur sa taille et sa forme ainsi que sur le trajet à effectuer pour le saisir correctement. Plutôt que de chercher à décrire les informations inhérentes à l'objet sous forme verbale avec une description angulaire ou horaire de sa position ou encore en utilisant des techniques du type "tu chauffes ... tu refroidis ... tu brules", nous souhaitons placer un son virtuel

sur la position de l'objet à détecter et ajouter des informations sémantiques à ce son afin de rendre compte de sa taille et de sa forme.

Les expériences qui vont suivre ont été mises en place afin d'évaluer les possibilités de substitution de la vision par l'audition pour la détection et la saisie d'objets. Elles se focalisent sur l'étude de la précision de localisation en champ proche afin d'évaluer la précision de localisation et de saisie que l'on peut obtenir avec un tel dispositif. L'ajout d'informations sur la taille et la forme de l'objet ainsi que sur les obstacles potentiels entre la main et l'objet n'a pas été traité pendant cette thèse.

4.2.2 Performances de localisation des sons en champ proche

Nous avons vu dans la section 2.2.3 que les performances de localisation des sons ont été relativement bien étudiées dans la littérature. Cependant pour les sources proches (inférieures à un mètre), les capacités du système auditif sont moins connues et mènent en général à des performances plus faibles.

Au niveau des indices de localisation, [Brungart et Rabinowitz, 1999, Shinn-Cunningham *et al.*, 2000] ont montré que les plus grosses différences entre le champ proche et le champ lointain interviennent sur l'ILD. Ils mettent en évidence, à partir de mesures acoustiques un accroissement conséquent de l'ILD lorsque les sources ont une distance inférieure à un mètre. Cette augmentation est la conséquence de deux facteurs. Le premier correspond à l'augmentation de l'effet d'ombre acoustique par la tête. Plus la source est proche de la tête, plus le trajet acoustique contralatéral subit des atténuations des hautes fréquences. Le second est une conséquence de la loi de décroissance de l'intensité en $1/r^2$. Pour les petites distances, la différence de longueur du trajet acoustique de la source à chacune des deux oreilles est proportionnellement plus grand que pour les grandes distances et conduit à des différences d'intensités plus grandes et donc plus facilement perceptibles. En ce qui concerne l'ITD, il reste quasiment indépendant de la distance même pour les sources très proches. Bien qu'il existe une légère augmentation de l'ITD pour les distances les plus proches, cette augmentation se produit aux positions latérales où les auditeurs sont relativement insensibles aux changements d'ITD ([Hershkowitz et Durlach, 1969]). Au niveau de la variation des indices spectraux en champ proche, [Brungart et Rabinowitz, 1999] ont montré que la partie haute fréquence des HRTF est plus sensible aux changements d'élévation qu'aux variations de la distance. [Brungart, 1999] a cependant montré que dans le plan horizontal, les indices hautes fréquences des HRTF ne varient plus en fonction de l'angle entre la source et le centre de la tête, mais varient en fonction de l'angle entre la source et chacune des deux oreilles, entraînant ainsi un effet de parallaxe auditive.

L'évaluation des performances de localisation en champ proche a été principalement réalisée par Brungart. Dans [Brungart *et al.*, 2000], les auteurs ont évalué plusieurs méthodes de report des positions perçues (pointage égocentré avec un bâton, pointage allocentré autour d'une tête de mannequin, report verbal) et montré que, comme dans le cas des sources lointaines, la méthode la plus efficace est la méthode de report égocentrée. [Brungart *et al.*, 1999] décrivent les résultats de

l'expérience la plus conséquente en terme de performances de localisation de sons réels en champ proche. Réalisée sur 4 sujets, pour des sources situées dans l'hémisphère droit à des distances comprises entre 0.1 et 1 mètre, avec un stimulus composé d'une répétition de 5 bursts de 150 ms, cette expérience a été réalisée en quatre sessions de 2 heures (soit un total d'environs 2000 sources à localiser). Leurs résultats montrent qu'au niveau angulaire, l'erreur augmente lorsque la source approche de la tête. Elle est d'environs 20° pour les sources dont la distance est inférieure à 25 cm et d'environs 15° pour les distances supérieures. L'erreur en élévation, quant à elle, est plus faible pour les sources très proches et pour les positions latérales. Les plus grosses différences de performances interviennent pour la perception de la distance. En champ lointain, [Coleman, 1962, Gardner, 1968] ont montré que les auditeurs ne sont pas capable de déterminer de façon précise la distance absolue d'une source en champ libre. En champ proche, étant donné que l'ILD augmente lorsque la distance diminue alors que l'ITD reste constant, il est possible d'estimer la distance en comparant les valeurs de l'ITD et de l'ILD. Les résultats de [Brungart *et al.*, 1999] montrent que l'erreur de perception de la distance augmente lorsque la source s'éloigne, diminue pour les positions latérales et est plus grande pour les élévations supérieures à 20°.

4.3 Étude des mouvements de saisie vers des cibles sonores réelles

La mise en place du dispositif de suppléance visio-auditif nécessite une bonne connaissance des performances de localisation sonore et des mouvements de saisie dans l'espace qu'il devra couvrir. Étant donné qu'il doit permettre au sujet de localiser et d'attraper des objets à portée de main souvent placés à hauteur de ventre (poignée de porte, objets placés sur une table, ...), une plateforme expérimentale a été spécifiquement conçue pour évaluer la précision de localisation dans cette zone. Nous allons présenter dans cette section le dispositif mis en place ainsi que les trois expériences réalisées sur la localisation de sons réels : les deux expériences réalisées à l'IRIT et l'expérience réalisée par nos soins dans le cadre de cette thèse.

4.3.1 Description du dispositif utilisé pour les expériences

Le dispositif permettant de tester la précision des mouvements de saisie guidés par des sons réels dans l'espace péripersonnel a été mis en place conjointement par le LIMSI et l'IRIT dans le cadre du projet PLOREAV (précurseur du projet NAVIG) à l'IRIT. Ce dispositif, présenté figure 4.1, est constitué de 35 haut-parleurs (réf : CB990, 8 Ohms, 3 Watt) répartis sur un plateau semi-circulaire de 1 m de diamètre et placés sous une grille acoustiquement transparente et entourés de mousse acoustique afin de réduire les réflexions sur le plateau (figure 4.1 à droite). Ces haut-parleurs sont répartis sur cinq arcs de cercles concentriques espacés de 13 cm (leur distance au centre des cercles est de : 33 cm, 46 cm, 59 cm, 72 cm et 85 cm) ; chaque arc de cercle contient sept haut-parleurs espacés de 30° ; cette répartition est schématisée sur la figure 4.1 (en bas). Pour chaque expérience, les haut-parleurs étaient orientés de façon à pointer en direction de la tête du sujet, ceci afin de s'affranchir des problèmes de directivité. Les haut-parleurs ont été égalisés de

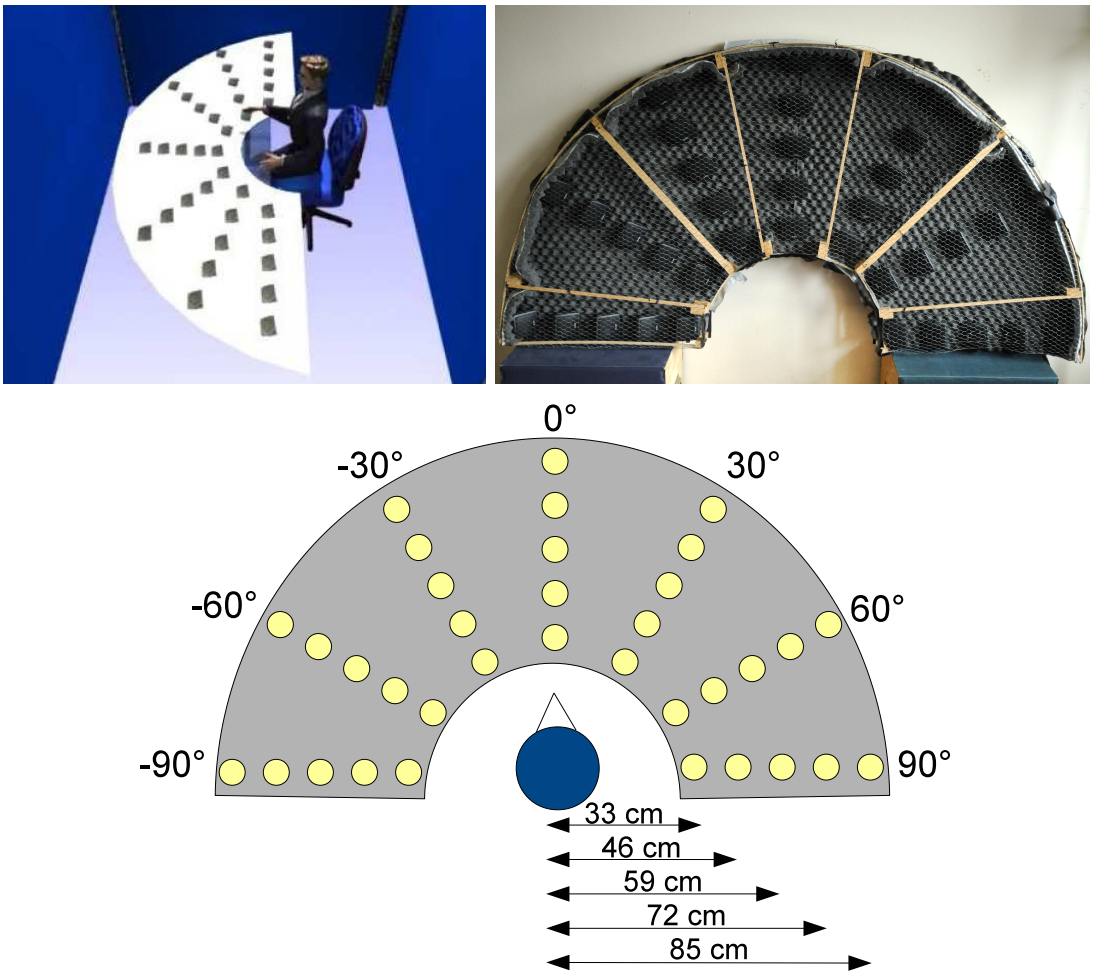


FIGURE 4.1 – A gauche : Schéma du dispositif utilisé (tiré de [Dramas, 2010]) ; à droite : photo du plateau de haut-parleurs ; en bas : Schéma du plateau avec le placement de chaque source sonore.

façon à avoir tous le même spectre et le même niveau au point d'écoute. L'égalisation du niveau sonore au point d'écoute supprime l'indice d'intensité sonore, mais permet d'éviter un jugement relatif au son précédent ("si le son est plus fort, c'est qu'il est plus près") et permet de se placer dans une condition où l'auditeur n'est pas familier avec le son (et où il ne connaît donc pas son niveau sonore). L'égalisation du spectre des haut-parleurs permet de s'assurer qu'aucune coloration spectrale due à des différences entre les haut-parleurs n'apporte des indices supplémentaires au sujet pour lui permettre de mémoriser des positions spécifiques.

4.3.2 Les expériences préliminaires menées à l'IRIT

La première expérience menée à l'IRIT avec des sujets voyants, s'est focalisée sur l'effet des stimuli utilisés ainsi que sur l'effet de la réverbération de la salle ; la deuxième a permis de comparer les performances de sujets voyants et non-voyants. Dans les sections qui suivent nous allons faire un bref résumé des résultats obtenus pour ces deux expériences. Pour plus d'informations, le lecteur pourra se rapporter à [Dramas *et al.*, 2008b] et [Dramas, 2010].

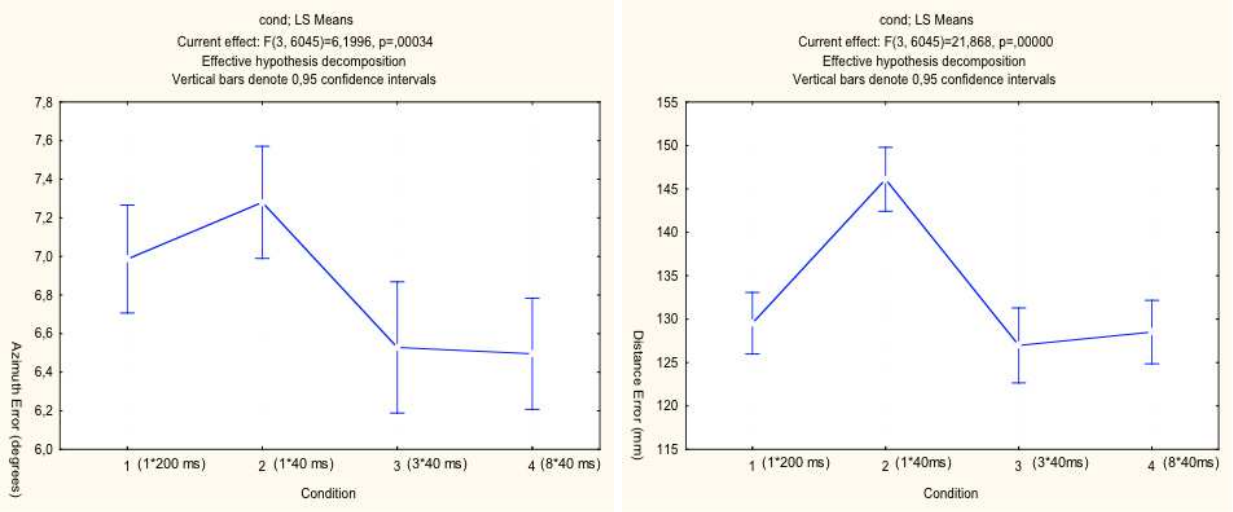


FIGURE 4.2 – A gauche : Erreur en azimut en fonction du stimulus ; à droite : Erreur en distance en fonction du stimulus. (Extrait de [Dramas, 2010])

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

10 sujets voyants (âgés de 20 à 60 ans), les yeux bandés ont participé à cette expérience. Quatre types de stimuli basés sur des bruits gaussiens ont été testés : 1 burst de 200 ms, 1 burst de 40 ms, 3 bursts de 40 ms et 8 bursts de 40 ms. Pour les deux derniers stimuli, chaque "burst" était espacé de 30 ms. Dans cette expérience, les 20 positions de la partie droite ont été testées (azimut de 0°, 30°, 60° et 90°) pour deux configurations de salle. Dans la première configuration, aucun traitement acoustique de la salle n'était effectué et son temps de réverbération était de 500 ms. Dans la deuxième configuration, la salle était acoustiquement traitée avec l'ajout de rideaux formant un cube de 2 mètres de côté dans lequel était placé le plateau et le sujet. Avec ce traitement acoustique, le temps de réverbération était de 350 ms.

Les résultats de cette expérience ont mis en évidence une meilleure précision de la perception de la distance sur le côté et de plus faibles performances pour les sources éloignées. En général, l'erreur en distance est de l'ordre de 13 cm soit environ 15% des distances à percevoir. La précision en azimut (comprise en moyenne entre 6 et 7°) est moins bonne sur les côtés (hormis à 90°) et ne semble pas varier en fonction de la distance. Une augmentation de la précision est apparue à 90°. Elle semble due à un effet d'apprentissage du dispositif. En effet, la rangée de haut-parleurs correspondant à 90° est située à l'extrémité du plateau. Les sujets ont donc pu mémoriser plus facilement ces positions en supposant qu'aucun son ne pouvait être émis plus à droite du dispositif.

Les figures 4.2 montrent l'effet des stimuli sur les performances en azimut et en distance. On remarque que la répétition des stimuli augmente significativement la précision en azimut et que seul le "burst" de 40 ms semble dégrader la précision de pointage en distance. Pour une même durée de stimulus (si l'on compare le "burst" de 200 ms à la répétition de 3 "bursts" de 40 ms espacés de 30 ms), le fait d'augmenter le nombre d'attaques en ajoutant des silences semble avoir un effet

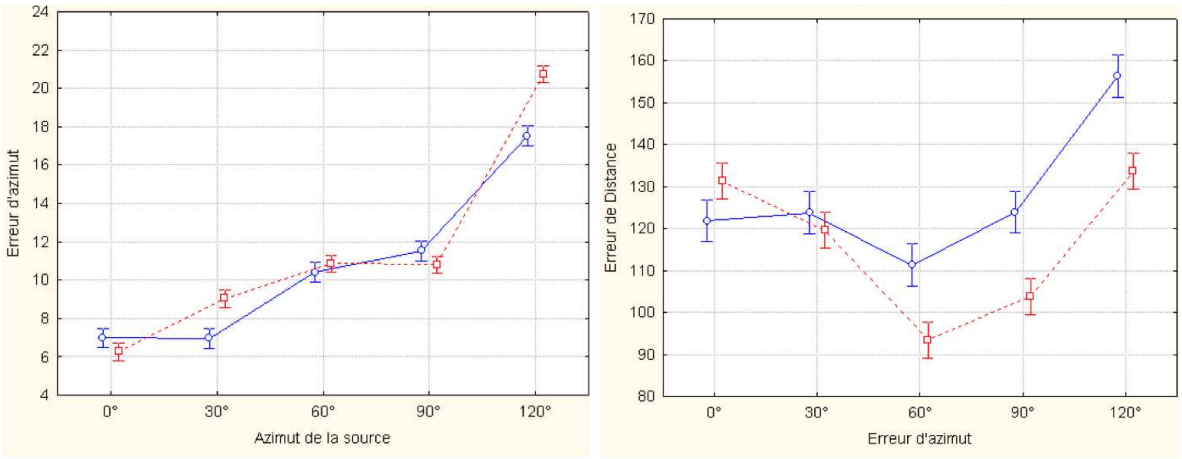


FIGURE 4.3 – A gauche : Erreur en azimut en fonction de l'azimut ; A droite : Erreur en distance en fonction de l'azimut. Rouge : Voyants, Bleu : Non-voyants. (Extrait de [Dramas, 2010])

positif sur la perception de la direction de la source et n'a pas d'effet négatif sur la précision en distance. Ce résultat est tout à fait en accord avec la littérature : les attaques franches permettent une meilleure perception de l'ITD entraînant ainsi une meilleure précision de la localisation en azimut.

Au niveau de l'effet de la réverbération sur la précision en distance, les résultats montrent que les performances en distance sont dégradées de façon significative par la diminution du temps de réverbération. Ce résultat est en accord avec la littérature qui montre que le rapport énergie directe/énergie réverbérée en espace clos est un bon indice de perception de la distance. La diminution du temps de réverbération n'a cependant aucune influence sur la précision angulaire de la localisation.

b) Expérience 2 : Comparaison voyants/non-voyants

19 sujets ont participé à la deuxième expérience (11 voyants et 8 non-voyants). Sept types de "bursts" ont été testés : 1x10 ms, 1x25 ms, 1x50 ms, 1x200 ms, 2x25 ms, 3x25 ms et 4x25 ms (pour les stimuli composés de une ou plusieurs répétitions, les "burst" étaient espacés par des pauses de 30 ms). Pour éviter les effets d'apprentissage du bord du plateau, les sujets étaient placés face à la rangée $-60°$ de la figure 4.1. 25 positions ont été testées pour cette expérience (azimut de 0°, 30°, 60°, 90° et 120°), aucun traitement acoustique de la salle n'a été effectué. Le temps de réverbération était donc d'environ 500 ms.

Concernant l'effet des stimuli, les résultats vont dans le même sens que la première expérience. Lorsque le "burst" n'est pas répété, la précision angulaire et la précision en distance augmentent avec la durée du stimulus. La répétition du "burst" a une influence significative sur les performances en azimut et en distance. Même si le stimulus est très court, sa répétition permet d'augmenter le nombre d'attaques et donc d'améliorer la perception de l'ITD. Ces résultats montrent que l'on peut obtenir les mêmes performances de localisation qu'avec un stimulus de 200 ms avec une répétition

de stimuli très court dont le temps total est inférieur à 200 ms.

Au niveau des différences entre les sujets voyants et non-voyants, les résultats montrent des performances quasi équivalentes pour la perception de l'azimut et une perception de la distance un peu plus faible chez les non-voyants. Cette différence de performance apparaît surtout pour les sources proches (les sources situées à 33 et 46 cm) et les positions latérales (angle de 60°, 90° et 120°). Ces résultats sont présentés figure 4.3.

4.3.3 Étude des performances en fonction de la main utilisée

L'étude menée au LIMSI dans le cadre de cette thèse, avait pour objectif de mesurer la précision de pointage vers une source sonore réelle dans un milieu peu réverbérant en fonction de la main utilisée pour le pointage et de la position de la source sur le plateau expérimental (localisation en azimut et en distance). L'objectif était de comparer les résultats de cette étude avec la même expérience réalisée avec des sons virtuels sur les mêmes sujets. Étant donné que certains problèmes techniques nous ont empêché à ce jour de réaliser l'expérience avec les sons virtuels, nous allons dans cette section décrire l'expérience réalisée avec les sons réels et analyser l'effet de la main de pointage sur la précision. Il existe peu d'études sur les capacités de localisation existantes dans cette zone. Face à ce constat, nous avons souhaité nous placer dans un cadre général en évaluant les performances de sujets voyants les yeux bandés.

a) Sujets

16 sujets voyants non-payés ont participé à cette expérience (3 femmes et 13 hommes, moyenne d'âge 25 ans, âge minimum 22 ans et âge maximum 30 ans). Un audiogramme a été effectué sur chacun des sujets afin de vérifier qu'aucun d'entre eux ne présentait de déficit auditif supérieur à 20 dB pour des fréquences comprises entre 125 et 8000 Hz. Ils étaient tous naïfs quant à la position des sources à localiser, avaient les yeux bandés pendant les sessions de localisation et le plateau était recouvert d'un drap opaque entre les sessions de localisation (ceci afin qu'ils ne voient pas le nombre et la répartition des haut-parleurs). Sur les 15 sujets, un seul était gaucher. Les résultats ont donc été traités en tenant compte de la latéralisation du sujet.

b) Matériel et méthode

Pour cette étude, nous avons utilisé le dispositif décrit dans la section 4.3.1. Les haut-parleurs étaient orientés de façon à tous pointer vers un point situé 65 cm au dessus de l'origine du demi disque. Des couettes et des mousses acoustiques ont été installées dans la salle d'expérimentation afin d'atténuer la réverbération du son sur les murs et le plafond. Avec ces matériaux absorbants, le temps de réverbération de la salle était de 300 ms. Cela est encore loin des conditions anéchoïques mais permet (comme nous l'avons vu dans les résultats de la première expérience décrite section 4.3.2) de diminuer l'effet de salle sur la perception de la distance.

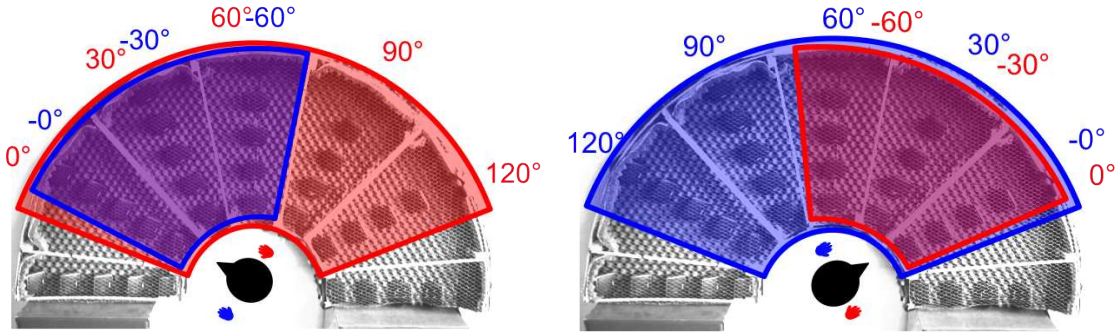


FIGURE 4.4 – Les deux configurations de placement du sujet dans le dispositif expérimental. Pour chaque configuration, la zone bleu correspond à la zone pointée avec la main gauche et la zone rouge, à la zone pointée avec la main droite.

L'expérience consiste à comparer la précision d'atteinte d'une cible auditive en fonction de sa distance et de son azimut pour la main "préférée" et pour la main "secondaire". Nous appelons main "préférée", la main avec laquelle le sujet effectue la plupart des tâches de la vie courante. S'il est droitier (respectivement gaucher), sa main "préférée" sera la main droite (respectivement la main gauche). Pour identifier cette main, nous avons demandé au sujet la main qu'il utilisait le plus, nous n'avons pas effectué de test de latéralisation plus poussé.

Pour chaque main, 35 positions ont été testées avec 5 distances différentes (33 cm, 46 cm, 59 cm, 72 cm et 85 cm par rapport au centre du plateau) et 7 azimuts différents ($-60°$, $-30°$, $0°$, $30°$, $60°$, $90°$ et $120°$). Afin d'éviter les effets d'apprentissage dus aux bords du plateau, seuls 25 haut-parleurs ont été utilisés sur les 35. L'expérience a donc été réalisée en quatre phases :

- Phase 1 (Figure 4.4(gauche) zone rouge) : le sujet est tourné vers la gauche du système et effectue la tâche avec sa main droite. Il doit localiser 25 positions pour des azimuts de $0°$, $30°$, $60°$, $90°$ et $120°$.
- Phase 2 (Figure 4.4(gauche) zone bleu) : le sujet est tourné vers la gauche du système et effectue la tâche avec sa main gauche. Il doit localiser 15 positions pour des azimuts de $0°$, $-30°$ et $-60°$.
- Phase 3 (Figure 4.4(droite) zone bleu) : le sujet est tourné vers la droite du système et effectue la tâche avec sa main gauche. Il doit localiser 25 positions pour des azimuts de $0°$, $30°$, $60°$, $90°$ et $120°$.
- Phase 4 (Figure 4.4(droite) zone rouge) : le sujet est tourné vers la droite du système et effectue la tâche avec sa main droite. Il doit localiser 15 positions pour des azimuts de $0°$, $-30°$ et $-60°$.

Pour chaque main, la localisation des sources positionnées sur la rangée située à $0°$ a été répétée deux fois. Cette répétition nous permettra plus loin d'analyser l'influence du découpage de l'expérience en plusieurs parties ne correspondant qu'à un seul côté à la fois. Le sujet était assis sur une chaise pivotante, la tête située 65 cm au dessus du centre du demi disque. L'orientation de la tête et la position du doigt du sujet étaient suivies avec un système de tracking infrarouge (Optitrack couplé au logiciel Arena). Le stimulus utilisé était le triple "burst" gaussien de 40 ms espacés de 30 ms, résultant de l'expérience décrite dans la section 4.3.2.a). Le spectre de ce stimulus est large

	Azimut	−60°	−30°	0°	30°	60°	90°	120°	Total
Erreur moyenne	Main P	9.3 (7.8)	6.7 (5.4)	6.5 (7.8)	7.1 (5.8)	11.1 (8.9)	15.2 (12.4)	36.1 (12.4)	12.3 (14.6)
	Main S	11.9 (9.7)	6.9 (5.9)	6.1 (7.4)	8.2 (7.3)	9.2 (8.0)	17.7 (12.0)	42.0 (23.5)	13.6 (16.2)
Conf AV/Ar %	Main P	0	0	0.1	0	3.5	0	56.5	7.5
	Main S	4.3	0	0.1	0.3	0.8	0	66.3	9.1

TABLE 4.1 – Moyenne de l'erreur angulaire absolue et pourcentage de confusions avant/arrière en fonction de l'angle de la cible.

bande (20–20000 Hz), son niveau sonore était de 60 dBA au point d'écoute. Le protocole expérimental était totalement contrôlé par l'ordinateur. Pour déclencher un son, le sujet devait placer la main équipée du capteur sur sa poitrine (dans une zone de ± 10 *cm* située autour ! d'un point préalablement enregistré) et garder la tête droite en direction de la rangée correspondant à la phase de l'expérience. L'orientation du sujet était vérifiée par l'expérimentateur au début de chaque phase (permettant ainsi l'enregistrement d'une position de référence) puis contrôlée par l'ordinateur (avec une tolérance de $\pm 3°$) sur les angles yaw et pitch. Étant donné le peu de probabilité que les sujets penchent leur tête sur les côtés, nous n'avons pas ajouté de vérification de l'angle roll. Lorsque le sujet était mal placé, une voix "suave" diffusée par un haut-parleur n'appartenant pas au plateau lui indiquait dans quel sens tourner sa tête pour se placer correctement. Une fois la main et la tête correctement placées et stables pendant un délai aléatoire (compris entre 100 et 700 ms), le stimulus était présenté sur des haut-parleurs appartenant au plateau. Le sujet devait alors pointer son doigt vers la position perçue du son et la valider avec un bouton placé dans son autre main ; puis revenir en position initiale pour passer à un autre essai. Chacune des 25 positions étant répétée 5 fois, le nombre d'essais était de 125 pour les phases 1 et 3 et de 75 pour les phases 2 et 4, soit un total de 400 sources à localiser. Pour chaque phase, l'ordre des stimuli était divisé en cinq blocs (pour les cinq répétitions) complètement aléatoires. La durée de ce test était d'environ 1h30.

c) Résultats

Les positions testées et les positions perçues sont calculées par rapport au centre de la tête en coordonnées sphériques (azimut, élévation et distance). Etant donné la configuration du plateau, la distance et l'élévation de la source par rapport à la tête sont interdépendantes, nous n'analyserons donc que les résultats concernant la distance ainsi que ceux concernant l'azimut. Pour ces deux variables nous commencerons par évaluer les performances générales obtenues avec la main préférée puis nous étudierons l'effet de la main utilisée. Enfin, nous étudierons les biais par zone de pointage et tenterons de les relier aux capacités mécaniques des bras.

i/ Performances de localisation en azimut

L'azimut de la cible variait entre -60 et $120°$ par pas de $30°$. La figure 4.5 présente la moyenne de l'azimut perçu en fonction de l'azimut généré, la valeur absolue de l'erreur moyenne ainsi que le pourcentage de confusion avant/arrière en fonction de l'angle et de la main utilisée sont présentés

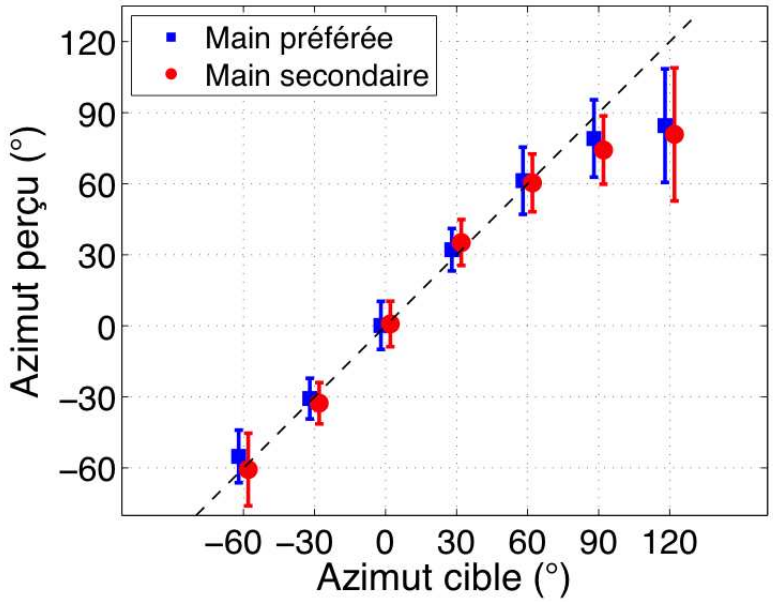


FIGURE 4.5 – Moyenne de l'azimut perçu en fonction de l'azimut cible.

dans le tableau 4.1.

Ces résultats mettent en évidence une bonne précision de localisation dans le plan médian (entre −30 et 30°) avec une erreur absolue moyenne de 6.7° et des performances plus faibles sur les côtés avec une erreur absolue moyenne de 17.8°. On remarque sur les courbes de la figure 4.5 que pour les positions médianes, l'azimut perçu correspond relativement bien à l'azimut cible et que l'erreur standard est assez faible. Pour les positions latérales la précision est plus faible, l'azimut a tendance à être sous-estimé et la variabilité des réponses est plus grande. L'analyse des confusions avant/arrière (répertoriées dans le tableau 4.1) met en évidence un grand pourcentage de confusion à 120° et quelques confusions à -60, 0 et 60°. L'azimut 120° est perçu une fois sur deux à 60° cela explique les erreurs très grandes observées pour cet angle. En corrigeant les confusions avant/arrière pour cet angle, on obtient une moyenne d'erreur absolue à 15.9°. Il est surprenant de noter que quelques confusions ont eu lieu de l'avant vers l'arrière (à 0 et 30°) alors que les sujets étaient conscients de la géométrie du plateau (sans connaître la position exacte des sources). Dans ces cas là, les sujets pointaient derrière eux dans le vide.

L'analyse des performances en fonction de la main de pointage met en évidence des erreurs similaires dans le plan médian et des différences un peu plus grandes sur les côtés. En général, les performances sont légèrement meilleures avec la main préférée. Une analyse ANOVA à mesures répétés réalisée sur l'erreur en azimut (sans corriger les confusions avant/arrière) en prenant en compte la main de pointage ne met pas en évidence de différence significative [$F(1,14)=2.76$, $p=0.12$]. Lorsque l'on prend aussi en compte l'azimut de la cible [$F(6,84)=2.44$, $p<0.05$], on obtient une différence significative entre les mains de pointage à 120° ($p < 10^{-5}$). L'analyse en fonction de la distance et de la main de pointage [$F(4,56)=2.86$, $p<0.05$] met en évidence une influence significative de la main de pointage pour toutes les distances (hormis la plus proche).

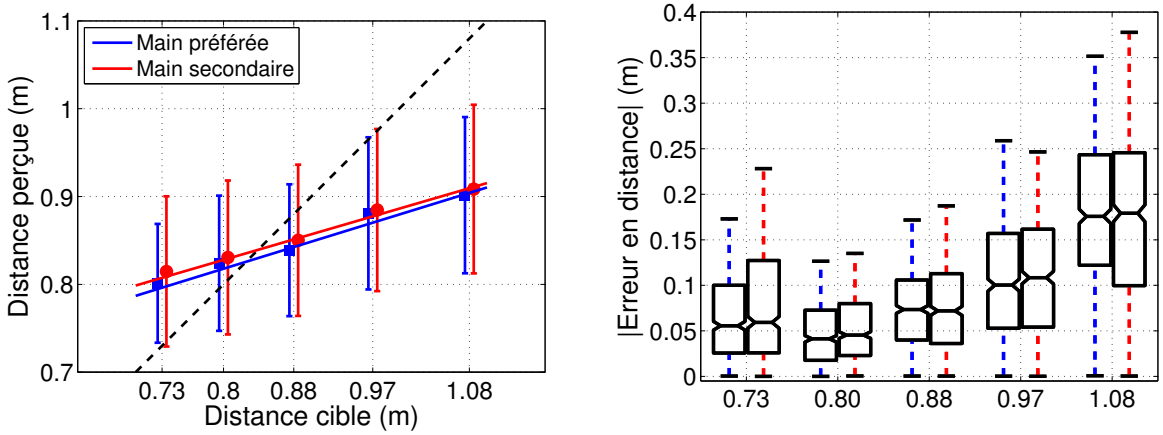


FIGURE 4.6 – Résultats pour la localisation en distance. Distance perçue en fonction de la distance cible à gauche. Valeur absolue de l'erreur de distance en fonction de la distance cible à droite. Résultats pour la main préférée en bleu et pour la main secondaire en rouge.

	Azimut	−60°	−30°	0°	30°	60°	90°	120°	Total
Erreur en	Main P	9.5	9.9	10.7	10.5	8.7	8.6	10.0	9.8 (7.7)
distance (cm)	Main S	9.8	10.6	10.5	10.1	9.0	9.4	12.1	10.3 (8.2)
Coefficient	Main P	0.31	0.29	0.21	0.25	0.38	0.39	0.30	0.30 (0.11)
directeur	Main S	0.34	0.24	0.24	0.26	0.35	0.34	0.22	0.28 (0.11)
Qualité	Main P	0.78	0.77	0.77	0.80	0.90	0.79	0.63	0.95 (0.05)
d'ajustement r^2	Main S	0.87	0.74	0.82	0.77	0.80	0.77	0.59	0.92 (0.11)

TABLE 4.2 – Coefficient directeur de la droite de régression linéaire et coefficient de qualité d'ajustement r^2 pour chaque angle et pour l'ensemble des données.

En général, les performances de localisation en azimut sont autour de 7° dans le plan médian et supérieur à 10° sur les côtés. Hormis pour quelques positions spécifiques, la main utilisée n'a pas d'influence sur la précision de pointage.

ii/ Performances de localisation en distance

Pour chaque direction, cinq distances comprises entre 0.7 et 1.1 mètres (calculées entre le centre de la tête et la source) devaient être localisées. La figure 4.6 présente la moyenne de la distance perçue en fonction de la distance générée, avec pour chaque condition, la droite s'ajustant le mieux aux données. La valeur absolue de l'erreur en distance en fonction de l'angle et de la main utilisée est présentée dans le tableau 4.2 avec les coefficients directeurs calculés avec une régression linéaire pour chaque angle ainsi que le coefficient de l'ajustement.

Les résultats mettent en évidence une plus grande difficulté à percevoir la distance. L'erreur globale est de 10 cm, soit 11% d'erreur relative (calculée par rapport au centre du plateau). Elle est légèrement plus faible sur les côtés (9.2 cm) qu'à l'avant (10.4 cm). Bien que compressée entre 0.8 et 0.9 mètres, la figure 4.6(gauche) montre que la perception de la distance est linéaire. Ceci est vérifié en faisant une régression linéaire sur la distance perçue en fonction de la distance cible.

On obtient un coefficient directeur moyen de 0.30 ± 0.11 avec une qualité d'ajustement de 0.95. Ce résultat montre que la perception de la distance est faible (dans le cas idéal on devrait obtenir un coefficient directeur égal à 1) mais qu'elle est bien linéaire (le facteur de qualité d'ajustement est quasiment égal à 1). La variabilité des résultats est très grande, avec un écart type de l'ordre de 10 cm, ce qui tend à montrer que la perception de la distance avec les indices sonores fournis est très faible et qu'il existe de grandes disparités d'un sujet à l'autre. La figure 4.6(droite) montre les boxplot de la valeur absolue de l'erreur en distance en fonction de la distance cible. On remarque que cette erreur est minimum pour la deuxième position la plus proche et maximum pour les sources les plus éloignées du sujet. L'analyse de l'évolution des coefficients directeurs moyens en fonction de l'angle de pointage (tableau 4.2) met en évidence l'amélioration de la perception de la distance sur les côtés. En effet, les plus grands coefficients directeurs sont obtenus pour les angles -60, 60 et 90°. Ce résultat confirme les observations sur l'évolution de l'erreur en distance en fonction de l'angle.

L'analyse des performances en fonction de la main de pointage montre peu de différences entre les deux conditions (main préférée ou secondaire). Les résultats présentés sur la figure 4.6(gauche) montrent que la perception de la distance est quasiment la même dans les deux cas avec cependant une variabilité légèrement supérieure pour le pointage avec la main secondaire. Les coefficients directeurs des régressions sont quasiment égaux (0.30 ± 0.11 pour la main préférée contre 0.28 ± 0.11 pour la main secondaire) et les facteurs de qualité d'ajustement aussi (0.95 pour la main préférée contre 0.92 pour la main secondaire). Une analyse ANOVA à mesures répétées réalisée sur l'erreur en distance en prenant en compte la main de pointage met en évidence une différence significative [$F(1,14)=5.97$, $p<0.05$]. Cette différence, qui n'est pas visible sur les courbes ou dans les moyennes peut s'expliquer lorsque l'on prend en compte l'azimut de la cible [$F(6,84)=2.17$, $p=0.053$]. En effet, un test post-hoc de Duncan met en évidence une différence significative entre les deux conditions pour la direction 120° ($p < 0.001$). Cette différence se retrouve dans le tableau 4.2 qui montre que les sujets obtiennent une erreur moyenne de 10.0 cm avec la main préférée et de 12.1 cm avec la main secondaire. L'analyse de la variance en prenant comme facteur la condition et la distance ne montre quant à elle aucune différence significative entre la main préférée et la main secondaire [$F(4,56)=0.90$, $p=0.47$].

La perception de la distance dans la zone étudiée est plus mauvaise que la perception de l'azimut. Bien que linéaire, cette perception est compressée dans une zone correspondant au milieu du plateau (coefficient directeur de la droite de régression égale à 0.3). L'erreur globale est de 10 cm et les sujets sont plus performants sur les côtés. Les performances sont un peu meilleures avec la main préférée, surtout pour les positions extrêmes (à 120°).

iii/ Biais du geste de saisie

Nous avons analysé dans les sections précédentes les erreurs en azimut et en distance de la même façon que dans les études de localisation classique. Afin de faire le parallèle entre ces résultats et la finalité de notre dispositif, nous allons présenter ici les moyennes des biais de localisation sur le

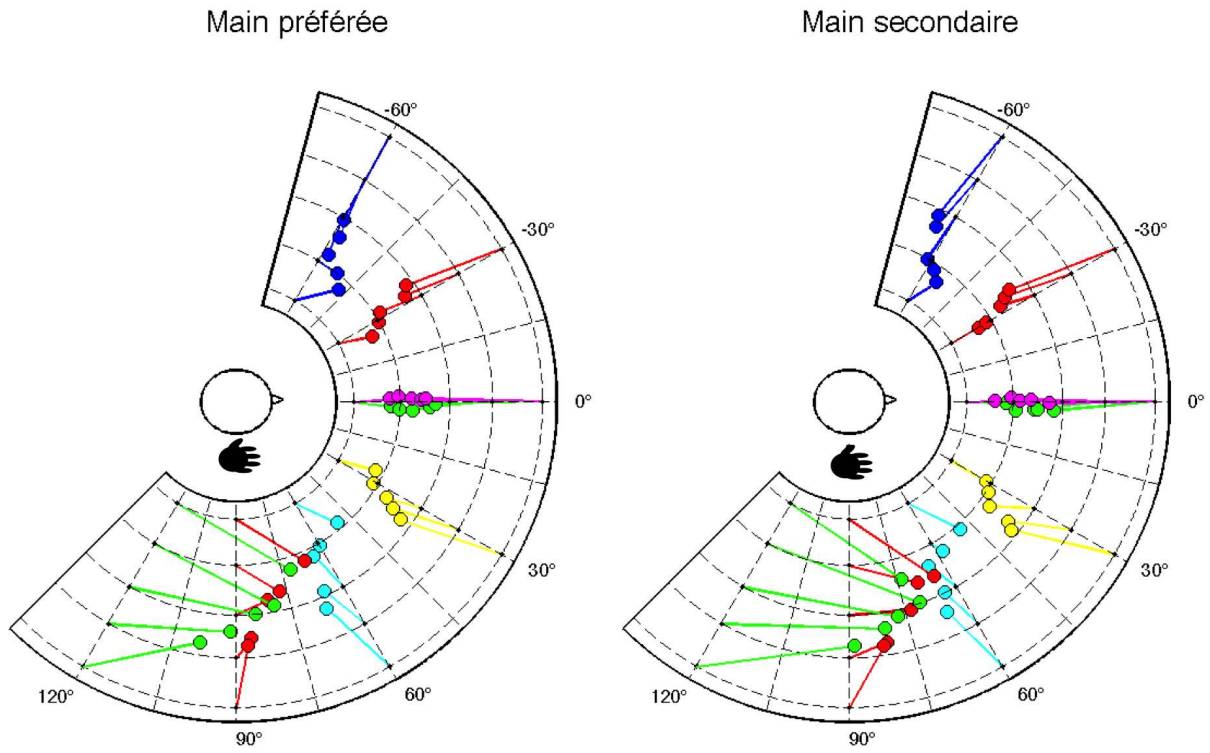


FIGURE 4.7 – Représentation des positions moyennes pointées avec la main préférée (à gauche) et la main secondaire (à droite).

plateau et comparer ces biais aux capacités de pointage de sujets vers des cibles visuelles.

La figure 4.7 présente les positions moyennes pointées avec la main préférée (à gauche) et la main secondaire (à droite). Étant donné la grande variabilité des réponses, nous n'avons pas tracé les ellipses de confidences qui rendaient ces figures illisibles. Les deux figures ont été placées dans le même sens afin de faciliter la comparaison. La présence de deux couleurs de points pour la rangée médiane (0°) est due à la décomposition de l'expérience en deux phases (voir section 4.3.1). Les points violets correspondent au pointage vers la partie droite du plateau (de 0 à 120°), les points verts correspondent au pointage vers la partie gauche (de 0 à $-60°$). Ces figures permettent de faire les mêmes constats que dans les sections précédentes (compression de la distance, erreur angulaire qui augmente sur les côtés, grand nombre de confusions avant/arrière à 120°, ...). On observe de légers décalages en azimut à 0° qui sont à chaque fois opposés à la partie du plateau utilisé. Ce petit décalage a également été observé dans les deux expériences présentées section 4.3.2, il semble être causé par le protocole expérimental. En effet, pour chaque phase, les sources n'étaient positionnées que d'un côté du plateau (à gauche ou à droite) créant ainsi une asymétrie qui a pu potentiellement entraîner un décalage perceptif vers le côté opposé pour la ligne médiane. Pour vérifier cette hypothèse, il serait intéressant d'effectuer le même test en répartissant les sources des deux côtés de l'utilisateur.

Certains biais apparaissant sur les côtés, pour les sources les plus proches, semblent inhérents à des problèmes mécaniques corporels. Il est effectivement difficile de placer correctement la main

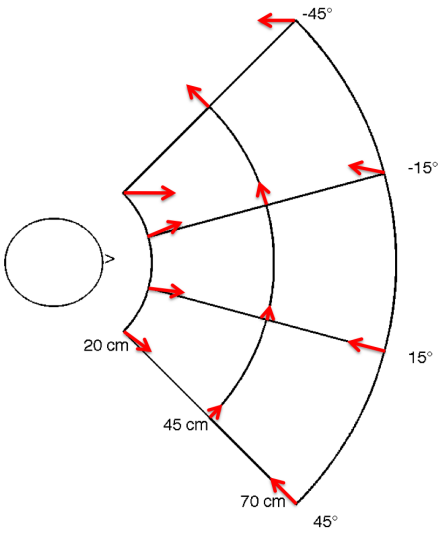


FIGURE 4.8 – Distorsion spatiale lors du pointage dans le noir vers des positions visuelles mémorisées (d'après l'étude de [Soechting et Flanders, 1989]). Les cibles visuelles correspondent aux intersections entre les rayons et les arcs du plateau qui était situé 40 cm en dessous de la tête du sujet.

près du corps sur les côtés (notamment pour l'azimut $-60°$) et cela peut expliquer les décalages vers le plan médian observés à ces positions. La figure 4.8 représente les biais de pointage vers des positions visuelles mémorisées tirés de l'étude de [Soechting et Flanders, 1989]. Dans cette étude, les auteurs ont quantifié les erreurs de pointage de sujets placés dans le noir pour des cibles reparties dans une zone de -45 à $45°$ d'azimut, de 20 à 70 cm de distance et pour des plateaux placés à des hauteurs de -40, -20, 0, 20 et 40 cm par rapport à la tête du sujet. Les flèches rouges de la figure montrent les distorsions observées lorsque le plateau était 40 cm en dessous de la tête (donc 25 cm au dessus de notre plateforme expérimentale). Leurs résultats mettent en évidence une tendance à compresser la distance ainsi qu'un léger décalage vers le centre des cibles latérale proches. Étant donné que la configuration de leur expérience n'est pas la même que la nôtre et que leurs cibles ne vont que de -45 à $45°$ d'azimut, il est difficile d'extrapoler leur résultat à notre étude. Nous pouvons cependant émettre l'hypothèse que les erreurs observées lors de notre expérience ne sont pas toutes dues à un problème de localisation auditive. Pour aller plus loin, il serait intéressant de répéter l'expérience de [Soechting et Flanders, 1989] réalisée avec des cibles visuelles, pour les mêmes positions et dans la même configuration que notre expérience.

4.3.4 Discussion

Le but de cette expérience était de déterminer la faisabilité d'un système donnant des informations de position avec des sons 3D dans l'espace peripersonnel et de quantifier la précision qu'il est possible d'atteindre dans le "cas idéal" : l'utilisation de sons réels.

Les résultats mettent en évidence une grande variabilité des réponses dans la zone étudiée. Malgré cette disparité dans les réponses des sujets, les résultats montrent qu'ils sont capables de percevoir

et d'indiquer la direction du son sur une table avec une précision d'environs 13° en général et de 6 à 7° dans la zone frontale. Au niveau de la distance, les performances sont assez médiocres. Les sujets ont une légère perception de la distance mais celle-ci est compressée dans une partie restreinte de la zone étudiée. Si cette perception reste bien linéaire (avec des facteurs de qualité d'ajustement de 0.95 et 0.92), le coefficient directeur moyen des droites de régression de 0.3 montre que cette compression rend difficile la tâche d'atteinte d'une cible sonore avec la main.

Les résultats médiocres obtenus au niveau de la perception de la distance peuvent être expliqués de plusieurs manières. Tout d'abord, l'étendue des variations de distance dans cette expérience est assez faible (les distances vont de 0.72 à 1.08 mètres, soit une variation de 36 cm), ce qui limite les variations des indices permettant de percevoir la distance. Ensuite, la suppression de l'indice d'intensité sonore (pour se placer dans le cas d'une source inconnue) et l'environnement peu réverbérant limite les indices de perception de la distance aux variations des différences binaurales.

La comparaison des résultats avec la littérature est complexe étant donné que le dispositif utilisé dans cette expérience est différent de tous ceux utilisés dans les expériences précédentes. Nous pouvons néanmoins faire quelques comparaisons avec les résultats de l'étude de [Brungart *et al.*, 1999] qui a étudié la localisation de sources sonores dans toute la sphère péripersonnelle pour des distances comprises entre 0.1 et 1 mètre. Afin de se placer dans la même zone, nous ne rappellerons que leurs résultats obtenus pour les zones frontales et latérales, avec des élévations inférieures à $-20°$ et des distances comprises entre 0.5 et 1 mètre. Concernant la perception de l'azimut, l'erreur moyenne obtenue par [Brungart *et al.*, 1999] est de 13.4° sur le côté (de -120 à $-60°$) et de 16.1° pour la zone frontale ($<-60°$). Nos résultats (de respectivement 18° pour la zone latérale et 7° pour la zone frontale) affichent une tendance inverse des résultats de Brungart. La première explication est que nous n'avons ni supprimé ni corrigé les confusions avant/arrière alors que Brungart les a supprimées. Dans notre étude, l'erreur moyenne en azimut dans la zone latérale lorsque l'on supprime les confusions avant/arrière est de 13°, nous obtenons donc des résultats similaires à Brungart dans cette zone. La différence de précision dans la zone frontale est assez surprenante. Dans notre étude, les sujets sont plus performants à l'avant que sur les côtés, ce qui est similaire aux résultats obtenus dans les études de localisation classiques (recencées dans [Blauert, 1997] et mentionnées dans la section 2.2.3b)). Dans l'étude de Brungart, quelque soit la zone pointée, l'erreur est plus grande dans la zone frontale, ce qui semble aller à l'encontre des résultats généraux. Au niveau de la perception de la distance, l'étude de Brungart affiche des résultats plus précis que nôtre étude. Avec des coefficients de régression de 0.90 dans la zone latérale et de 0.69 dans la zone frontale, la distance est mieux perçue dans leur étude que dans la notre (coefficient de régression de 0.30). La principale raison de cette différence semble être l'étendue de variation des distances présentées dans chaque expérience et les indices acoustiques présentés aux sujets. Dans notre expérience, les distances varient entre 0.7 et 1.1 mètre alors que dans l'expérience de Brungart la distance varie entre 0.1 et 1 mètre donc l'étendue couverte est plus que deux fois plus grande ce qui entraîne plus de variations des indices de distance. Dans notre expérience, nous avons supprimé l'indice d'intensité afin d'éviter que le jugement soit relatif au son précédent. Le niveau d'intensité n'a pas été normalisé dans l'expérience de Brungart ce qui ajoute un indice de perception de la

distance.

La transmission d'une information spatiale avec uniquement des indices acoustiques réels très courts est donc possible mais peu précise au niveau de la profondeur. Dans la prochaine section, nous allons étudier les dégradations dues à l'utilisation d'indices acoustiques virtuels ainsi que les solutions envisagées pour améliorer la perception de la distance.

4.4 Localisation et saisie de cibles virtuelles et sonification de la distance

Nous avons étudié dans la section précédente les performances du geste de saisie vers des cibles sonores réelles dans l'espace proche. Les résultats ont montré qu'il est possible de transmettre une position spatiale en utilisant des indices sonores, mais que les performances ne sont pas optimales en ce qui concerne les informations sur la distance de la cible. Dans cette section, nous étudions les performances du geste de saisie vers des cibles sonores virtuelles (synthétisées en binaural et présentées sur un casque stéréophonique). Étant bien conscient que les artefacts de la synthèse binaurale risquent de diminuer les performances observées pour les cibles sonores réelles, une stratégie de sonification a été mise en place pour tester l'amélioration de la perception de la distance avec l'ajout de nouveaux indices sonores.

Afin de répondre aux besoins des utilisateurs (présentés dans l'annexe A), plusieurs métaphores de sonification de la distance ont été mises en place. Basées sur des effets sonores, ces métaphores peuvent être appliquées à tout type de son et ainsi augmenter les chances de satisfaire les exigences esthétiques de tous les utilisateurs. Nous allons commencer par présenter les raisons qui nous ont incitées à mettre en place un nouveau paradigme de sonification basé sur des effets sonores. Puis nous discuterons de la création de plusieurs métaphores basées sur ce concept. Enfin, nous présenterons une évaluation perceptive de ces métaphores à travers une expérience de localisation de sons en champ proche. Cette expérience permettra de comparer les performances de localisation obtenues avec les différentes métaphores à celles obtenues avec un rendu en binaural seul afin de quantifier la contribution de chaque métaphore sur la perception de la distance. Les résultats seront d'abord comparés aux performances obtenues avec des sons réels (section 4.3) afin de quantifier les dégradations induites par la synthèse binaurale et les améliorations entraînées par la sonification de la distance.

4.4.1 Sonification des indices de localisation

Malgré la multiplicité des indices permettant la perception de la distance, la synthèse de ceux-ci dans un VAE reste compliquée à mettre en oeuvre et mène à des résultats assez pauvres en particulier pour le rendu de sources proches. Nous allons, dans cette section, présenter les différentes solutions envisagées dans la littérature pour améliorer le rendu sonore dans les VAE ainsi que les techniques de sonification dont nous nous sommes inspiré pour mettre en place nos métaphores.

Outre les tentatives de reproduction des indices naturels et leur adaptation à plusieurs situations (indices mesurés à une certaine distance et adaptés à d'autres distances), plusieurs études ont abordé l'amélioration du rendu sonore des VAE par distorsion ou exagération des indices acoustiques. La première étude adoptant cette approche, réalisée par [Durlach *et al.*, 1993], consistait à diminuer l'angle minimum audible en azimut à l'avant (donc améliorer la précision de localisation) en changeant la relation entre la direction à simuler et les HRTF correspondantes. Les auteurs montrent qu'avec un apprentissage perceptif (développé dans [Shinn-Cunningham *et al.*, 1998a, Shinn-Cunningham *et al.*, 1998b] et décrit dans la section 3.3.1.b)), les utilisateurs sont capables de s'adapter à ces indices "supranormaux" et que leur précision de localisation à l'avant est significativement améliorée.

Dans une tentative de linéariser les relations entre la distance physique et la distance perçue mises en place par [Zahorik *et al.*, 2005] et [Bronkhorst, 2002], l'équipe de Rocchesso a mené plusieurs études dans lesquelles les auteurs exagèrent les indices de réverbération. Dans la première, [Fontana *et al.*, 2002b, Fontana et Rocchesso, 2003] ont testé l'effet de la simulation de la propagation acoustique du son dans un tube. À partir d'une expérience perceptive, les auteurs ont montré que la perception de la distance est meilleure lorsque la source et le récepteur sont placés dans un tube. Étudiée aussi avec des sons réels (donc dans un vrai tube de 10 mètres de long) [Fontana et Rocchesso, 2008], cette méthode consiste à exagérer les indices de réverbération en simulant la propagation d'une onde dans un environnement particulièrement réverbérant et facile à contrôler. Elle a le grand avantage de ne pas être dépendante des indices d'intensité et permet donc de fonctionner avec des sources inconnues. Cette équipe a ensuite testé plusieurs formes 3D et montré que les indices de réverbération fournis par une membrane trapézoïdale ayant des propriétés d'absorption spécifiques à ses limites sont les plus efficaces pour linéariser la relation entre la distance physique et la distance perçue [Devallez *et al.*, 2008].

D'autres études telles que [Lokki et Grohn, 2005], [Picinali *et al.*, 2010] ou [Ortega-González *et al.*, 2010a, Ortega-González *et al.*, 2010b] utilisent une méthode basée sur l'ajout d'effet audio dépendant de la position (ou de la distance) pour améliorer la perception de certains indices. Dans [Lokki et Grohn, 2005], les auteurs utilisent une métaphore de compteur Geiger pour améliorer la perception de la distance et un filtre passe-bande pour l'élévation. [Picinali *et al.*, 2010] quant à eux utilisent la fréquence fondamentale du stimulus pour représenter la distance. Dans leur expérience, plus la distance entre l'objet et la tête est courte, plus le stimulus est aigu. Enfin [Ortega-González *et al.*, 2010a, Ortega-González *et al.*, 2010b] ont cherché à améliorer un jeu d'HRTF avec l'ajout de plusieurs indices. Ainsi, ils ajoutent un filtre passe-bas lorsque les sources sont situées à l'arrière (pour réduire les confusions avant/arrière), un filtre passe-haut lorsque les sources ont une élévation supérieure à 10° et une reverb artificielle lorsque l'élévation est inférieure à $-10°$. Si cette approche permet d'améliorer significativement la localisation, il est regrettable que les auteurs n'aient pas choisi un mapping des paramètres qui soit dépendant de la position. En effet, leur méthode permet à l'auditeur de savoir dans quelle région se trouve la source (avant/arrière et haut/milieu/bas) mais ne donne pas plus d'informations sur la position de cette source dans la zone.

Dans un contexte de guidage en champ proche (pour des distances inférieures à 1.5 mètres), la perception de la distance est très limitée comparée à la précision requise. Au lieu de chercher à linéariser ou à exagérer les indices acoustiques de distance, cette étude a pour but d'explorer l'influence de l'ajout de nouveaux indices acoustiques pour la perception de la distance. Notre approche consiste (comme [Ortega-González *et al.*, 2010a, Ortega-González *et al.*, 2010b]) à simuler de nouveaux indices de distance plutôt que de chercher à simuler les indices physique utilisés en situation réelle. Ceci peut être réalisé grâce à l'utilisation de techniques de sonification.

Nous avons vu dans le chapitre 2.3 les différentes approches de sonification existantes. Dans cette étude, une approche de sonification basée sur un mapping de paramètre a été utilisée. Cette méthode consiste à représenter les variations des données à sonifier par des variations de paramètres acoustiques [Kramer, 1993, Walker et Kramer, 1996]. Les applications basées sur cette approche utilisent en général la fréquence, le tempo, l'amplitude et le timbre comme paramètres principaux de mapping appliqués à des sons de synthèse. Alors que la fonction de transfert entre des données à sonifier et les paramètres de la synthèse sonore sont très faciles à mettre en place, un des problèmes majeur est que les sons produits ne sont pas forcément agréables, et peuvent être agaçants lors d'une utilisation prolongée.

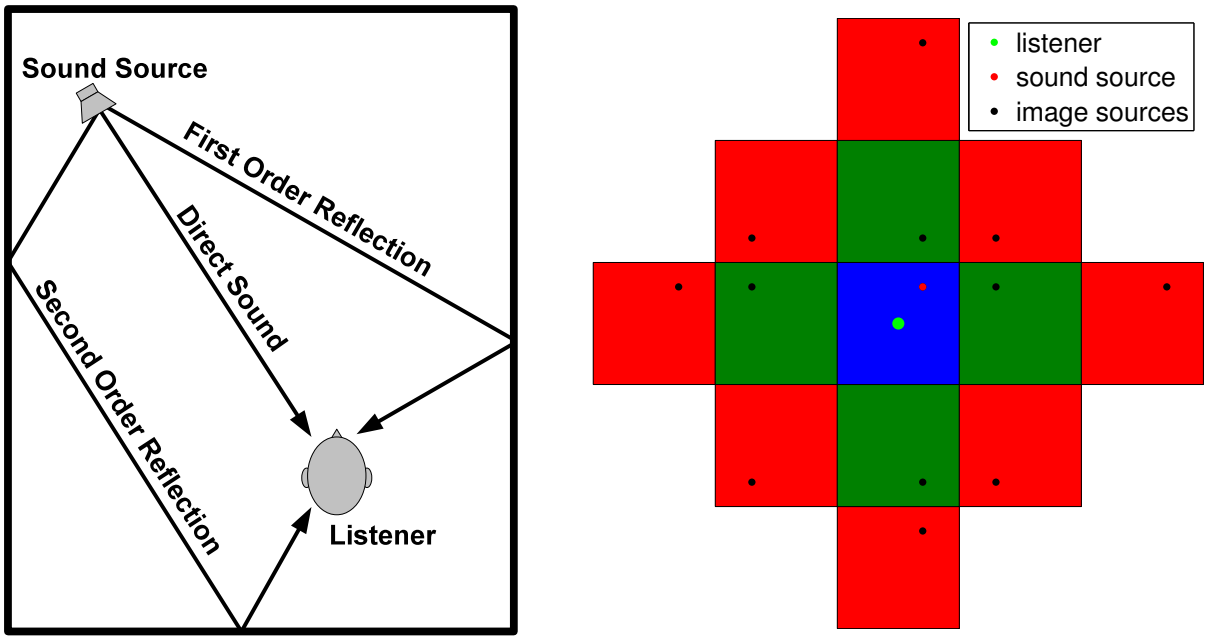
Ces dernières années, malgré le développement de nombreuses interfaces sonores, les notions d'esthétique et d'acceptation par les utilisateurs ont été absents du champ d'investigation de la plupart des recherches. Très peu d'études ont travaillé sur la personnalisation des informations sonores par l'utilisateur et son impact sur l'efficacité et l'efficience du système.

Dans cette étude, le concept de sonification par mapping de paramètres acoustiques est étendu à l'utilisation de tout type de signal audio par un mapping de paramètres d'effets sonores (qui sont ensuite appliqués au son). Avec ce concept, les données à sonifier ne sont plus directement reliées à des paramètres de synthèse sonore mais sont reliées à des variations de paramètres acoustiques des sons via des paramètres d'effets. Cette technique permet d'appliquer les métaphores de sonification à tout type de sons tout en maintenant une cohérence avec les données à présenter.

4.4.2 Métaphores de sonification basées sur des effets audio

Nous avons vu dans la section 5.2.2 que dans un contexte de projet commercial, plusieurs contraintes sont imposées au développement du prototype. L'utilisation de la synthèse binaurale et la variabilité des désidératas des utilisateurs doivent être pris en compte dans la conception de la sonification de la distance.

Dans cette section, nous allons commencer par décrire la technique de sonification développée dans le cadre du projet pour répondre à la demande de personnalisation des sons par les utilisateurs. Puis nous détaillerons les trois métaphores que nous avons développé à partir de ce concept.



(a) (b)

FIGURE 4.9 – (a) Les différents trajets du son dans un environnement clos. (b) Schéma 2D de la méthode source-image. La pièce simulée (en bleu) contient la source (en rouge) et l'auditeur (en vert), les réflexions du premier ordre proviennent des zones vertes, celles du second ordre des zones rouges.

a) Sonification basée sur des effets sonores

Afin de prendre en compte les contraintes imposées par l'utilisation de la synthèse binaurale ainsi que celles imposées par les utilisateurs, la sonification de la distance a été conçue comme un effet audio appliqué au son. Avec ce concept, la distance est reliée à un ou plusieurs paramètres d'un effet audio appliqué au son qui devient donc dépendant de la distance. Cette méthode permet la mise en place de plusieurs métaphores de sonification de la distance, tout en laissant à l'utilisateur la possibilité de personnaliser les sons de l'interface. Elle a en plus l'avantage, une fois que la métaphore est comprise et apprise, de permettre à l'utilisateur de changer de son sans avoir à réapprendre le mapping de la sonification (les variation acoustiques dues au mapping de l'effet restent les mêmes).

b) Réflexions précoces (*ER* pour *Early Reflection*)

Comme nous l'avons vu dans la section 4.2.2, plusieurs études ont mis en évidence l'amélioration de la perception de la distance avec des indices de réverbération [Mershon et King, 1975, Nielsen, 1992]. [Begault, 1992] a quant à lui montré l'avantage de l'utilisation d'une réverbération artificielle dans un environnement audio virtuel. Cet ajout conduit à une meilleure perception de la distance et augmente l'effet d'externalisation du son. Il a par contre un effet néfaste sur la

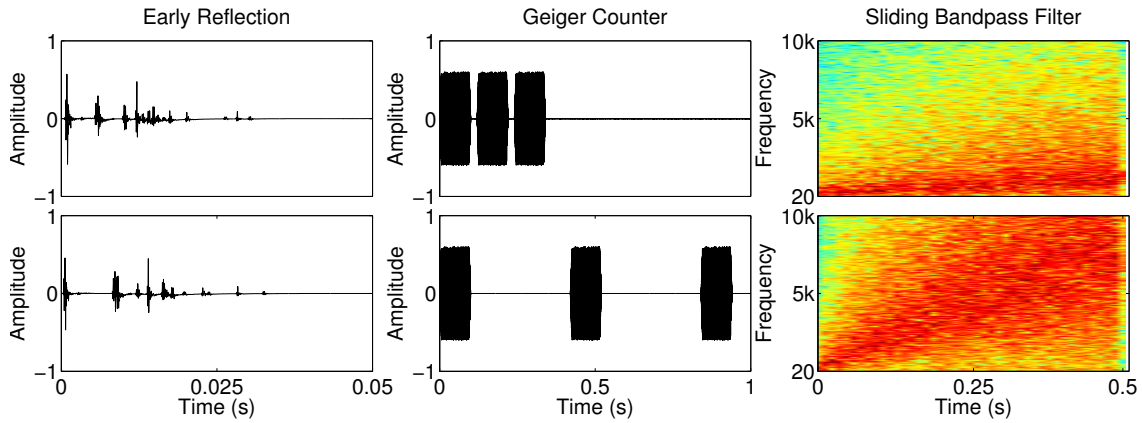


FIGURE 4.10 – Représentations des sons résultants de l'application de chacune des trois métaphores pour deux distances (figures du haut, dist=0.6 m ; figures du bas, dist=1.5 m). Gauche : Réponses impulsionnelles de la métaphore ER. Centre : Représentation temporelle de la métaphore GC appliquée à un "burst" de 10 ms. Droite : spectrogramme du son résultant de l'application de la métaphore SBF à un "burst" de 0.5 sec.

latéralisation des sources étant donné qu'il a tendance à augmenter la largeur apparente de la source.

À partir de la littérature sur la perception de distance de sources proches, nous avons fait l'hypothèse que la perception de distance de sources sonores dans l'espace péripersonnel peut-être améliorée par l'ajout des réflexions précoces [Kearney *et al.*, 2012]. Le concept de cette métaphore est donc de créer un effet basé sur la simulation de réflexions précoces spatialisées calculées au deuxième ordre (i.e., se réfléchissant sur un ou deux murs en considérant une source omnidirectionnelle, voir figure 4.9) pour une salle donnée. Afin de ne pas dégrader les performances de localisation en azimut, nous avons décidé de ne simuler que les premières réflexions. Une méthode de source-image a été utilisée pour calculer les réflexions précoces [Allen et Berkley, 1979]. Chaque réflexion (appelée source-image) est un double de la source sonore, arrivant d'une position différente et atténuée en fonction de la distance parcourue et des obstacles rencontrés. L'ajout de ces réflexions permet une multiplication des informations spatiales à travers la spatialisation binaurale du son direct et de chaque source-image.

Pour l'expérience, les réflexions ont été construites à partir d'une salle de taille $5 \times 5 \times 3m^3$. La tête de l'auditeur est placée au centre de cette salle virtuelle à une hauteur de 1m40. 24 sources-images (6 du premier ordre et 18 du second ordre) sont nécessaires pour simuler les réflexions du premier et second ordre. Leurs positions sont calculées en temps réel en fonction de la position de la source. Chaque source-image est filtrée une ou deux fois (en fonction du nombre de murs rencontrés), puis retardée en fonction de la différence entre la longueur de son trajet vers l'auditeur et la longueur du trajet du son direct. Pour le filtrage, nous avons utilisé un filtre passe-bas à réponse impulsionnelle infinie d'ordre 2. Ces coefficients ont été calculé de manière à obtenir une absorption correspondant à un bureau traditionnel. Afin de réduire la charge de calcul imputée à l'ordinateur pour la synthèse binaurale, les 24 sources ont été spatialisées en utilisant un système ambisonique

du troisième ordre diffusé sur 12 haut-parleurs virtuels. Ces haut-parleurs virtuels entourant l'auditeur étaient spatialisés en binaural aux positions classiques sur une sphère (pour plus de détails voir [McKeag et McGrath, 1996, Noisternig *et al.*, 2003]). Le signal résultant est combiné au son direct spatialisé en binaural. La figure 4.10 (gauche) représente une réponse impulsionnelle de cette métaphore pour deux distances différentes (0.6 m et 1.5 m).

c) Compteur Geiger (*GC* pour *Geiger Counter*)

Une des premières applications utilisant la sonification est le compteur Geiger inventé par Hans Geiger au début des années 1900. Il consiste à avertir un utilisateur de la présence et de l'intensité de radiations invisibles en générant un "beep" dont la fréquence de répétition augmente proportionnellement avec l'intensité des radiations. Cette méthode de sonification bien connue, qui a été testée avec succès dans un grand nombre d'applications de sonification, est maintenant bien intégrée dans la vie de tous les jours et utilisée dans un certain nombre d'applications commerciales. Elle déjà utilisée dans les voitures pour aider le conducteur à éviter les obstacles lorsqu'il se gare. Plus la voiture se rapproche d'un objet, plus la fréquence de répétition du signal sonore augmente.

Afin d'améliorer la perception de la distance, cet effet consiste à répéter trois fois le stimulus et à faire varier l'intervalle temporel entre chaque répétition en fonction de la distance. Plus la distance est proche, plus l'intervalle temporel est court.

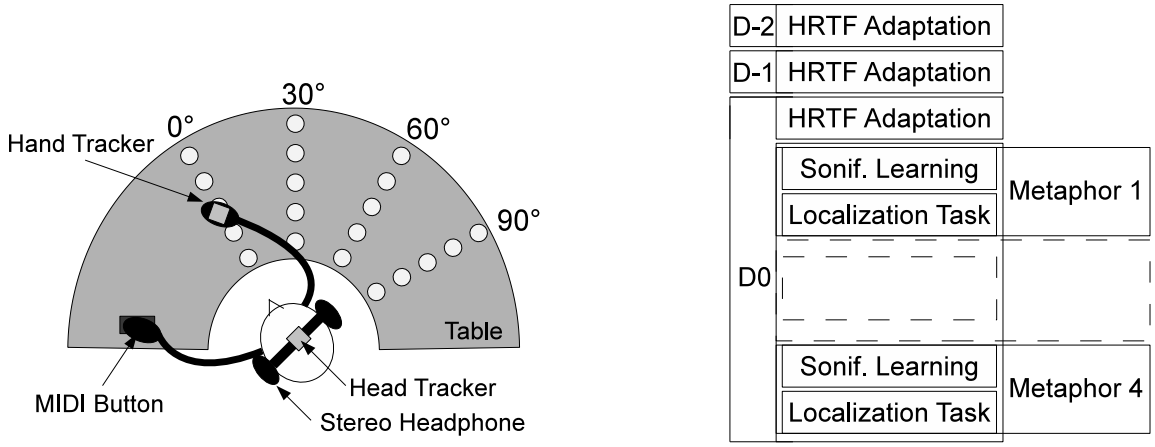
Pour cette expérience, nous avons choisi un mapping qui permet d'éviter un recouvrement entre les sons et qui crée des variations suffisamment perceptibles. Les intervalles de répétitions sont donc de 20 ms pour une distance de 0.6 mètre et de 320 ms pour une distance de 1.5 ms, l'évolution entre ces deux distances est linéaire. Le signal sonore résultant de l'application de cette métaphore sur un "burst" de 10 ms pour deux distances (0.6 m et 1.5 m) est présenté au centre de la figure 4.10.

d) Filtre passe-bande glissant (*SBF* pour *Sliding Bandpass Filter*)

Plusieurs études ont montré que l'utilisation de la fréquence pour la sonification de données était facilement compréhensible et efficace [Brown *et al.*, 2003]. L'idée de cette métaphore est de transposer ce concept de sonification (généralement appliqué à des sons de synthèse) en un effet audio applicable à tout type de son.

Cet effet est créé en utilisant un filtre passe-bande dont la fréquence centrale et la bande passante, varient en fonction du temps. La variation de la bande passante est calculée de façon à ce que le facteur de qualité du filtre reste constant $Q = \Delta f / f$ (où Δf est la largeur de bande et f la fréquence centrale).

La fréquence centrale initiale du filtre (à T=0 sec, début du son) est fixée à 200 Hz indépendamment de la distance. La fréquence centrale finale du filtre (à T= à la longueur du son) augmente proportionnellement par rapport à la distance. Avec cet effet, un "burst" de bruit sera perçu comme un "chirp" bruité dont la fréquence finale dépend de la distance.



(a) (b)

FIGURE 4.11 – (a) Configuration du système expérimental. Les petits cercles correspondent à la position des sources (b) Déroulé de l'expérience.

Pour l'expérience, le facteur de qualité a été fixé à $Q = \Delta f/f = 2$, et la fréquence finale à 1 kHz pour les sources placées à 0.6 m et 8 kHz pour les sources placées à 1.5 m. L'évolution de la fréquence centrale finale entre ces deux distances est linéaire. Le choix de la polarité du mapping a été réalisé après plusieurs pré-tests. Si ce mapping paraît non intuitif aux acousticiens, il n'a pas dérangé les sujets qui l'associaient à une "distance fréquentielle" plutôt qu'à une analogie au filtrage spectral dû à l'absorption par l'air. La figure 4.10 (droite) présente les spectrogrammes des sons résultants de cette métaphore appliquée à un bruit blanc de 0.5 sec pour deux distances différentes (0.6 m et 1.5 m).

4.4.3 Expérience

a) Participants

16 sujets adultes rémunérés (3 femmes et 13 hommes, moyenne d'âge 28 ± 6 ans) ont participé à l'expérience. Un audiogramme a été effectué sur chaque sujet afin de s'assurer qu'il n'avait aucun déficit auditif (audition $>$ 20 dB). Ils ne connaissaient ni le but de l'expérience, ni la répartition spatiale des sources sonores choisies pour l'expérience.

b) Matériel et méthode

Le schéma du déroulement de l'expérience est présenté figure 4.11(b), avec le déroulement temporel de la procédure expérimentale. Les trois premières phases avaient pour but d'adapter le sujet aux HRTF non-individuelles qui seront décrites dans la section 4.4.3.c). Elles utilisaient la plateforme audio-kinesthésique décrite dans le chapitre 3. Les phases suivantes correspondent à

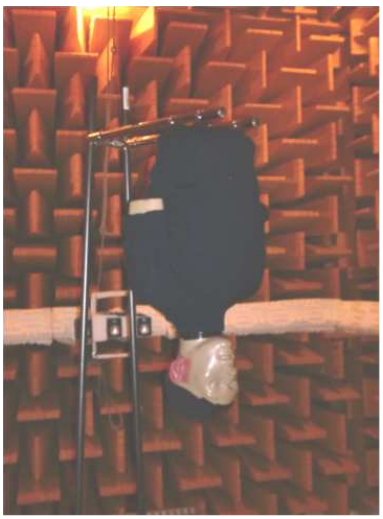


FIGURE 4.12 – Photo de l'installation du mannequin KEMAR à l'envers dans la chambre anéchoïque de l'IRCAM pour réaliser la mesure des HRTF comprises entre 0 et $-90°$ d'élévation.

l'évaluation de chaque métaphore de sonification de la distance avec une tâche de localisation. Pour ces phases, le sujet était assis sur une chaise pivotante placée au centre d'un plateau en bois circulaire de 90 cm de diamètre (voir figure 4.11(a)).

Les sujets portaient un casque stéréo ouvert (modèle Sennheiser HD570) équipé d'un capteur de position et d'orientation magnétique placé sur le dessus du casque. Ils tenaient un capteur de position dans leur main dominante et interagissaient avec le système en pressant sur un bouton midi positionné dans leur autre main. Pour les différentes phases du test, la position des sons était générée par rapport au centre de la tête (en prenant comme centre du référentiel la position du capteur placé sur le casque, décalé au centre de la tête). Aucune égalisation de la réponse fréquentielle du casque n'a été effectuée.

Le stimulus utilisé était spatialisé en utilisant des HRTF non-individuelles mesurées sur un mannequin KEMAR dont les particularités sont décrites dans la section 4.4.3.c). Il était suffisamment court pour éviter que le sujet bouge la tête pendant sa présentation et consistait en une répétition de trois "burst" de bruit gaussien large bande (50–20000 Hz) fenêtrés avec des rampes de Hamming de 2 ms au début et à la fin afin d'éviter les "clics". Un silence de 30 ms est inséré entre chaque "burst". Le niveau global mesuré dans l'oreille pour une source binaurale positionnée en face du sujet à 50 cm (0° en azimut et 0° en élévation) était de 60 dBA.

c) HRTF utilisées

Les HRTF du mannequin KEMAR ont été mesurées dans la chambre anéchoïque de l'IRCAM. Afin de pouvoir générer toutes les positions du test, il a été nécessaire de mesurer ces HRTF sur toute la sphère (alors qu'habituellement, les mesures ne descendent pas en dessous de $-40°$). Pour ce faire, nous avons réalisé deux séries de mesures qui ont ensuite été combinées : une avec le buste à l'endroit et une avec le buste à l'envers (photo figure 4.12). Les HRTF résultantes contiennent

des mesures de $-90°$ à $+90°$ en élévation par pas de $5°$ et de $-180°$ à $+180°$ en azimut par pas de $15°$. Hormis une précision plus fine en élévation, ces mesures ont les mêmes caractéristiques que les HRTF de la base [LISTEN, 2003]. Étant donné que nous ne disposions que d'un seul jeu d'HRTF mesuré sur toute la sphère, la modélisation de l'ITD n'a pas pu être réalisée pour cette expérience. Contrairement à l'expérience du chapitre 3, l'expérience décrite ici a été réalisée avec des indices ITD non-individualisés ce qui (nous le verrons plus loin) influe sur les performances en azimut. Les indices de distances synthétisés pour cette expérience étaient : l'intensité (qui décroit en $1/r^2$ avec l'augmentation de la distance), l'ILD (qui est corrigé pour les distances proches) et l'effet de parallaxe (en dessous de 1 mètre, l'angle de la source n'est plus indépendant de la distance. Le moteur de synthèse utilise alors les HRTF correspondants aux angles entre la source et chacune des deux oreilles).

d) Procédure

L'expérience de localisation était divisée en quatre blocs de 80 positions, chaque bloc correspondant à une condition de sonification de distance. La durée de chaque bloc était d'environ 15 minutes. Afin d'évaluer l'effet d'amélioration de perception de la distance, les trois conditions de sonification ont été comparées à une condition contrôle générée seulement avec de la synthèse binaurale et servant de référence pour les performances de localisation. Les quatre blocs sont appelés *Control* (pour le blocs sans sonification), *Geiger Counter (GC)*, *Sliding Band-pass Filter (SBF)* et *Early Reflections (ER)*. Pour chaque sujet les blocs étaient présentés dans un ordre aléatoire afin de contrebalancer un effet potentiel d'apprentissage de la tâche. Chaque bloc débutait par une petite session d'apprentissage de la sonification consistant à habituer le sujet à la métaphore de distance. À cet effet, un son était répété toutes les 2 sec et positionné virtuellement sur la main du sujet qui pouvait donc interagir avec la distance via un processus audio-kinesthesique. Premièrement, le sujet devait déplacer sa main dominante sur le plateau devant lui de l'intérieur vers l'extérieur puis revenir (2 fois) et faire la même chose dans une autre position. Cette partie permettait de s'assurer que le sujet prenne conscience de toute la gamme de distances et l'associe aux différences sonores. Il avait ensuite 45 sec pour explorer le plateau comme il le désirait.

La tâche de localisation consistait à reporter la position perçue d'un son spatialisé statique en utilisant une technique de pointage avec la main validée par bouton midi. Chaque sujet devait s'orienter dans une direction donnée et garder sa tête fixe dans une position de référence correspondant au centre du système, 0.65 mètres au dessus du plateau, pendant la présentation du stimulus. Avant chaque essai, l'orientation de la tête du sujet était contrôlée automatiquement et il lui était demandé de corriger sa position si le décalage dépassait $5°$. Après la présentation du stimulus, les sujets devaient pointer leur main dans la direction perçue du son à localiser et valider leur réponse avec le bouton midi. Les sujets étaient placés dans le système de façon à pointer avec leur main dominante. La position perçue était calculée à partir du couple position/orientation de la tête lors de la présentation du stimulus et de la position de la main lors de la validation de la position perçue. Aucun feedback n'était fourni au sujet sur la position des cibles sonores.

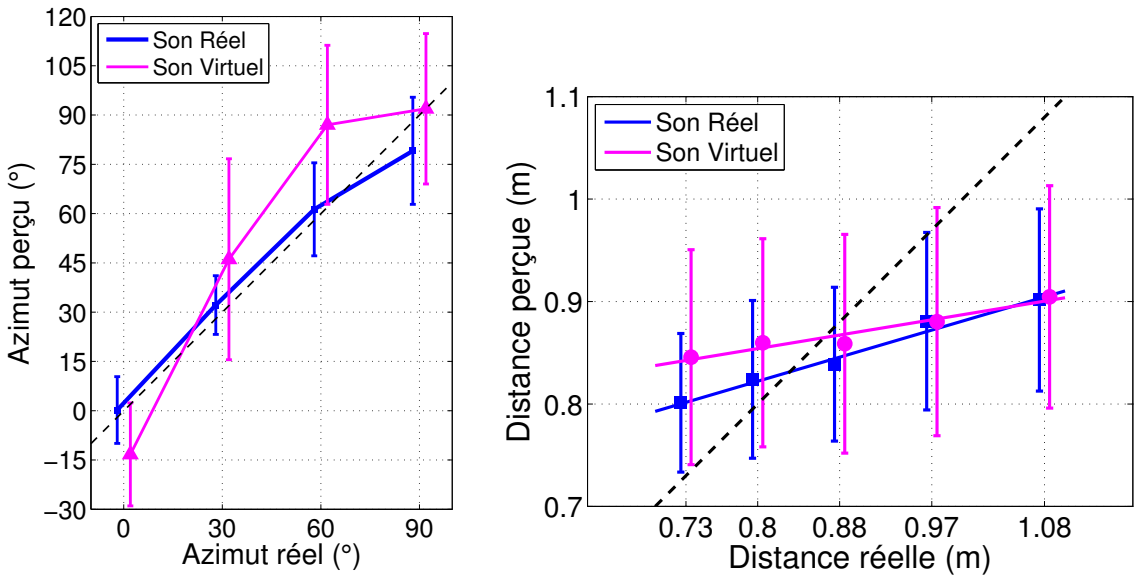


FIGURE 4.13 – Gauche : Comparaison de l'azimut perçu en fonction de l'azimut réel pour la localisation de sons réels et virtuels ; Droite : Distance perçue en fonction de la distance réelle pour la localisation de sons réels et virtuels.

Chaque bloc était composé de 20 positions (5 distances par rapport à la tête : 0.73 m, 0.80 m, 0.88 m, 0.97 m, and 1.07 m et 4 azimut : 0°, 30°, 60° et 90°, voir figure 4.11) présentés de façon aléatoire 4 fois chacune.

e) Résultats

La contribution de chaque métaphore de sonification sur la distance perçue a été analysée en comparant les erreurs de distance et d'azimut pour ces métaphores aux erreurs obtenues pour la condition *Control*. Étant donné que certains participants ont rencontré des problèmes de validation sur certains essais, toutes les positions validées en dehors du plateau expérimental ont été supprimées (cela représente 2% du nombre total d'essais). Sur l'ensemble des essais et des conditions, 7.8% confusions avant/arrière ont été observées pour les sources positionnées à 30° et 60°. Étant donné que nous sommes surtout intéressés par l'évolution des performances en distance, ces confusions ont été corrigées.

i/ Comparaison des performances sons réels, sons virtuels

Avant d'analyser l'effet des métaphores de sonification sur l'amélioration de la perception de la distance, commençons par comparer les résultats obtenus dans cette expérience avec la condition *Control* correspondant au binaural sans sonification de la distance, aux résultats de l'expérience sur les sons réels (décrite dans la section 4.3) dans la condition de pointage avec la main préférée. La figure 4.13 présente l'évolution de l'azimut perçu en fonction de l'azimut réel, et de la distance perçue en fonction de la distance réelle pour la localisation de sons réels et de sons virtuels.

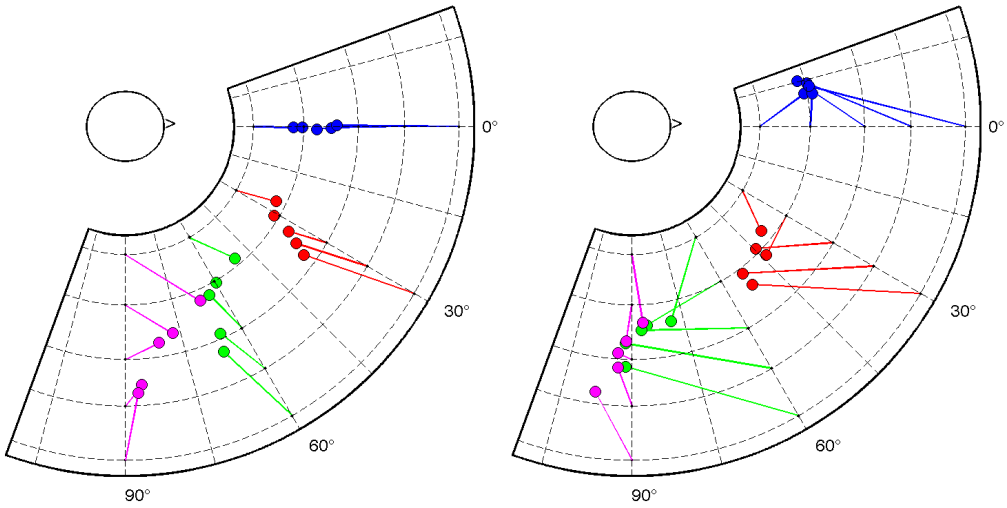


FIGURE 4.14 – Représentation des positions moyennes pointées sur le plateau pour les sons réels (à gauche) et les sons virtuels (à droite).

La figure de gauche (sur l'azimut), met en évidence une grande différence de performances entre les deux conditions. Dans le cas des sons réels, la moyenne d'erreur absolue globale est de $9.3 \pm 9.4°$ alors que dans le cas des sons virtuels, elle est de $23.7 \pm 17.5°$. Cette dégradation des résultats n'est pas surprenante au regard de la littérature (cf. section 2.2.3.b)), sachant aussi que les indices ITD des HRTF du mannequin KEMAR n'étaient pas individuels. Le résultat surprenant est par contre la différence d'évolution de l'erreur signée entre les deux conditions. À 0°, cette erreur est de 0.2° pour les sons réels alors qu'elle est de $-13.3°$ pour les sons virtuels, tandis qu'à 90°, nous avons 10.9° d'erreur pour les sons réels et 1.9° pour les sons virtuels (avec une variance toujours beaucoup plus élevée dans le cas de la localisation de sons virtuels). L'évolution de la courbe de perception de l'azimut pour les sons virtuels met en évidence un gros décalage des sources vers les côtés. Ce décalage est dû à la non-individualisation des indices ITD ainsi qu'au problème de localisation intracrânienne qui a été rapporté par la majorité des sujets (les sujets n'entendent pas le son devant eux mais dans leur tête). Cette latéralisation explique que l'erreur signée à 90° est quasiment nulle alors qu'elle est assez grande pour les sons réels. Le décalage de 13° du côté opposé aux sources à localiser apparu pour les sources de la ligne médiane avec les sons virtuels est plus difficile à expliquer. Nous avons déjà remarqué dans la section 4.3 un léger décalage des positions perçues pour la ligne médiane avec les sources réelles lorsque nous ne prenions en compte qu'un seul bloc de localisation (avec des sources sur un seul côté). Ce décalage pour la localisation de sons virtuels a pu être amplifié par le manque d'externalisation de la source.

Concernant la résolution des erreurs dues au cône de confusion, nous avons vu dans la section 2.2.3.b) que la synthèse binaurale avec des HRTF non-individuelles entraînait une augmentation des confusions avant/arrière. Cette constatation est bien vérifiée dans le cas de notre expérience avec un taux de confusion général passant de 0.7% pour les sons réels à 9.6% pour les sons virtuels (avec une augmentation de 3.5% à 33.3% à 60°).

Condition	Control	GC	SBF	ER
Regression slope	0.14 (.09)	0.64 (.29)	0.50 (.20)	0.06 (.13)
Goodness-of-fit	0.66 (.26)	0.96 (.05)	0.92 (.12)	0.27 (.31)

TABLE 4.3 – Moyennes des coefficients directeurs des droites de régression linéaire et coefficients de qualité de l'ajustement r^2 pour chaque métaphore. Variance montrée entre parenthèse.

La figure 4.13 (à droite) met en évidence une augmentation de l'erreur en distance pour les cibles virtuelles avec une variance beaucoup plus grande. L'erreur globale qui était de 9.8 cm (soit une erreur d'environ 11.0% par rapport aux distances à localiser) pour les sons réels passe à 12.0 cm (soit 13.5%) pour les sons virtuels. On observe pour la synthèse binaurale la même augmentation des performances pour les positions latérales que pour les sons réels (erreur en distance dans la zone frontale ($\leq|30°|$) de 13 cm et de 11 cm dans la zone latérale ($>|30°|$) pour les sons virtuels, contre respectivement 10.6 cm et 8.6 cm pour les sons réels). Une régression linéaire sur l'ensemble des résultats en distance, de la localisation des sons virtuels, donne un coefficient de régression de 0.16 ± 0.09 avec une qualité d'ajustement de 0.65 (contre 0.29 ± 0.11 et 0.95 pour les sons réels). Ce résultat montre qu'avec les sons virtuels, la distance n'est quasiment pas perçue par les sujets. Ce constat est confirmé par l'analyse de la position moyenne perçue sur le plateau représentée pour les deux conditions figure 4.14.

ii/ Effet des métaphores sur la distance perçue

La figure 4.15 montre la moyenne des distances perçues en fonction de la distance de la source calculée sur tous les sujets pour chaque condition. Elle met en évidence une tendance à surestimer la distance des sources les plus proches et à sous estimer celle des sources les plus éloignées. Les résultats obtenus pour les conditions *control* et *early reflection* sont plus mauvais que ceux obtenus pour les conditions *geiger counter* et *sliding bandpass filter*. Les droites de la figure correspondent aux régressions linéaires effectuées sur les résultats. Les régressions linéaires ont été effectuées sur les résultats de chaque sujet, pour chaque condition. La moyenne (et l'écart type) de coefficients directeurs obtenus ainsi que les critères de qualité de l'ajustement r^2 sont donnés pour chaque condition dans le tableau 4.3. Ces coefficients directeurs sont loin de l'unité attendue dans le cas d'une perception parfaite de la distance de la source virtuelle. Pour les conditions *control* et *early reflection*, les résultats montrent que les sujets n'ont fait aucune distinction entre les différentes distances et que quelque soit la position de la cible, le son était perçu au milieu du plateau. Les résultats des conditions *sliding bandpass filter* et *geiger counter* mettent en évidence une amélioration de la perception de la distance avec des coefficients directeurs plus élevés. Par contre, ces résultats montrent une variabilité inter-sujet plus large (mise en évidence par les grands écarts types).

Ces résultats sont confirmés par les boxplots de l'erreur relative de distance de la figure 4.16. En effet, la moyenne d'erreur obtenue pour les *geiger counter* et *sliding bandpass filter* est approximativement inférieure de 5 cm à celles obtenues pour les conditions *control* et *early reflection*. Une

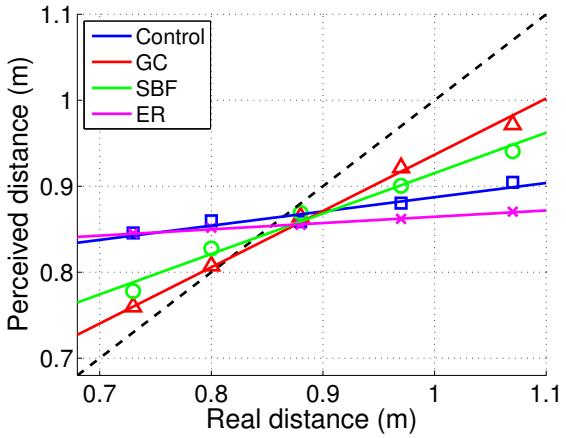


FIGURE 4.15 – Distance perçue en fonction de la distance réelle pour chaque condition de sonification. «□, △, ○, ×» : Moyenne de la distance perçue selon chaque condition. Lignes : Moyenne de la régression linéaire.

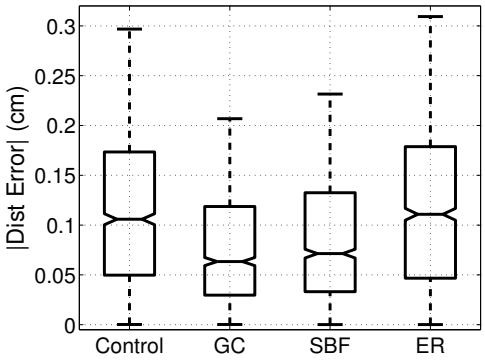


FIGURE 4.16 – Boxplot de la valeur absolue de l'erreur de distance pour chaque métaphore.

ANOVA à mesures répétées a été effectuée sur la moyenne des erreurs de distance, en prenant en compte trois facteurs intra-sujets : la condition (facteur fixe de 4 niveaux), la distance des cibles (facteur fixe de 5 niveaux) et l'azimut des cibles (facteur fixe de 4 niveaux). Cette analyse met en évidence un effet significatif de la condition ($F(3,42) = 19.76, p < 0.001$), de la distance ($F(4,56) = 12.01, p < 0.001$) et de l'azimut ($F(3,42) = 9.32, p < 0.001$). Un test de Duncan sur les conditions montre une différence significative entre les conditions *control* et *geiger counter* ($p = 6.10^{-5}$) ainsi qu'entre les conditions *control* et *sliding bandpass filter* ($p = 2.10^{-4}$). La comparaison entre les conditions *control* et *early reflection* ne montre aucune différence entre les résultats de ces deux conditions ($p = 0.59$). Pour les positions des sources, un test de Duncan sur la distance montre une différence significative entre l'erreur pour la source la plus éloignée et les autres (mettant en évidence de moins bons résultats pour les plus grandes distances). Un test de Duncan sur l'azimut montre une différence significative des résultats entre les sources positionnées à 90° et les autres (mettant en évidence de meilleures performances pour les sources latérales).

Une étude plus approfondie de l'erreur sur la distance perçue en prenant en compte l'azimut de

Angle	0°	30°	60°	90°
Control	0.13 (.10)	0.11 (.09)	0.11 (.08)	0.09 (.08)
GC	0.07 (.08)	0.06 (.07)	0.06 (.07)	0.06 (.06)
SBF	0.09 (.09)	0.08 (.08)	0.07 (.07)	0.06 (.06)
ER	0.11 (.10)	0.10 (.09)	0.12 (.09)	0.10 (.08)

TABLE 4.4 – Moyennes de l'erreur en distance (en mètre) par angle et par métaphore. Variance montrée entre parenthèse.

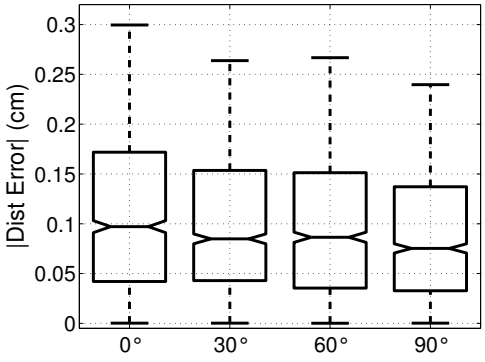


FIGURE 4.17 – Boxplot de la valeur absolue de l'erreur de distance pour toutes les conditions réunies en fonction de l'angle de la source.

la source est montrée pour toutes les conditions réunies figure 4.17 et pour chaque condition dans le tableau 4.4. On remarque que la perception de la distance est plus précise pour les sources latérales que pour les sources frontales. Les résultats du tableau 4.4, montrent que cette amélioration des performances pour les sources latérales apparaît surtout pour les conditions *control* et *sliding bandpass filter* (mais avec un écart type très grand).

Ces résultats sont confirmés par l'analyse de la position moyenne des réponses pour chaque condition, présentées figure 4.18

iii/ Effet des métaphores sur l'azimut perçu

Bien que ce ne soit pas le premier but de notre étude, il est intéressant de regarder l'effet de la condition sur la perception de l'azimut afin de s'assurer que les métaphores de sonification ne dégradent pas trop la latéralisation de la source. La figure 4.19 montre la moyenne sur tous les sujets de l'azimut perçu en fonction de l'azimut de la cible pour toutes les conditions. Elle met en évidence une grande déviation standard (principalement à 30° et 90°) et un décalage vers les angles négatifs pour les sources frontales.

En analysant chaque condition, il apparaît qu'aucune métaphore n'affecte la perception de l'azimut hormis la condition *early reflection* pour les sources latérales.

L'erreur moyenne sur l'azimut est de $20 \pm 15°$. Une ANOVA à mesures répétées sur l'erreur en azimut pour chaque métaphore en combinant toutes les positions ne montre pas d'effet significatif

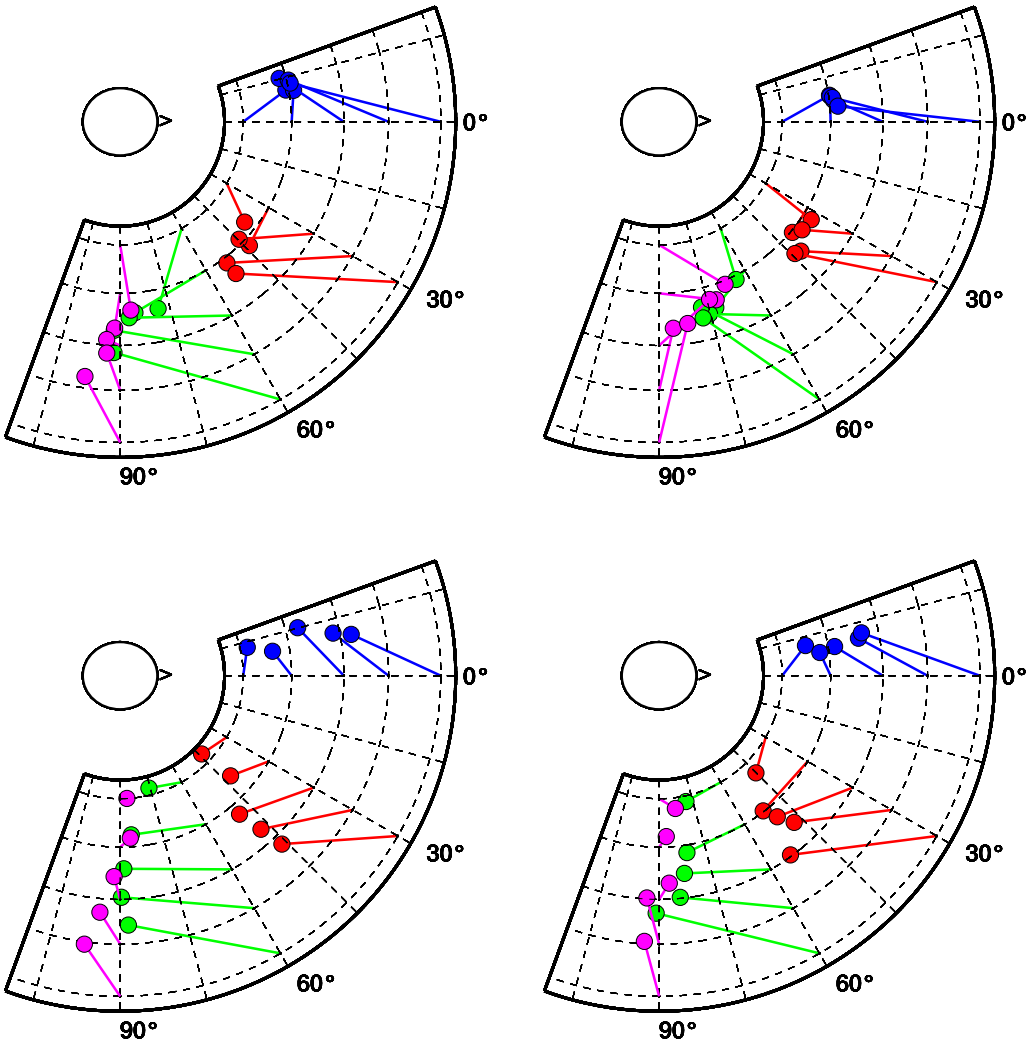


FIGURE 4.18 – Représentation des positions moyennes pointées sur le plateau en fonction de la condition : en haut, à gauche : *Control* ; en haut, à droite : *Early Reflection* ; en bas, à gauche : *Geiger Counter* ; en bas, à droite : *Sliding Bandpass Filter*.

de la condition ($F(3, 45) = 0.206, p = 0.89$).

4.4.4 Discussion

Au regard de ces résultats, sur les trois métaphores mises en places, seules deux ont été efficaces (les métaphores *geiger counter* et *sliding bandpass filter*). Comparées au rendu avec du binaural seul (condition pour laquelle la perception de la distance est nulle pour la configuration de cette expérience), ces deux métaphores ont permis d'améliorer significativement la perception de la distance. La supériorité de la métaphore *geiger counter* sur la métaphore *sliding bandpass filter* peut être expliquée par le mapping utilisé pour chacune des conditions. En effet, pour la condition *sliding bandpass filter*, nous avons utilisé un mapping linéaire alors que la perception des fréquences est

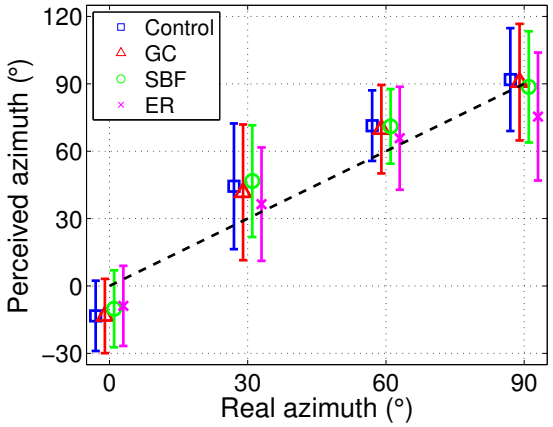


FIGURE 4.19 – (a) Azimut perçu en fonction de l'azimut réel pour chaque condition de sonification. «□, △, ○, ×» : Moyenne selon chaque condition. Lignes verticales : Déviation standard selon chaque modalité. Pour un souci de lisibilité, les résultats pour chaque condition ont été légèrement décalés en abscisse.

plutôt logarithmique. Il semble que l'étendue de variations des fréquences n'était pas suffisamment grande pour un rendu complet de la distance.

Contrairement à ce qui était attendu, la métaphore *early reflection* n'a pas permis d'améliorer la perception de la distance et a même dégradé les performances (comparée à la condition *control*). De plus, cette métaphore a tendance à dégrader la localisation directionnelle (surtout à 90°), ce qui n'est pas le cas avec les autres conditions. Pour expliquer ce résultat, plusieurs hypothèses peuvent être faites. Tout d'abord, le modèle choisi basé uniquement sur la simulation des réflexions du premier et du deuxième ordre est peut-être trop simple et l'absence de réverbération a affecté la perception en créant une situation d'écoute anormale. Deuxièmement, les études ayant constaté une amélioration de la perception de la distance avec l'ajout de réflexions précoces ont toutes été réalisées pour des distances supérieures à 1 mètre. Il est possible que cet indice ne soit pas effectif pour des distances plus faibles ou qu'il ne soit pas assez saillant pour de faibles variations de distance.

Pour toutes les conditions (et principalement pour les conditions *control* et *sliding bandpass filter*), l'erreur de distance est plus faible pour les sources latérales (surtout à 90° d'azimut). Cette amélioration, apparaissant pour toutes les conditions, semble être spécifique au rendu binaural. En effet, pour la configuration de cette expérience, la distance était connectée à l'élévation (étant donné que les sujets étaient positionnés 0.65 m au dessus du plateau). L'élévation variait donc entre $-37°$ pour la plus grande distance et $-63°$ pour la plus petite. Pour ces élévations, l'influence du torse est beaucoup plus grande pour les sources latérales que pour les sources frontales. Ceci a probablement ajouté des indices permettant la perception de la distance. Ces résultats sont confirmés par les résultats de l'étude de [Kopčo et Shinn-Cunningham, 2011] qui ont montré de meilleures performances pour la perception de la distance pour les sources latérales avec des sons réels. Ce résultat est principalement expliqué par les variations de l'ILD en fonction de la distance

pour les sources latérales (variations principalement dues aux effets d'ombre de la tête) et à l'absence de variations pour les sources frontales (étant donné que l'ILD est nul).

Hormis pour la condition *early reflection* à 90°, les performances de localisation en azimut n'ont pas été affectées par les métaphores de distance. L'erreur en azimut en moyenne égale à 20° est plus grande que pour les sources réelles (pour des sources situées entre 0.5 et 1 m avec des élévations inférieures à −20°, [Brungart *et al.*, 1999] obtient des erreurs de l'ordre de 10°). Ces performances ne sont pas si mauvaises si on considère le jeu de mesures HRTF utilisé pour générer le rendu binaural de cette expérience. Ces mesures étaient tout d'abord non-individuelles et étaient espacées de 15° en azimut.

Étant donné que le dispositif utilisé pour cette expérience diffère des études de la littérature, une comparaison précise des résultats est impossible. Dans leur étude sur la localisation de sources proches dans un environnement anéchoïque, [Brungart *et al.*, 1999] obtiennent des performances de perception de la distance allant d'un coefficient de régression de 0.3 pour les sources frontales à 0.8 pour les sources latérales. En simulant des sources proches avec des réponses impulsionnelles de salle binaurale enregistrées dans un environnement reverbérant, [Kopčo et Shinn-Cunningham, 2011] obtiennent de meilleures performances avec des coefficients de régression de 0.6 pour les sources frontales et de 0.8 pour les sources latérales. Dans cette étude, il n'y a pas eu de perception de la distance pour la condition avec du rendu binaural seul (coefficient de régression de 0.14). Ceci peut être expliqué par les HRTF non-individuelles utilisées qui ont été mesurées à une distance de 2 m et qui ne contiennent donc aucun indice de champ proche. Avec les métaphores *geiger counter* et *sliding bandpass filter* les performances de localisation (avec des coefficients de régression de 0.64 et 0.50) approchent les performances obtenues avec des sons réels par [Brungart *et al.*, 1999], mettant ainsi en évidence l'efficacité de la méthode adoptée pour la sonification.

4.5 Conclusion

Le but de ce chapitre était d'étudier les capacités de substitution de la vision en champ proche avec des sons réels puis avec des sons virtuels et de trouver un moyen d'augmenter la précision de perception de la distance dans cette zone, ceci avec des sons très courts.

L'expérience réalisée avec les sons réels a permis de quantifier les performances "idéales" qu'il est possible d'atteindre avec un tel système si le sujet a la tête fixe. Cette expérience a aussi mis en évidence que la main utilisée pour l'action de pointage vers l'objet n'a quasiment pas d'influence. Si dans le cas "idéal" les performances de localisation en azimut sont très bonnes, il n'en n'est pas de même pour la perception de la profondeur pour laquelle nous avons observé une grande compression de la distance perçue et une grande variabilité dans les résultats.

L'utilisation de sons virtuels pour un tel système a pour effet d'augmenter l'erreur de perception de la direction et de quasiment anéantir les capacités de perception de la distance (ce qui semble normal étant donné que la synthèse est réalisée dans des conditions anéchoïques pour des sources

très proches dont la distance varie assez peu). Dans cette étude, nous avons utilisé des stimuli très courts afin de nous placer dans un cadre fondamental. Dans le cadre d'un système d'aide à la saisie d'objet en champ proche, les stimuli pourront être répétés à une fréquence définie par l'utilisateur jusqu'à ce qu'il ait attrapé l'objet cible. Cette répétition permettra à l'utilisateur de mettre en relation les variations d'indices acoustiques en fonction de ses mouvements, et donc de localiser les cibles avec plus de précision. Il est toutefois intéressant avant de passer à cette phase, d'essayer d'acquérir les performances maximales avec des sons très courts.

Afin d'augmenter la précision de perception de la distance, nous avons mis en place plusieurs métaphores de sonification de la distance pour un environnement audio virtuel. Pour répondre aux besoins des utilisateurs, la technique de sonification employée doit être indépendante du type de son et facile à apprendre. Sur la base de ces contraintes, un nouveau concept de sonification a été introduit. Ce concept consiste à appliquer un effet audio (dont les paramètres varient en fonction des données à sonifier) au son choisi par l'utilisateur pour représenter les objets qu'il souhaite attraper. Avec cette méthode, les informations sont transmises par les variations dues à l'effet et non par le son lui-même.

Sur la base de ce concept, trois métaphores de sonification de la distance ont été créées et évaluées, de la même façon que la localisation des sons réels, avec une expérience de localisation sonore. Les résultats de cette expérience, comparés aux résultats de localisation avec des sons virtuels (synthèse binaurale sans ajout de sonification) ont permis de mettre en évidence une amélioration significative de la perception de la distance pour deux des trois métaphores réalisées (le *Compteur Geiger* et le *Filtre Passe-bande Glissant*) et ce malgré un très court apprentissage de ces métaphores (une minute d'entrainement audio-kinesthésique). Ces deux métaphores ont même permis d'obtenir de meilleurs résultats que pour le cas "idéal" (i.e. les sons réels).

Ces deux métaphores permettant d'améliorer la perception de la distance en champ proche, cette étude montre l'équivalence entre le concept de métaphores basées sur des effets audio et la traditionnelle sonification par mapping de paramètres qui est appliquée à des sons de synthèse. Ce résultat est positif au regard de l'acceptation par les utilisateurs de la sonification qui est souvent critiquée pour son défaut d'esthétique.

Étant donné que cette étude s'est focalisée sur l'efficacité des métaphores basées sur des effets appliquées à des "sons de laboratoire" ("burst" de bruit blanc), de nouvelles expériences sont requises pour valider leur efficacité avec des sons réels (sons naturels, électroniques ou instrumentaux) afin d'approcher la situation réelle du système d'aide à la navigation et de vérifier si ce concept de sonification comble les attentes des utilisateurs.

Chapitre 5

Les morphocons : une sonification personnalisable basée sur des earcons morphologiques

Sommaire

5.1 Introduction

Dans ce chapitre, nous allons présenter le concept d'une nouvelle méthode de sonification permettant de prendre en compte les besoins des utilisateurs en terme d'adaptabilité des interfaces

sonores, mis en place à partir de concepts existants déjà bien étudiés dans la littérature. Les morphocons (contraction de "morphologique" et de "earcons") sont des petites entités sonores héritant des propriétés d'icônes sonores appelées "earcons" et dont la construction se fait sur la base de description de l'évolution temporelle du son. Ce nouveau type d'earcon a été mis en place dans le cadre du projet NAVIG afin de répondre aux besoins des utilisateurs en terme d'esthétique et de personnalisation des messages sonores. Bien que développé pour le projet, ce type d'icônes sonores a été pensé et développé de façon à être utilisable dans de nombreuses autres applications nécessitant la sonification de plusieurs types d'informations.

Nous allons, dans ce chapitre, décrire le processus d'aller/retour entre concept et application, sur lequel a été fondée la mise en place des "morphocons". Dans la section 5.2 nous commencerons par rappeler brièvement le contexte applicatif de ce chapitre : la navigation en champ lointain. Nous ferons un point sur les informations que le système doit fournir, puis nous présenterons un bilan des sessions de réflexion qui ont été menées avec les utilisateurs (décrites en détail dans l'annexe A). La section 5.3 rappellera quelques études mentionnées dans le chapitre 2 qui ont inspiré la mise en place des morphocons. Nous y verrons : les approches de présentation existant dans les systèmes d'aide à la navigation utilisant des messages sonores ; les techniques de sonification permettant de présenter plusieurs types d'informations ; puis nous étudierons les différentes façons d'aborder la notion d'ergonomie dans les interfaces sonores. Le concept des morphocons ainsi que leur application au projet NAVIG sera décrit dans la section 5.4. Enfin, l'expérience perceptive menée sur plusieurs palettes de morphocons créées pour le système NAVIG sera décrite dans la section 5.5. Les résultats de cette évaluation mettent en évidence l'efficacité des morphocons à transmettre une même information avec plusieurs types de sons sans nécessiter d'apprentissage supplémentaire.

5.2 La navigation en champ lointain

Nous avons vu dans le chapitre 1 que le projet NAVIG a pour objectif de développer un système d'aide pour deux actions particulières : la navigation en champ lointain et la préhension d'objet en champ proche. Pour cette étude nous nous focalisons sur la partie d'aide à la navigation en champ lointain. Pour ce type d'action, nous souhaitons mettre en place une approche de guidage intuitif en relation avec les processus de la perception spatiale, de la cognition spatiale, et de la navigation urbaine. Avec cette approche, le système doit pouvoir donner les informations selon trois grandes fonctions : avertir, décrire, guider. Ici, nous étudierons la manière dont le système doit donner les informations pour avertir l'utilisateur sur les étapes principales de son trajet et sur le mobilier urbain qui l'entoure.

Dans cette section, nous commencerons par rappeler les différentes informations que le système doit fournir aux non-voyants puis nous ferons un résumé des constatations amenés par les différentes sessions de réflexion avec les non-voyants qui ont été abordées dans l'annexe A.

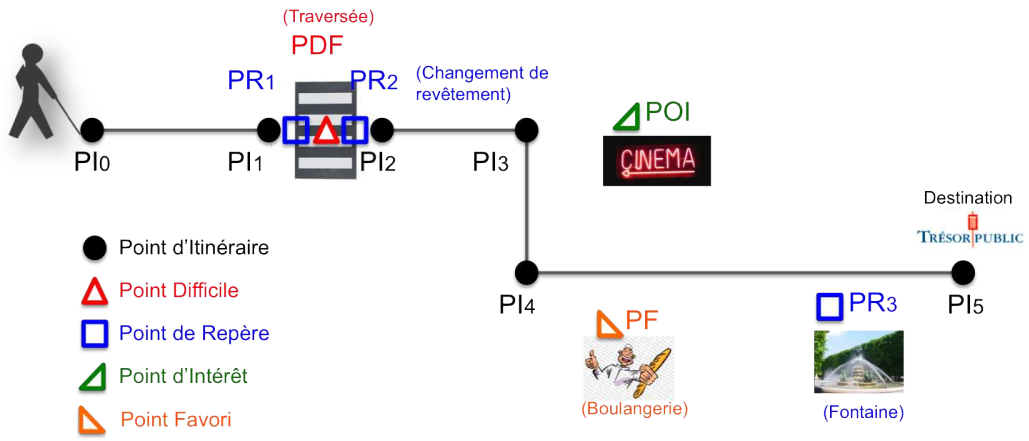


FIGURE 5.1 – Exemple de trajet et du mobilier urbain qui doit être signalé aux utilisateurs.

5.2.1 Les informations à fournir

La revue de la littérature sur la cognition spatiale chez les non-voyants [Gallay *et al.*, 2012], les séances de conceptions participatives ainsi que les expérimentations avec un panel de non-voyants et de rééducateurs en mobilité ont permis de dégager les informations nécessaires pour permettre le guidage d'un utilisateur le long d'un trajet inconnu.

Deux types d'informations sont considérés comme nécessaires pour guider un non-voyant de façon sûre et fiable. Le premier correspond aux informations sur la trajectoire à suivre, il doit donner des indications sur les directions à prendre, les distances à parcourir et les noms des rues. Ce type d'informations est similaire aux informations présentées pour la navigation "pas-à-pas" développée dans la majorité des systèmes GPS du commerce. S'il est donné suffisamment à l'avance, ce type d'informations permet à l'utilisateur de se faire une représentation mentale de son trajet, d'anticiper son trajet et de se déplacer avec plus de flexibilité. Le deuxième type correspond aux informations additionnelles permettant à l'utilisateur de mieux connaître le mobilier urbain qui l'entoure afin de placer son trajet dans le contexte plus global d'un quartier et de devenir plus confiant lors du déplacement qu'il effectue. Ces informations peuvent être, par exemple, la taille des trottoirs, la présence ou l'absence de passages piétons, les points d'intérêts le long du parcours ainsi que les points de repères détectables par les non-voyants. La figure 5.1 présente un exemple de trajet et les différentes informations qui peuvent y être rencontrées.

Le recensement de ces informations a permis de définir un ensemble d'objets qui devraient être inclus dans le système d'information géographique utilisé par le dispositif de guidage. Cinq grandes catégories d'objets ont été identifiées (certaines catégories pouvant contenir plusieurs sous-catégories, le tout définissant un ensemble de 15 objets) :

– **Point d'Itinéraire (PI)** : Cette catégorie permet de définir l'itinéraire. C'est l'ensemble des points importants constituant la trajectoire. Ils doivent être définis de façon à être situés sur les trottoirs ou à des endroits permettant aux piétons de se déplacer de manière sûre. Ils

marquent les changements de directions et les changements de sections au sein d'un trajet. L'écoute d'un son parcourant spatialement tous ces points doit permettre à l'utilisateur de se faire une idée de la forme globale de son trajet. Ces points ont une fonction de guidage.

– **Point Difficile (PDF)** : Ils correspondent aux traversées, aux intersections ou carrefours ainsi qu'à tous les passages du trajet qui peuvent être considérés comme problématiques pour les non-voyants et dont l'utilisateur doit être informé de façon spécifique. Ils ont le même rôle que les points d'itinéraire mais sont différenciés afin de permettre à l'utilisateur de connaître les difficultés qu'il peut rencontrer lors du trajet. Ils ont donc pour fonction d'alerter l'utilisateur. Comme nous l'avons vu précédemment, une fois alerté, l'utilisateur arrivant à proximité d'un PDF, pourra demander une description de ce point, puis enclencher le guidage. Dans cette section nous travaillerons uniquement sur la manière de signaler la présence de points difficiles le long du trajet.

– **Point de Repère (PR)** : Ce sont les différents objets utiles à la compréhension de l'espace que le non-voyant doit pouvoir détecter sans le système (avec sa canne, ses oreilles, son nez, etc.). Ces points permettent à l'utilisateur de confirmer sa position le long de l'itinéraire. Étant donné que les GPS ne sont pas toujours fiables, la détection d'un de ces points par l'utilisateur lui permet (si le point a préalablement été annoncé) de confirmer sa position et donc d'être sûr qu'il se trouve au bon endroit. Si les PR sont suffisamment nombreux, un non-voyant ayant préparé son itinéraire aura la possibilité de ne pas utiliser le système pendant le trajet tout en ayant régulièrement des confirmations sur sa position. Les PR sont divisés en quatre sous-catégories :
  - *Tactile* : les éléments que l'utilisateur peut repérer avec la canne ou avec les pieds. (ex : un changement de revêtement du sol, des plots en bétons, ...) ;
  - *Vestibulaire* : les éléments repérables grâce à la perception du mouvement et à l'orientation par rapport à la verticale (ex : escaliers, rue qui descend, ...) ;
  - *Sonore* : les éléments que l'utilisateur peut entendre de façon naturelle (ex : une fontaine, un jardin, un grand espace, ...) ;
  - *Olfactif* : les éléments que l'utilisateur peut sentir (ex : une boulangerie, un espace vert, ...).

  Notons que si certains de ces points peuvent être facilement accessibles dans les SIG pour piétons existants, d'autre (tel que les PR sonores ou olfactifs) nécessitent un travail de réflexion sur l'environnement urbain qui n'a pas encore été réalisé à ce jour.

– **Point d'Intérêt (POI)** : Comme dans tous les systèmes GPS ces points représentent les lieux ayant un intérêt potentiel pour l'utilisateur. En plus de les avertir sur des lieux utiles, ils leur permettent de se faire une meilleure compréhension de l'environnement dans lequel ils se déplacent. Ces points sont extrait du SIG où ils sont stockés accompagnés de métadonnées comprenant un nom de sous-catégorie (ex : commerce alimentaire) et une description détaillée

(boulangerie + nom de la boulangerie). Pour le système NAVIG, nous avons restreint à sept le nombre de sous-catégories. Celles-ci pourront être définies par l'utilisateur, voici un exemple de sept sous-catégories possibles :

- *Restauration*
- *Hôtellerie*
- *Commerces divers*
- *Commerces alimentaires*
- *Services*
- *Sport/Loisir*
- *Tourisme*

– **Point Favori (PF)** : Il s'agit des POI personnels de l'utilisateur. L'utilisateur peut les ajouter au SIG du dispositif au cours de ses trajets par l'intermédiaire d'un bouton ou d'une commande vocale spécifique. L'utilisateur a la possibilité de définir deux catégories de PF, comme par exemple :
  - *professionnel*
  - *personnel*

Éléments fondamentaux du système d'aide à la navigation, les PI sont répétés tout le long du trajet afin de maintenir l'orientation de l'utilisateur vers sa destination (ils permettent un guidage "point-à-point"). L'utilisateur a la possibilité de modifier la fréquence de répétition (en demandant une présentation de la balise sonore toutes les n secondes ou tous les x mètres). Les autres points (PDF, PR, POI et PF) ont pour fonction d'informer ou d'alerter l'utilisateur. Ils ne sont joués qu'une seule fois, lorsque l'utilisateur arrive à une distance prédéfinie (paramétrable pour chaque sous-catégorie) de l'objet. Pour ces catégories, différents niveaux d'informations sont possibles (fournir uniquement la catégorie, ajouter un son correspondant à la sous-catégorie ou ajouter la description totale avec de la synthèse vocale).

5.2.2 Les besoins des utilisateurs

Afin de concevoir une sonification accessible, plaisante et ergonomique, les besoins des utilisateurs en terme d'interface de sortie ont été évalués avec des questionnaires et avec une session de design participatif (voir [Brunet, 2010] pour plus de détail). Voici un résumé des constatations résultant des échanges avec les utilisateurs potentiels du système :

– En général les non-voyants ne sont pas favorables à l'utilisation du son comme moyen de guidage. En plus du problème de masquage des sons réels par le casque, ils se plaignent d'une grande fatigue auditive due aux types de sons généralement utilisés (comme les "bips") dans les interfaces et à la longueur des messages dans le cas des systèmes utilisant la synthèse vocale.

– Comme les informations sonores peuvent interférer avec les indices sonores naturels de l'environnement et causer une charge cognitive supplémentaire, la quantité d'informations présentées doit être minimale et afficher uniquement ce qui est nécessaire et suffisant pour aider l'utilisateur. Les messages doivent être efficaces et les moins intrusifs possibles. Le niveau de détail et la fréquence d'apparition des messages doivent être ajustables par l'utilisateur. Les sons doivent être brefs et différents des sons de l'environnement direct.

– Enfin, chaque utilisateur a sa propre façon d'imaginer les sons d'un système de guidage idéal. Si certains souhaitent utiliser des sons électroniques (comme par exemple des sons de jeux vidéo) afin de facilement les différencier des sons naturels ambiants, d'autres préfèrent les sons naturels (sons d'animaux, de la mer, de la forêt) ou les sons instrumentaux. Au regard de ces résultats, il est difficile de trouver un consensus sur le type de sons à utiliser pour le système d'aide à la navigation.

Pour satisfaire les utilisateurs, il est donc nécessaire de concevoir une sonification qui soit paramétrable par l'utilisateur et qui puisse répondre à ses exigences esthétiques.

5.3 Contexte bibliographique

Afin de répondre de la façon la plus appropriée aux exigences des futurs utilisateurs, nous allons, dans cette section, faire un bref rappel des travaux décrits dans le chapitre 2 qui ont influencé notre démarche. Nous commencerons par rappeler les méthodes de présentation audio de l'information utilisées dans les systèmes d'aide à la navigation, en détaillant leurs avantages et leurs inconvénients. Nous verrons ensuite les techniques de sonification pouvant être utilisées pour présenter les informations décrites dans la section 5.2.1. Enfin, nous décrirons les travaux traitant de la notion d'adaptabilité dans les interfaces sonores.

5.3.1 Utilisation du son dans les systèmes d'aide à la navigation

Pour les systèmes d'aide à la navigation, le moyen le plus évident de donner les informations à l'utilisateur est d'utiliser des instructions vocales basées sur un langage spatial et synthétisées avec un logiciel de synthèse vocale traditionnelle. Les produits existant dans le commerce tel que le Trekker [1] (UmanWare) ou le Kapten [2] (Kapsys) utilisent ce moyen de transmission pour les instructions en addition avec des indices sonores permettant de rendre compte de l'interaction de l'utilisateur avec l'interface (déplacement dans les menus, changement de mode, etc ...). Malgré la facilité de compréhension des messages vocaux, il existe plusieurs désavantages dans l'utilisation d'instructions vocales pour une application en navigation. Tout d'abord, l'homme a une capacité

1. url : http ://www.humanware.com/
2. url : http ://www.kapsys.com/

limitée pour traiter plusieurs flux de parole. L'interface ne peut donc pas afficher plusieurs informations en même temps. Deuxièmement, utiliser des messages vocaux pour donner des instructions spatiales (comme "tourner à gauche dans 100 mètres") n'est pas efficace et prend du temps. Cela conduit à des systèmes utilisant de longues phrases, avec une synthèse vocale dont la qualité n'est pas toujours au rendez-vous, ce qui a pour effet d'augmenter l'irritation de l'utilisateur. De plus, il est difficile de se focaliser sur les instructions verbales pendant un déplacement tout en effectuant simultanément une autre tâche, telle que le contournement d'un obstacle, la détection de dangers potentiels ou simplement la discussion avec une autre personne.

Un second moyen de transmettre des informations spatiales sous forme sonore, étudié dans plusieurs projets de recherche [Loomis *et al.*, 1998a, Helal *et al.*, 2001, Holland *et al.*, 2002, Wilson *et al.*, 2007], consiste à utiliser les capacités de perception sonore spatiale de l'homme et de transmettre les informations en utilisant la synthèse binaurale. Plutôt que de chercher à décrire la trajectoire ou la position d'un objet avec des mots, l'utilisation de la synthèse binaurale permet de placer des balises sonores virtuelles permettant de guider les mouvements de l'utilisateur vers des objets ou des cibles spécifiques. Précurseur dans ce domaine, l'équipe de Loomis a démontré la supériorité du son 3D sur la synthèse vocale, au niveau de la charge cognitive et de la précision d'orientation pour les tâches de déplacement. Dans [Klatzky *et al.*, 2006], les auteurs ont comparé trois types de rendu sonore (balises sonores virtuelles, instructions verbales et synthèse vocale spatialisée) et mis en évidence une amélioration significative des performances avec l'utilisation de balises sonores. Dans une étude sur l'effet du type de balise sonore sur les performances de navigation, [Walker et Lindsay, 2006] ont testé trois types de sons spatialisés (du bruit rose, un "bip" de sonar et un ton pur de 1000 Hz) et montré leurs effets respectifs sur la vitesse et la précision de déplacement dans les tâches audio guidées.

Bien que ces projets de recherche aient démontré l'habileté de l'homme à se déplacer en utilisant des indices de localisation sonore et qu'ils aient mis en évidence les bénéfices de ce type de rendu sur les performances de déplacement sans la vision, les non-voyants sont réticents à l'idée d'utiliser ce type de système. Le premier reproche qui peut être fait est le masquage des sons réels. En effet, la création de balises sonores 3D avec la synthèse binaurale nécessite l'utilisation d'un casque audio et celui-ci masque les sons de l'environnement réel, empêchant ainsi le non-voyant de "voir". Ce problème a été résolu par l'utilisation de casques osseux. Ces casques, qui fonctionnent par transmission osseuse, se placent contre les os de la tête, devant où derrière les oreilles. Étant donné qu'ils ne couvrent pas les oreilles, leur utilisation permet un masquage minimal des sons naturels, les oreilles de l'utilisateur étant parfaitement libres lorsque le système n'émet pas de sons. [Walker et Lindsay, 2005] ont démontré que leur utilisation n'entraine pratiquement pas de dégradation de la localisation des sons virtuels et qu'elle a peu d'effets sur les performances de navigation. Le second reproche qui peut être fait à ce type de système est le manque de convivialité et d'esthétique des balises sonores. La plupart des études n'ont porté que sur des sons simples (tels que des "bruits blancs" ou des "tons purs") qui peuvent être très irritants pour une utilisation journalière et ne peuvent être appliqués qu'à un seul type d'information.

Étant donné que les systèmes d'aide à la navigation doivent inclure plusieurs types de messages, tels que des balises sonores, des points d'intérêts, des points de repères, il est important d'étudier la façon dont ces différents types d'informations vont être transmis avec un rendu sonore ergonomique et intuitif. Il est donc important d'adapter les techniques de sonification aux besoins des utilisateurs afin de construire un système d'aide à la navigation utilisable par les non-voyants.

5.3.2 Sonification

L'état de l'art de la section 2.3, a présenté les différentes méthodes de sonification traditionnellement utilisées en VAD. Dans le cas d'un système d'aide à la navigation, les fonctions de la sonification sont de fournir des messages d'états, des avertissements et de permettre l'exploration de données cartographiques. La méthode de sonification la plus appropriée à notre application est l'utilisation d'icônes sonores. En effet, cette méthode préconise l'utilisation de sons très court et permet de manière assez simple de constituer un ensemble de sons permettant de transmettre plusieurs types d'informations (balises sonores à atteindre, information sur la présence d'un objet, alerte sur un point difficile, etc.). Trois catégories d'icônes sonores ont été définies dans la section 2.3.2.d) : les auditory icons, les earcons et les spearcons.

Les avantages et les inconvénients de ces différentes méthodes de sonification ont été relativement bien étudiés en Auditory Display. Plusieurs études ont exploré la facilité d'apprentissage de tels types de rendu et la façon dont ils affectent l'utilisation de toute l'interface [Lucas, 1994, Absar et Guastavino, 2008, Dingler *et al.*, 2008, Garzonis *et al.*, 2009]. Dans [Absar et Guastavino, 2008], par exemple, les auteurs ont comparé les auditory icons et les earcons dans différentes applications et montré que les auditory icons sont plus appropriés pour surveiller des tâches et permettre la navigation dans un ordinateur alors que les earcons sont préférés pour le déplacement dans les menus de téléphones, les alarmes et les systèmes d'alertes. [Dingler *et al.*, 2008] ont quant à eux montré la supériorité des spearcons comparés aux auditory icons et aux earcons en terme d'apprentissage. D'autres études ont exploré les performances de navigation en comparant plusieurs types de sons pour les balises sonores et en étudiant l'effet de la fréquence de l'affichage (voir [Walker et Lindsay, 2006, Tran *et al.*, 2000] pour une évaluation ergonomique des caractéristiques des balises sonores et des différences entre la parole et le son).

Malgré les résultats de toutes ces études, il semble que la faculté d'apprentissage et les performances de navigation ne sont pas des critères suffisants pour rendre compte de l'utilisabilité du rendu sonore d'un système d'aide à la navigation.

5.3.3 La notion de satisfaction des utilisateurs dans les interfaces sonores

Dans le domaine de l'ergonomie, la notion d'utilisabilité d'un système ou d'un produit, quel qu'il soit, fait référence à la mesure de la façon dont des utilisateurs spécifiques peuvent atteindre un but défini avec efficacité, efficience et satisfaction, dans un contexte d'utilisation spécifiée [Kraig, 2010].

- La notion d'efficacité fait référence à la capacité du dispositif de permettre d'atteindre un objectif donné. Elle porte sur la mesure de la qualité du résultat obtenu, elle correspond à la notion de performance évoquée dans la section précédente.
- L'efficience correspond à la capacité à produire une tâche donnée avec le minimum d'efforts. Cette notion renvoie en ergonomie à la charge de travail plus ou moins imposée à l'utilisateur. Pour les études sur les interfaces sonores cette notion est intimement liée à l'apprenabilité (ou facilité d'apprentissage) et la mémorisation des messages du système (un système facile à utiliser est un système facile à apprendre).
- La satisfaction se réfère au niveau de confort ressenti par l'utilisateur lorsqu'il utilise le système, à son plaisir et son bien-être. C'est une évaluation subjective provenant d'une comparaison entre ce que l'usage apporte à l'individu et ce qu'il s'attend à recevoir. Elle découle directement du degré d'efficacité et d'efficience atteint, mais ces deux critères ne suffisent pas à la définir.

Alors que les notions d'efficacité et d'efficience des messages sonores ont été régulièrement étudiées, la notion de satisfaction de l'utilisateur dans les interfaces sonores souffre toujours d'un manque de connaissances. En effet, les études sur la sonification se concentrent essentiellement sur l'amélioration de la performance des tâches ; bien souvent les sons utilisés ne sont pas agréables et la notion de satisfaction de l'utilisateur est très rarement abordée. La principale raison de ce manque d'investigation est le manque de critères objectifs pour mesurer la satisfaction des utilisateurs. Si la mesure de l'efficacité et de l'efficience d'un système est une notion quantifiable bien contrôlable en ergonomie, la mesure de la satisfaction des utilisateurs passe par des questionnaires ou par d'autres mesures subjectives qui rendent compliquée l'extraction des résultats.

En plus de dépendre de l'efficacité et de l'efficience, la notion de satisfaction de l'utilisateur face à un rendu sonore est liée à plusieurs critères tels que l'esthétique de l'interface et les préférences sonores de l'utilisateur.

Dans [Gaver, 1997], l'auteur affirme que l'un des grands axes de travail à explorer pour que le domaine de l'Auditory Display devienne mature, est la conception de sons esthétiques, aussi subtils et beaux que ceux que nous pouvons entendre dans un orchestre ou pendant une promenade dans les bois. Dans le cadre de sa thèse, [Leplâtre, 2002] a recensé trois critères pour la conception de sons esthétiquement satisfaisants :

- L'homogénéité de la conception - Souvent les différences entre les sons sont maximisées afin que ces derniers soient facilement reconnaissables. Ceci conduit à des interfaces auditives peu homogènes ce qui réduit leur esthétique.
- L'enveloppe temporelle du son - Les sons utilisés doivent être brefs et ils doivent pouvoir être facilement stoppés. L'information doit se situer au début du son. Un lissage de l'attaque et de la fin du son doit être effectué afin d'éviter les "clics" dus aux transitions trop abruptes.
- La densité du son - La densité perçue d'un son dépend de plusieurs facteurs tel que sa durée, son intensité, la largeur de son spectre, ... Une interface esthétique ne doit pas contenir de sons trop denses. Une étude de [Steele et Chon, 2007] sur l'ennui causé par les sons, montre que plus le son a une grande largeur fréquentielle plus il est considéré comme ennuyant par l'utilisateur.

Plus tard, dans une tentative de s'attaquer au problème de l'esthétique dans les interfaces audio, [Leplâtre et Mcgregor, 2004] mettent en évidence la forte dépendance entre la nature de la tâche et l'esthétique du rendu sonore. Leur principale conclusion est que les propriétés fonctionnelles et esthétiques de l'interface ne peuvent pas être traitées indépendamment.

D'autres études telles que [Cohen, 1993, Sikora *et al.*, 1995, James, 1996, Peres *et al.*, 2007] ont traité la notion d'esthétisme en étudiant la satisfaction des utilisateurs pour une application donnée face à un rendu composé de bruits, de "bip", de sons naturels ou de sons musicaux. Il résulte de ces études que les sons naturels et les sons musicaux sont bien souvent plus appréciés même lorsqu'ils entraînent des dégradations de performances. Dans le cadre d'études sur la sonification interactive d'activités sportives, [Schaffert *et al.*, 2009] et [Barrass *et al.*, 2010] ont comparé plusieurs méthodes de sonification d'un même type d'activité et mis en évidence les différences de préférences entre les utilisateurs. L'étude de [James, 1996] va aussi dans ce sens et émet l'hypothèse que l'utilisateur préfère les sons qui, pour lui, sont facilement reconnaissables en fonction de son expérience et que cette familiarité permet une meilleure acceptation de l'interface.

Ces dernières années, le développement considérable des interfaces utilisateurs pour les ordinateurs ou les téléphones portables a conduit à la mise en place d'interfaces personnalisables (comme les icônes visuels des bureaux Windows, ou les palettes de couleurs dans le design graphique). Si de nombreux brevets existent sur la personnalisation des messages sonores dans les interfaces, il n'existe, à notre connaissance, pas d'étude sur l'effet de cette paramétrisation des sons par l'utilisateur sur l'utilisabilité de l'interface. La personnalisation se fait bien souvent par le biais de palettes et l'utilisateur a la possibilité de changer de thème de palette ce qui affecte l'ensemble des icônes. Si cette méthode a déjà fait ses preuves dans les applications où les informations sonores sont peu nombreuses, elle reste difficile à appliquer dans des cas où les messages sont nombreux car elle nécessite de réapprendre la relation entre les icônes sonores et l'objet auquel ils font référence à chaque changement de palette.

Le but de cette étude est de proposer une nouvelle méthode permettant d'améliorer la satisfaction de l'utilisateur, en développant cette notion de paramétrisation et de personnalisation des interfaces sonores par les utilisateurs. Cela permettra ainsi à l'utilisateur de choisir une palette qui convienne à ses critères esthétiques sans toutefois nécessiter de réapprentissage.

5.4 Les morphocons

Dans cette étude, le concept d'earcons (motifs de notes) a été étendu au concept de morphocons (motifs de paramètres acoustiques). Les morphocons (contraction de morphologie et de earcons) permettent la construction d'un langage sonore hiérarchique basé sur la variation temporelle de plusieurs paramètres acoustiques. Ces variations morphologiques peuvent être appliquées à différents types de sons (naturels, musicaux ou de synthèse) et permettent ainsi de construire une infinité de palettes sonores tout en conservant une cohérence entre les sons et les objets ou messages à transmettre.

Dans cette section, nous allons introduire le concept des morphocons puis nous détaillerons le langage sonore mis en place pour le projet NAVIG et appliqué à différents types de sons pour construire trois palettes sonores.

5.4.1 Concept

Considérant que les sons de synthèse et les sons "midi" largement utilisés en sonification ont une structure interne simple qui entraîne trop rapidement un ennui de l'utilisateur, nous posons l'hypothèse que l'utilisation de sons "écologiques" paramétrisables par l'utilisateur va permettre d'augmenter le sentiment d'esthétisme de l'interface et entraîner une meilleure utilisabilité. Nous avons donc cherché à mettre en place un moyen de sonification permettant de créer des palettes d'icônes sonores qui puissent être modifiées par l'utilisateur sans dégrader l'efficience du système.

Le concept des morphocons est inspiré des travaux effectués ces cinquante dernières années pour l'analyse de la musique électroacoustique. Dans ses travaux en partie formalisés dans le "Traité des objets musicaux", [Schaeffer, 1977] a développé le concept d'objet sonore. Ce concept, par la suite repris par [Chion, 1983] est défini comme "tout phénomène ou événement sonore perçu comme un ensemble, comme un tout cohérent et entendu dans une écoute réduite qui le vise pour lui-même, indépendamment de sa provenance et de sa signification". Il s'oppose à l'écoute causale (qui est liée à la reconnaissance de la source physique du son) et à l'écoute sémantique (qui consiste à analyser le sens transmis par le son), il est basé sur une écoute liée aux caractéristiques inhérentes au son, indépendamment de sa cause et de sa signification. Pierre Schaeffer définit ensuite un vocabulaire descriptif et une grille de classification des objets sonores contenant sept critères morphologiques :

- la masse,
- le timbre harmonique,
- le grain,
- la dynamique,
- l'allure,
- le profil mélodique,
- le profil de masse.

Un son pourrait être comparé à une matière possédant une certaine forme définie par des variations morphologiques.

L'apparition de cette notion de "morphologie sonore", coïncide avec l'apparition des nouvelles technologies de traitement du son (enregistrement sur bande puis développement des outils informatiques). Dans [Tissot, 2010], l'auteur compare les écrits des compositeurs Pierre Schaeffer, Iannis Xenakis et Denis Smalley et donne un aperçu de l'évolution de la réflexion sur la morphologie sonore dans les musiques électroacoustiques. Pour [Xenakis, 1979], les morphologies se rapportent à un petit nombre de modèles qui peuvent être considérés comme universels, relevant de domaines aussi différents que la musique, les arts plastiques ou les mathématiques. Dans beaucoup de ses travaux, il fait le parallèle entre morphologie sonore et formes graphiques. Il construit même dans les années 1970, l'U.P.I.C. (acronyme de Unité Polyagogique Informatique du CEMAMu - Centre

d'Études de Mathématique et Automatique Musicales), un dispositif permettant de dessiner la micro et la macro-structure de la musique en traçant des formes d'ondes, des enveloppes d'intensité et des courbes associées à des oscillateurs. Plus tard, en s'appuyant sur les recherches de Pierre Schaeffer, [Smalley, 1986] développe le concept de spectromophologie. Il entreprend de mettre en place une typologie des morphologies basée sur trois figures simples, repérables de manière visuelle sur un sonagramme : attaque-impulsion, attaque-chute et continuité graduelle. A partir des variations de ces morphologies de base, l'auteur décrit plusieurs spectromorphologies (comme par exemple le mouvement "unidirectionnel/ascension" qui consiste en un glissando ascendant du son) particulièrement riche pour l'analyse.

Cherchant à étendre les travaux de Pierre Schaeffer à l'analyse de l'évolution temporelle des objets musicaux, des chercheurs et compositeurs du laboratoire de Musique et d'Informatique de Marseille (MIM) ont développé le concept d'Unités Sémiotiques Temporelles (UST) [Delalande *et al.*, 1996, Frémiot, 1999]. Les UST sont définies comme des segments sonores pouvant transmettre un sens à travers leur organisation dynamique et temporelle. A partir de ce principe, 19 UST ont été définies, en utilisant une méthodologie basée sur l'écoute collective de segments musicaux, pour établir des descriptions sémantiques et morphologiques de ces segments. Les travaux de [Frey *et al.*, 2009a, Frey *et al.*, 2009b] ont permis de valider la pertinence cognitive des UST, au niveau de la perception de formes temporelles, en utilisant des tâches de catégorisations libres et des expériences électrophysiologiques. Ces résultats sont confirmés par les résultats expérimentaux de [Minard *et al.*, 2010], qui, à travers une tache de catégorisation libre de sons environnementaux, met en évidence six classes de sons possédant des morphologies différentes et reconnaissables. Ces résultats confortent l'idée que les auditeurs peuvent dans certaines situations d'écoute, percevoir des formes sonores de la même façon que l'on peut percevoir des formes géométriques abstraites dans une image.

Afin de mettre en place des palettes d'icônes sonores indépendantes du type de son utilisé, nous nous sommes inspirés de ce concept de formes sonores basées sur des descriptions morphologiques. Comme il n'est pas évident d'établir un lien intuitif entre le son et l'information à présenter, l'utilisation des *auditory icons* est compliquée. De plus, les informations à présenter se divisant en plusieurs catégories et sous-catégories, il paraît plus intéressant de s'inspirer des *earcons* qui permettent la mise en place de langage sonore hiérarchique. Les *morphocons* ont donc été définis comme une extension des *earcons* à tous les types de sons. Nous avons vu dans la section 5.3.2 que les *earcons* sont construits à partir de motifs de notes dont l'arrangement de la succession de hauteur, de tempo, de dynamique et de timbre permet de créer des entités individuelles et reconnaissables. Bien que définis comme des sons abstraits, les *earcons* sont, de par leur définition, constitués de sons musicaux. Afin d'étendre leur définition à l'ensemble des sons possibles, nous nous basons sur des motifs de paramètres acoustiques (tel que la fréquence, le tempo, la dynamique) qui peuvent être applicables à tout type de son. Pour cela, nous tenons compte du fait que les utilisateurs sont capables de percevoir une forme sonore (ou un profil sonore morphologique) [Frey *et al.*, 2009a, Minard *et al.*, 2010] sans tenir compte du contexte du son ou du sens musical. Uniquement définis par un ensemble de profils morphologiques, les morphocons peuvent être appliqués à des sons

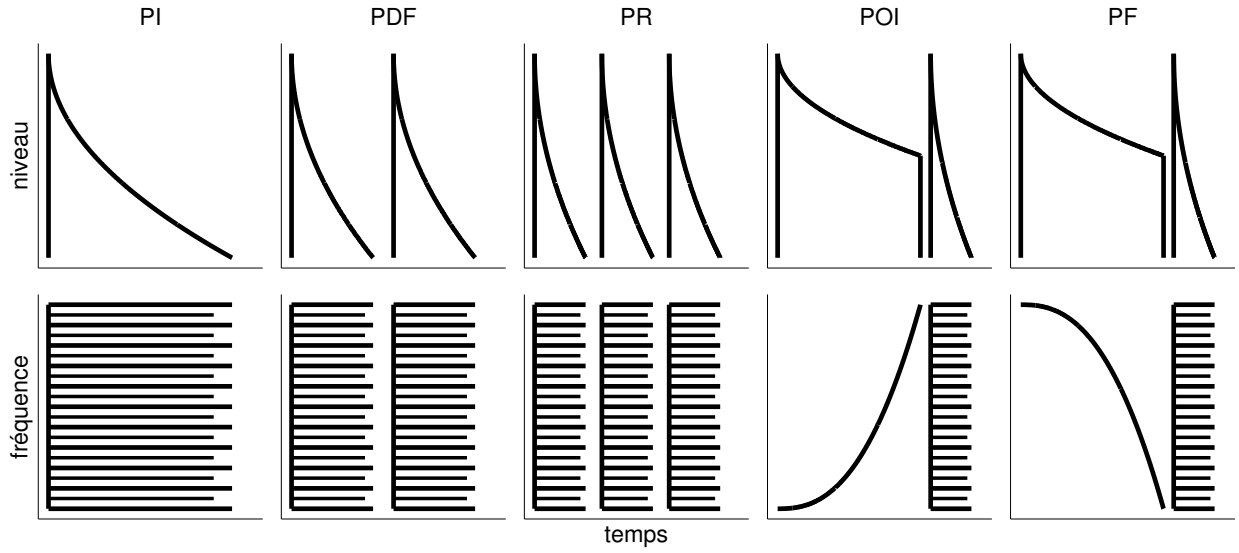


FIGURE 5.2 – Les fonctions temporelles élémentaires représentant le vocabulaire de morphocons mise en place pour représenter les cinq catégories d'informations à présenter avec le système NAVIG.

musicaux, comme à des sons environnementaux ou de synthèse. Il est donc possible de définir un ensemble de morphocons et de l'appliquer à plusieurs ensembles de sons pour créer différentes palettes sonores. L'utilisateur pourra alors choisir la palette qui lui convient le mieux et changer de palette quand bon lui semble ; ceci sans augmenter le temps d'apprentissage des messages du système.

5.4.2 Application au projet NAVIG

La conception d'un ensemble de morphocons pour la sonification des informations décrites dans la section 5.2.1 doit tenir compte des besoins des utilisateurs décrits dans la section 5.2.2 ainsi que des contraintes imposées par le dispositif de rendu sonore (son binaural sur casque osseux). Ces contraintes ont façonné les spécifications sur les formes à utiliser pour la composition des morphocons qui vont représenter les cinq catégories et leurs sous-catégories (pour un total de 15 morphocons pour chaque palette sonore) :

- *Contraintes relatives au casque osseux* : Utiliser le moins possible de variations dynamiques (à cause de la faible sensibilité des casques osseux).
- *Contraintes relatives au rendu binaural* : Attaques franches (afin d'améliorer la perception de l'ITD) ; spectre large (pour augmenter la perception des indices HRTF).
- *Contraintes relatives aux utilisateurs* : Sons brefs (pour éviter le masquage des sons réels, la surcharge cognitive et la fatigue auditive) ; sons plaisants (éviter les sons type bruits blancs ou tons purs) ; pas de sons trop distrayants ; sons robustes face à la contamination par le bruit de l'environnement extérieur.

En tenant compte de ces contraintes un vocabulaire de cinq catégories (PI, PDF, PR, POI et PF) et de 13 sous-catégories (4 PR, 7 POI et 2 PF) de morphocons a été construite. Les formes sonores pour les cinq catégories d'informations à transmettre sont représentées figure 5.2 avec les

variations temporelles de leurs variables acoustiques pertinentes (fréquence et enveloppe). Voici la description détaillée de ce langage sonore :

- PI : évènement bref
- PDF : séquence de deux évènements brefs.
- PR : motif rythmique de trois évènements brefs. Les variations rythmiques de ce motif permettent de différencier la sous catégorie de PR.
  - PR1 : 1 noire, 2 croches : ♩♫
  - PR2 : 2 croches, 1 noire : ♫♩
  - PR3 : 1 croche, 1 noire, 1 croche : ♪♩♪
  - PR4 : 1 croche pointée, 1 double croche, 1 noire : ♩.♬♩
- POI : combinaison de deux groupes de sons : le premier dont la fréquence augmente petit à petit, est commun à tous les POI ; le deuxième est constitué d'un ou plusieurs sons courts permettant de différencier les sept sous-catégories.
  - POI1 : son harmonique grave
  - POI2 : son harmonique aigu
  - POI3 : deux notes ascendantes
  - POI4 : deux notes descendantes
  - POI5 : 3 notes (ascendantes puis descendantes), e.g. Do-Sol-Do
  - POI6 : 4 sons courts répétés
  - POI7 : un son complexe (dont la hauteur n'est pas définie)
- PF : combinaison de deux groupes de sons : le premier dont la fréquence décroit petit à petit est commun à tous les PF ; le deuxième permet de différencier les deux sous-catégories.
  - PF1 : son court
  - PF2 : son long

A partir de ce langage, trois palettes de morphocons ont été construites. Les sons ont été fabriqués avec une combinaison de samples sonores (issus de bases de sons disponibles sur internet) et de sons synthétisés. Les samples ont été transformés en utilisant les algorithmes de "time stretch" (dilatation temporelle) et de "pitch shifter" (modification fréquentielle) du logiciel Audiosculpt [3] et assemblés avec le logiciel Audacity [4]. La première palette, appelée *instrumentale*, a été composée de sons d'instruments à cordes (violon, violoncelle, contrebasse et harpe). La seconde palette, appelée *naturelle* a été composée de sons naturels d'animaux (principalement des cris d'oiseaux). La troisième palette, appelée *électronique* a été composée avec des sons de synthèse. En plus de ces trois palettes, une palette *exemple* composée de bruit blanc et de fréquences pures a été conçue, afin de fournir une description sonore du vocabulaire mis en place aux utilisateurs non-voyants. Tous les icônes, conçus pour cette étude, sont disponible en ligne [5]. Le spectrogramme de la figure 5.3 met en évidence les similarités morphologiques en terme d'évolution fréquentielle pour les PI, les POI4, les PF1 et les PR3.

3. url : http ://forumnet.ircam.fr/691.html
4. url : http ://audacity.sourceforge.net/ ?lang=en
5. url : http ://groupeaa.limsi.fr/projets :navig :start

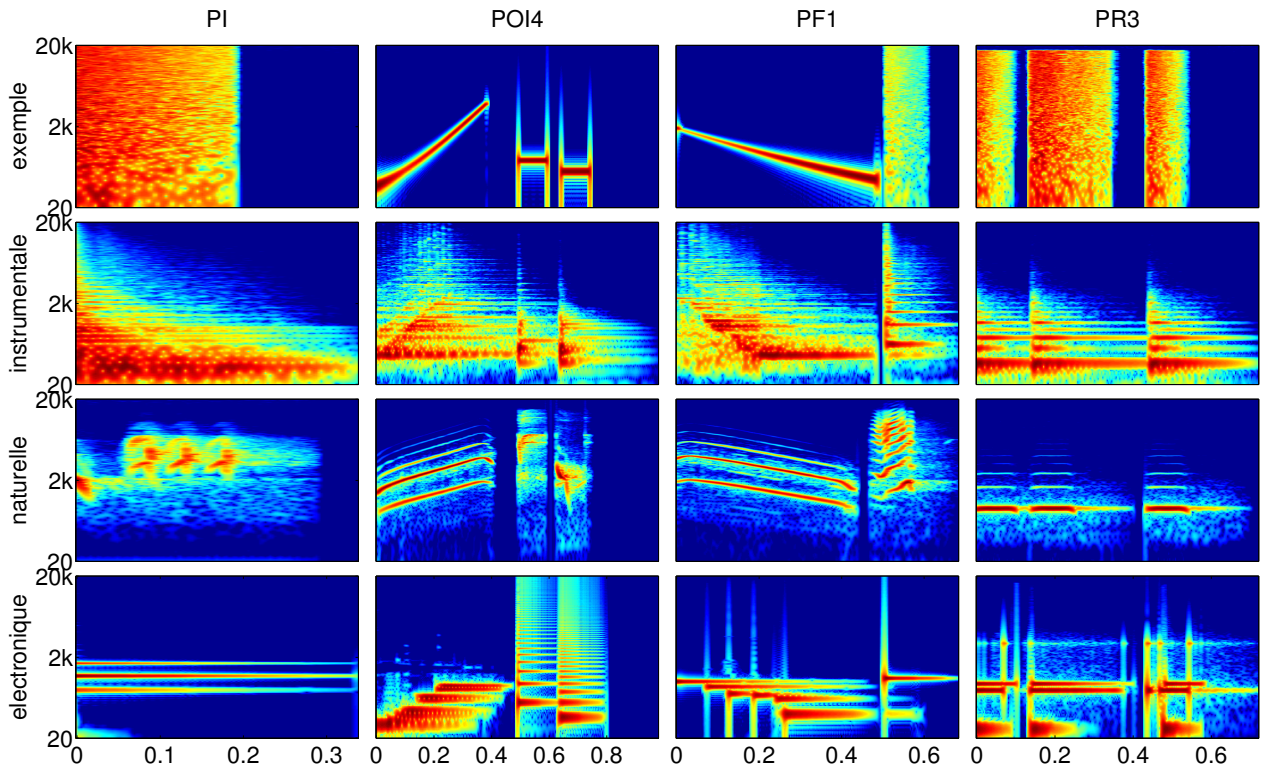


FIGURE 5.3 – Spectrogrammes des sons utilisés pour représenter (de gauche à droite) : les PI, les POI4, les PF1 et les PR3, pour chaque palette (de haut en bas : *exemple*, *instrumentale*, *naturelle*, et *électronique*. Abscisse : temps (en seconde) ; Ordonnée : fréquence (en Hz).

5.5 Expérience

Afin d'évaluer le bon fonctionnement du vocabulaire de morphocons, nous avons mis en place un test perceptif consistant à identifier chaque catégorie et sous catégorie indépendamment du choix de palette. Disponible en ligne [6], ce test a été conçu pour être accessible aux voyants et aux non-voyants. Son but était de déterminer si les sujets étaient capables de différencier et de reconnaitre les différentes catégories et sous-catégories d'informations sur la seule base de descriptions verbales et d'exemples sonores.

5.5.1 Méthode

a) Participants

Un total de 60 sujets a répondu au questionnaire (23 femmes et 37 hommes). Moyenne d'âge : 39 ± 15 ans (minimum 19 ; maximum 80 ans) ; 29 non-voyants, 5 mal-voyants et 26 voyants. Étant donné que les malvoyants étaient des personnes portant des lunettes mais n'ayant pas de problèmes de vision majeure ni de problèmes de mobilité ; ils ont été regroupés avec les voyants

6. url : http ://groupeaa.limsi.fr/projets :navig :start

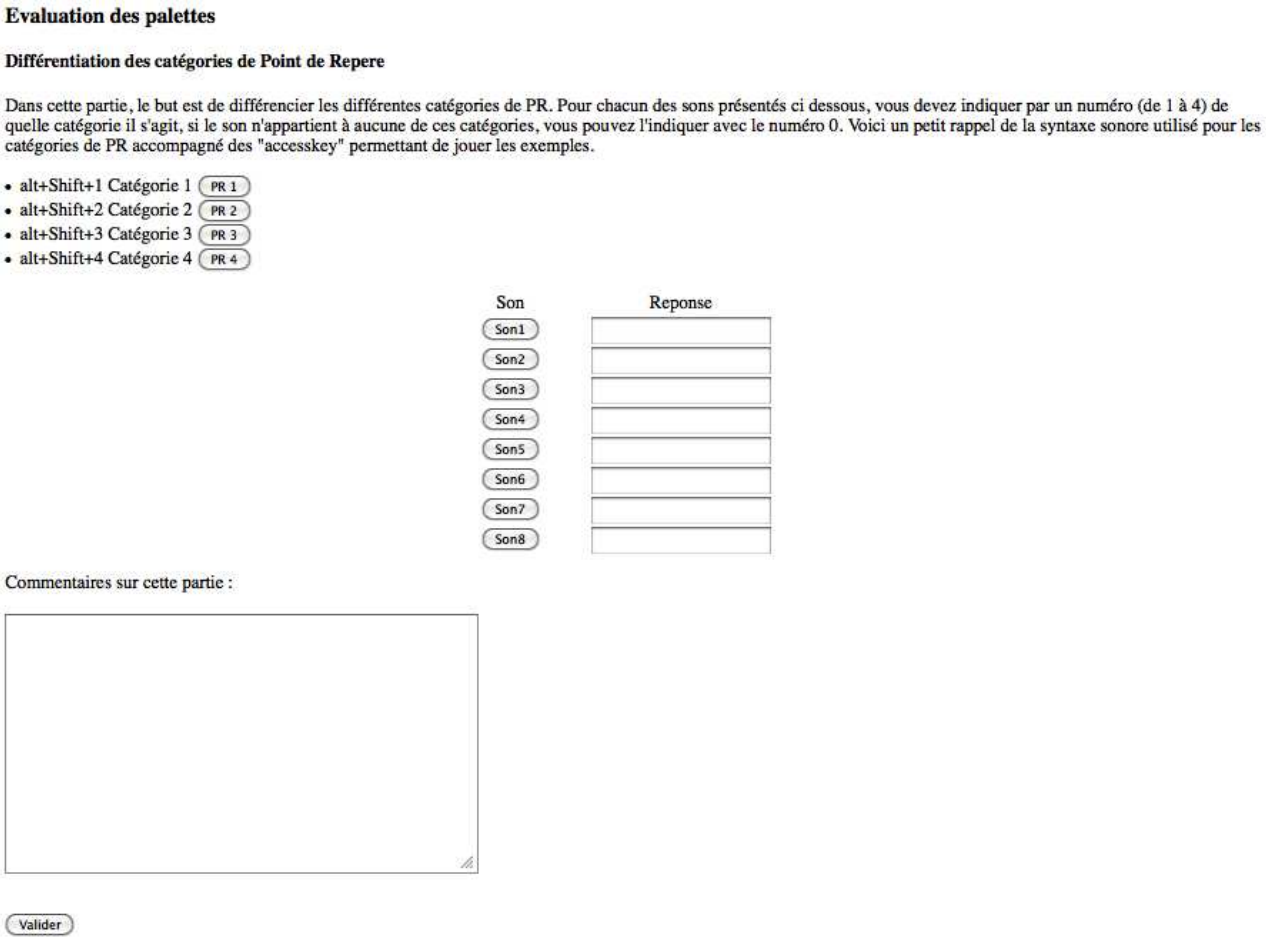


FIGURE 5.4 – Copie d'écran de l'interface mise en place pour la tâche d'identification des catégories de sons.

pour l'analyse des résultats. Il en résulte un groupe de 29 non-voyants et un groupe de 31 voyants.

b) Procédure

L'implémentation d'une tâche de catégorisation libre dans un formulaire internet conçu pour les non-voyants étant assez compliquée, l'expérience a été découpée en quatre tests d'identification. Une description verbale (voir section 5.4.2) et sonore (avec la présentation de la palette *exemple*) du vocabulaire de morphocons était fournie au début du questionnaire. Étant donné qu'il était principalement conçu pour les non-voyants, aucun graphique n'a été utilisé. La compatibilité de ce questionnaire a été vérifiée avec les logiciels de lecteurs d'écrans les plus couramment utilisés par les non-voyants (*Jaws*, *Window-Eyes* et *VoiceOver*). Des raccourcis claviers (*accesskey*) ont été ajoutés pour faciliter l'accés à la lecture d'exemples sonores aux non-voyants. Une copie d'écran de la troisième phase du questionnaire est présentée figure 5.4.

Première phase : identification de la catégorie d'objet décrite par les sons (PI, PDF, PR, POI ou

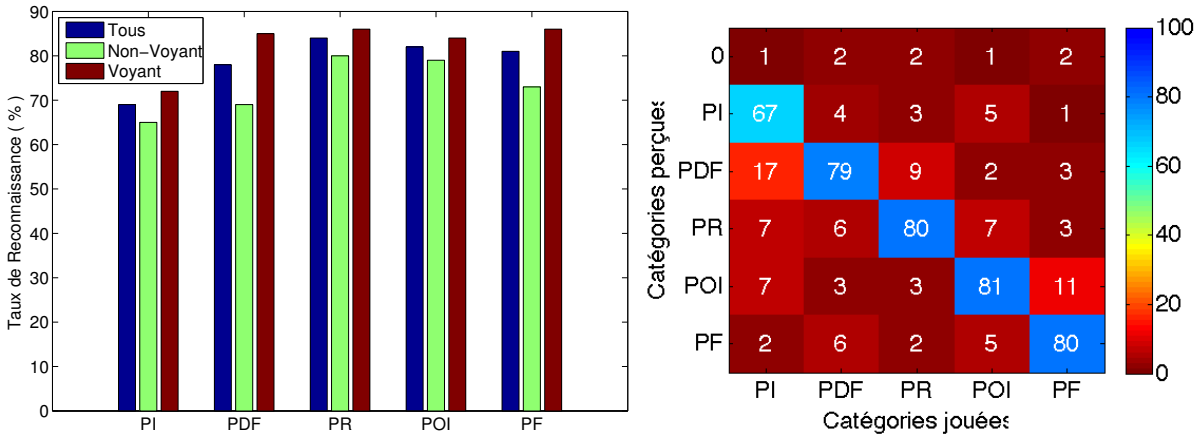


FIGURE 5.5 – *Gauche :* Moyenne du taux de reconnaissance obtenue pour chaque catégorie en fonction du type de cécité. *Droite :* Matrice de similarité entre les catégories jouées et les catégories reconnues pour tous les sujets. 0 ne correspond à aucune catégorie reconnue.

PF). Vingt sons étaient choisis aléatoirement dans une base de 24 sons (8 sons par palette, chaque palette étant réduite à 1 PI, 1 PDF, 2 PR, 2 POI et 2 PF afin d'avoir une répartition quasi égale des sons et des catégories par palettes). Pour chaque son, il était demandé au sujet d'identifier la catégorie d'information correspondante.

Deuxième phase : identification des différentes sous-catégories de points d'intérêt. 14 sons (choisis aléatoirement dans les 21 POI) étaient présentés au sujet qui, pour chacun, devait identifier le type de POI (POI1, POI2, . . . , POI7).

Les troisième et quatrième phases étaient les mêmes que la deuxième mais dédiées à l'identification des sous-catégories de PR et de PF. Le sujet devait identifier 8 sons (choisis de façon aléatoire dans les 12 PR) pour les sous-catégories de points de repères et 4 sons (choisis de façon aléatoire dans les 6 PF) pour les sous-catégories de points favoris.

Le test, ayant lieu sur l'ordinateur personnel des sujets, n'était pas supervisé. Pour chaque phase, le sujet avait la possibilité d'écouter les sons de la palette *exemple* afin de se remémorer la description morphologique de chaque élément.

5.5.2 Résultats

a) Test 1 : Différences entre les catégories

Sept sujets non-voyants n'ayant pas identifié l'intégralité des sons de ce test, les résultats ont été calculés sur 57 sujets (22 non-voyants et 31 voyants).

Pour tous les sujets rassemblés, le taux d'identifications correctes des catégories de sons présentées est de $78 \pm 22\%$ ($81 \pm 17\%$ pour les sujets voyants ; $74 \pm 27\%$ pour les sujets non-voyants). La figure 5.5 (à gauche) présente le taux d'identification correcte pour chaque catégorie d'informa-

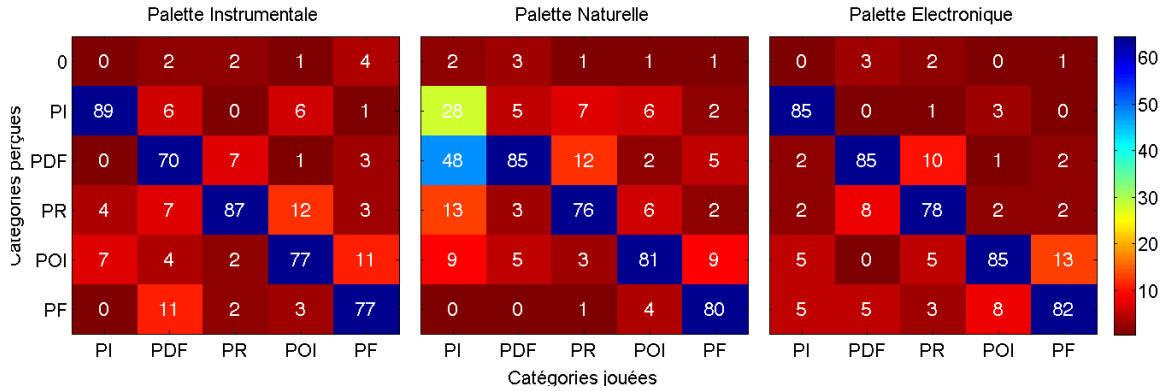


FIGURE 5.6 – Matrices de similarités entre les catégories jouées et les catégories reconnues pour chaque palette et tous les sujets.

tions en indiquant séparément les résultats des voyants et des non-voyants. On remarque que les performances des non-voyants sont légèrement inférieures à celles des voyants.

En ce qui concerne le type de catégorie, les sons utilisés pour les points d'itinéraires ont été moins bien reconnus que les sons des autres catégories. En général, le taux d'identification était très élevé et il semble que tous les sujets étaient capables d'identifier correctement chacune des cinq catégories d'informations. Une ANOVA à mesures répétées a été réalisées sur les résultats moyens par catégorie et par sujet en prenant en compte l'état de cécité. Elle met en évidence un effet *quasi*-significatif de l'état de cécité $[F(1, 52) = 3.77, p = 0.058]$ ainsi qu'une différence significative entre les différentes catégories $[F(4, 208) = 3.41, p < 0.01]$. Un test Posthoc de Duncan montre une différence significative entre la reconnaissance des PI et celle des autres catégories. Étant donné que chaque sujet n'a jugé qu'une partie des sons, une analyse statistique de l'effet du type de palette n'est pas possible.

La figure 5.5 (à droite) montre la répartition des taux d'identification pour chaque catégorie jouée. Cette figure permet d'identifier les différentes sources de confusions entre les types de catégories. On remarque par exemple que 17% des PI sont perçus comme des PDF. La figure 5.6 présente le même type d'information en fonction de la palette utilisée. Elle permet d'analyser l'effet de chaque palette. En effet, l'analyse du taux de reconnaissance des PI montre que les faibles performances pour l'identification des PI (observée dans la figure 5.5) est uniquement due à une mauvaise reconnaissance du PI de la deuxième palette (les PI des deux autre palettes étant respectivement correctement identifiés avec des taux de 89% et 85%). Cette figure permet de déterminer pour chaque palette avec quelles catégories elles ont été confondues. Le PI de la deuxième palette, par exemple, a été principalement reconnu comme un PDF (à 48%) et comme un PR (à 13%). En moyenne, les catégories ont été correctement identifiées à 80% pour la palette *instrumentale*, 70% pour la palette *naturelle* et 83% pour la palette *électronique*. On peut aussi remarquer sur la figure 5.5 (à droite), une légère confusion entre les PDF et les PR (9% des sons de PR étant identifiés comme des PDF et 6% des sons de PDF étant identifiés comme des PR). Des confusions ont aussi été remarquées entre les sons de POI et de PF (il est intéressant de noter que cette

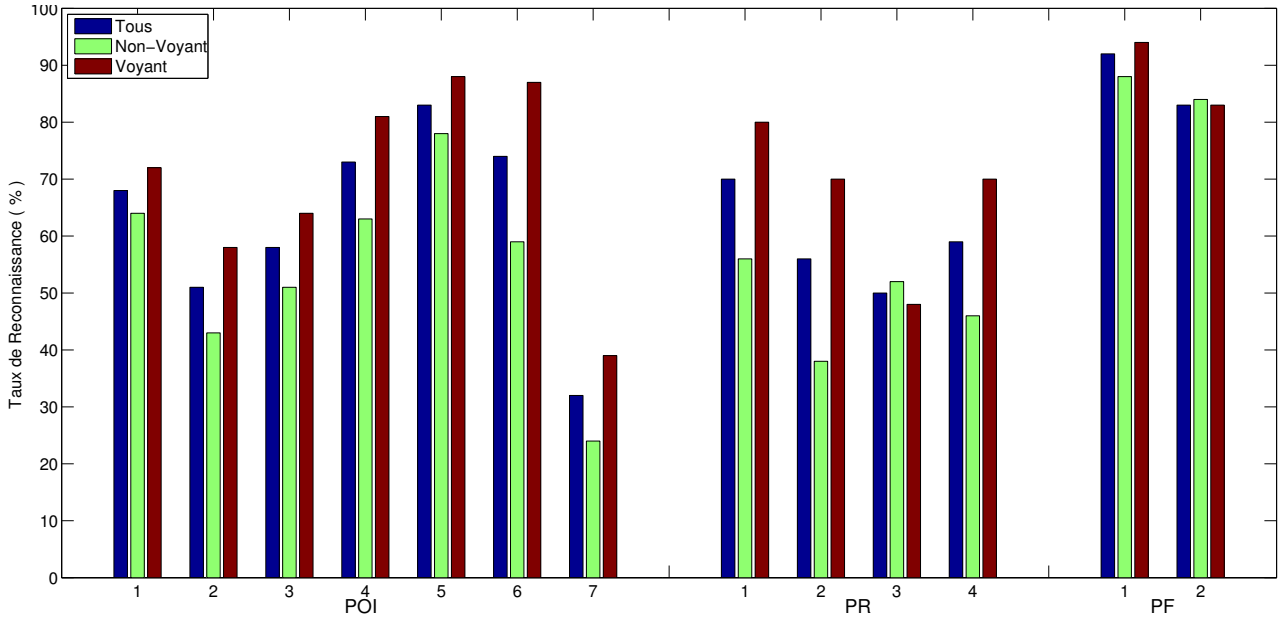


FIGURE 5.7 – Moyenne du taux de reconnaissance en fonction du type de cécité pour chaque sous-catégorie de POI (à gauche), de PR (au milieu) et de PF (à droite).

confusion n'était pas symétrique, les PF étant identifiés comme des POI mais pas l'inverse).

b) Différences entre les sous catégories

Test2 : Sous catégories de POI

Pour la deuxième phase qui consistait en l'identification des sept sous-catégories de POI, cinq sujets (3 non-voyants et 2 voyants) n'ont pas achevé le test et ont été exclus des résultats. Les résultats ont donc été calculés sur 26 sujets non-voyants et 29 sujets voyants.

Le taux d'identification moyen de la sous-catégorie de POI a été de $63 \pm 23\%$ ($68 \pm 20\%$ pour les voyants ; $56 \pm 26\%$ pour les non-voyants). La figure 5.7 (à gauche) présente les résultats par sous-catégories de POI, pour tous les sujets, en séparant voyants et non-voyants. Comme pour le premier test, on observe des performances inférieures chez les non-voyants comparées aux voyants (de 6% pour le POI1 à 30% pour le POI6). Les sons utilisés pour le POI7 (décrit comme un son complexe, sans hauteur définie) a produit des confusions chez tout les sujets (31% d'identification correcte en moyenne) alors que les sons du POI5 (décrits comme étant composés de trois notes) ont été très bien identifiés (la moyenne du taux d'identification étant de 81%).

La répartition détaillée des résultats pour chaque sous-catégorie de POI est représentée figure 5.8 (en haut), avec le taux d'identification pour chaque POI et les taux de confusions entre chaque POI, pour toutes les palettes confondues. Les résultats montrent un bon taux d'identification pour les sons des POI4, POI5 et POI6 (respectivement définis par deux notes descendantes, trois notes et un son court répété quatre fois. Le POI7 a été principalement confondu avec le POI1 (son harmonique grave) et plus légèrement avec le POI2 et le POI6. Une confusion entre les POI1 et POI2 (son

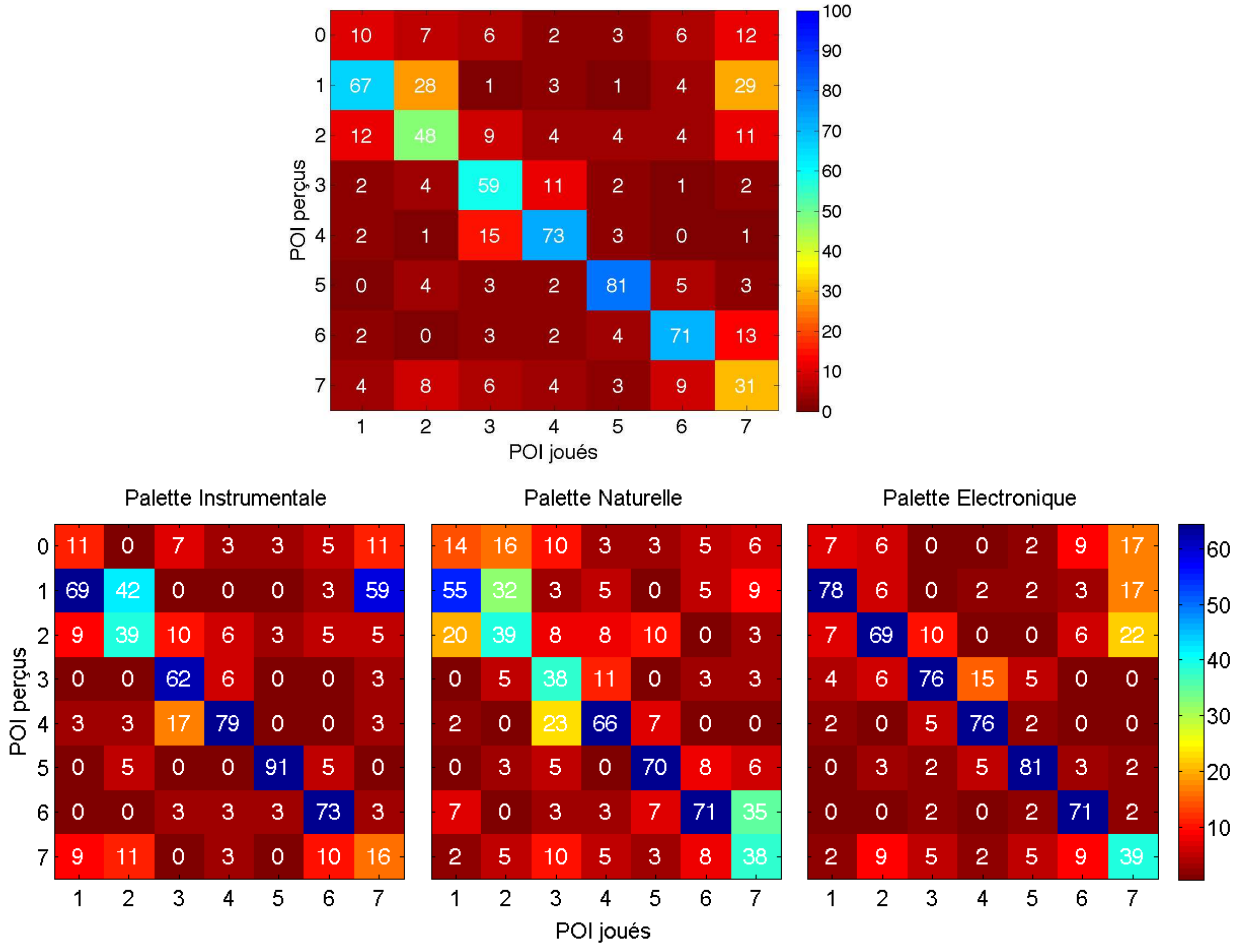


FIGURE 5.8 – Matrices de similarités entre les sous-catégories de POI jouées et les POI reconnues pour toutes les palettes confondues (en haut) et pour chaque palette (en bas).

harmonique grave et son harmonique aigu) a aussi été observée ainsi qu'entre les POI3 et POI4 (deux notes ascendantes et descendantes). Ces confusions peuvent être attribuées aux similarités morphologiques entre les sons de ces deux couples. Les matrices de confusion des figures 5.8 (en bas) permettent d'explorer les erreurs au sein de chaque palette. On remarque par exemple que la palette *électronique* a posé beaucoup moins de problèmes que les autres palettes (hormis pour le POI7). En moyenne, les sous-catégories de POI ont été correctement identifiées à 61% pour la palette *instrumentale*, 54% pour la palette *naturelle* et 70% pour la palette *électronique*.

Test 3 : Sous catégories de PR

Pour la troisième phase du test, qui consistait en l'identification des sous catégories de PR, huit sujets (6 non-voyants et 2 voyants) n'ont pas répondu à tout le test et ont été exclus des résultats. Les résultats ont donc été calculés à partir des réponses de 23 sujets non-voyants et 29 sujets voyants.

Pour ce test, le taux d'identification général a été de $58 \pm 29\%$ ($66 \pm 25\%$ pour les voyants et $47 \pm 29\%$ pour les non-voyants). Les résultats détaillés pour chaque sous-catégorie en fonction

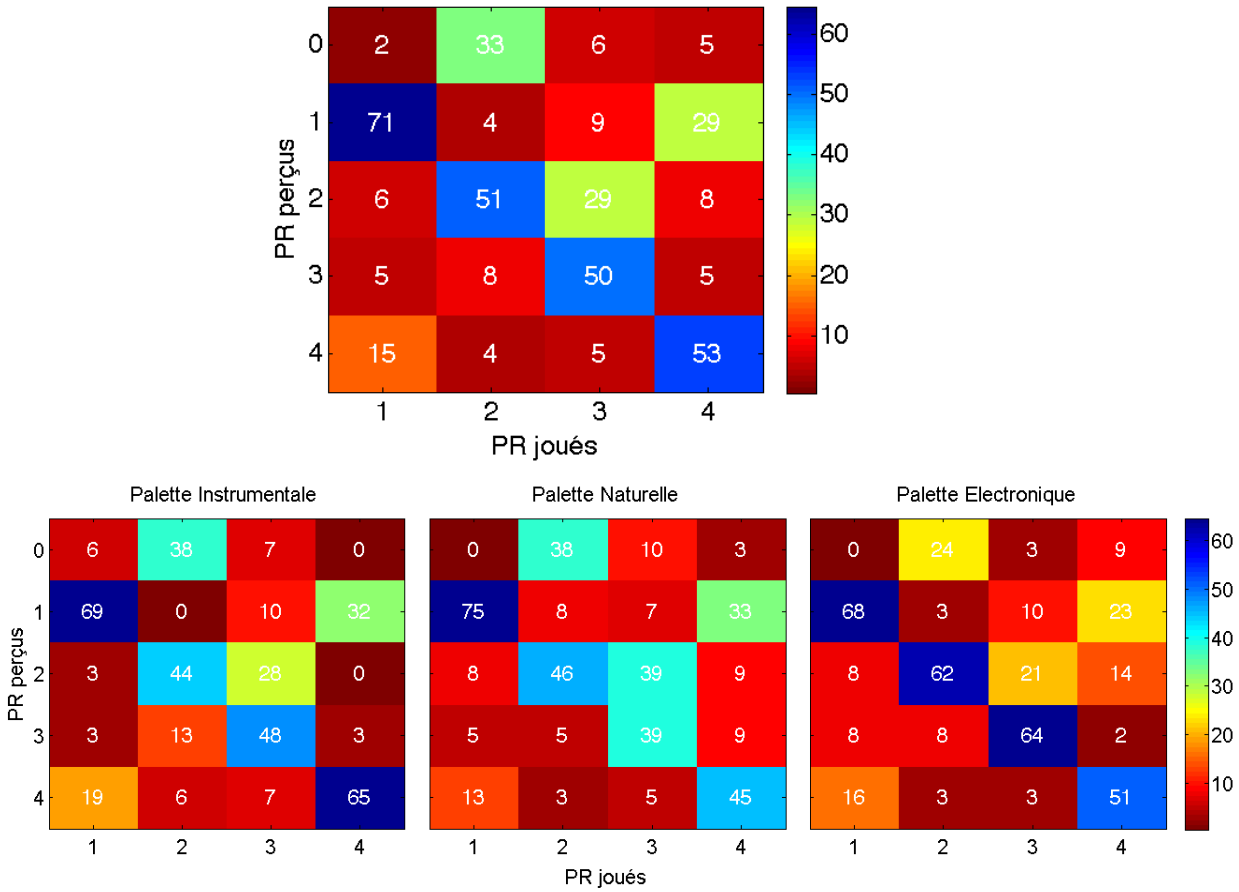


FIGURE 5.9 – Matrices de similarités entre les sous-catégories de PR jouées et les PR reconnues pour toutes les palettes confondues (en haut) et pour chaque palette (en bas).

du type de cécité sont présentés figure 5.7 (au centre). Une large différence de performance est observable entre les voyants et les non-voyants (excepté pour le PR3). Le taux d'identification moyen est plus faible que pour le test sur les sous-catégories de POI (alors que le nombre de PR à identifier était inférieur). Le détail de la répartition des identifications est présenté figure 5.9 (en haut) pour toutes les palettes confondues. Elle met en évidence une confusion entre les motifs rythmiques PR1 et PR4 (PR4 identifié comme un PR1) et entre PR2 et PR3 (PR3 identifié comme un PR2). On remarque aussi un taux élevé de non identification pour les sons des PR2 (33%). La figure 5.9 (en bas) met en évidence cette répartition en fonction du type de palette. On remarque comme pour les POI, que les éléments de la palette *électronique* ont été, en général, mieux reconnus que ceux des autres palettes et que la palette *naturelle* a posé plus de difficultés. En moyenne, les sous-catégories de PR ont été correctement identifiées à 56% pour la palette *instrumentale*, 51% pour la palette *naturelle* et 61% pour la palette *électronique*.

Test 4 : Sous catégories de PF

Pour la quatrième phase, qui consistait en l'identification des sous-catégories de PF, six sujets (5 non-voyants et 1 voyant) n'ont pas terminé tout le test et n'ont pas été inclus dans les résultats qui ont donc été réalisés sur 24 sujets non-voyants et 30 sujets voyants.

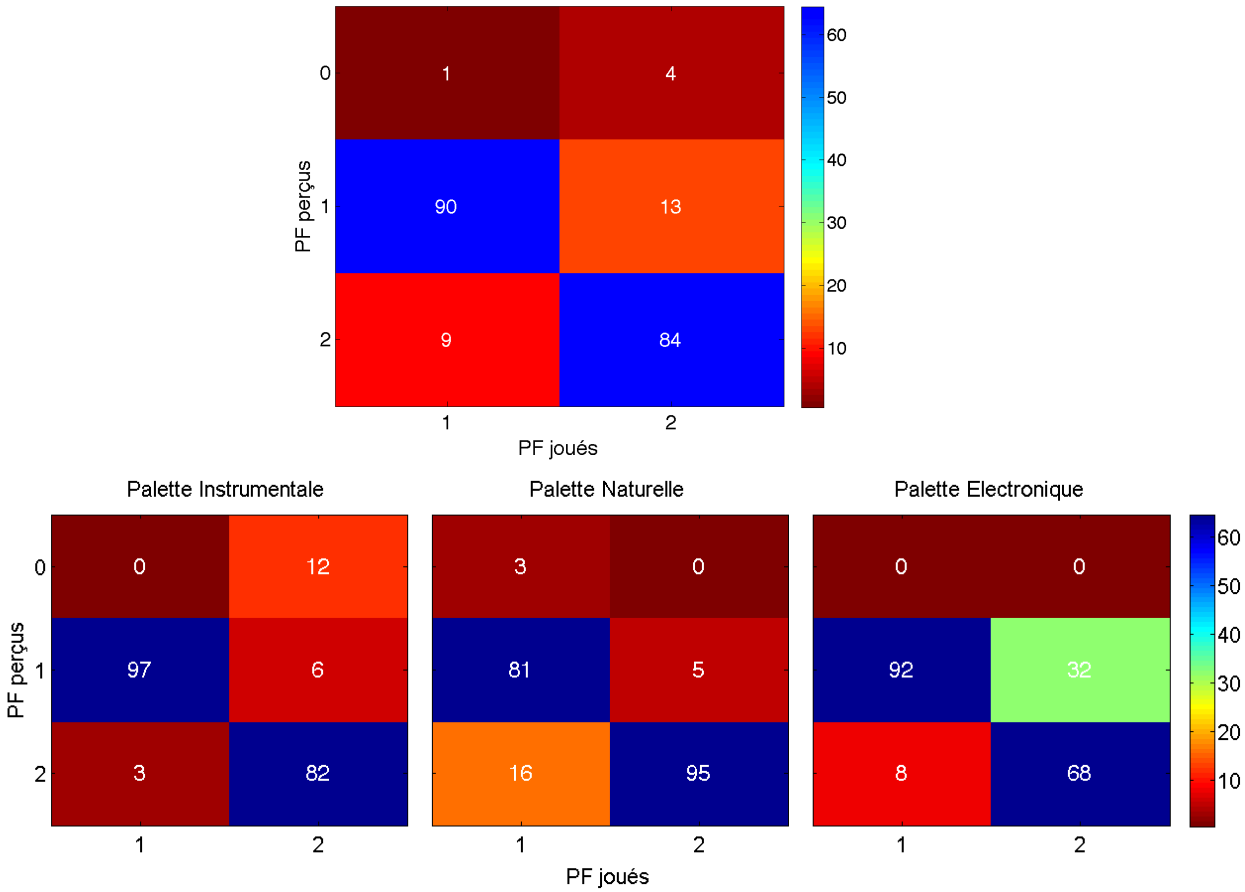


FIGURE 5.10 – Matrices de similarités entre les sous-catégories de PF jouées et les PF reconnues pour toutes les palettes confondues (en haut) et pour chaque palette (en bas).

Le taux d'identification général pour les sons de PF a été de $87 \pm 19\%$ ($87 \pm 17\%$ pour les voyants $86 \pm 21\%$ pour les non-voyants). Les résultats de ce test sont présentés figure 5.7 (à droite) et figure 5.10. Étant donné que les sujets n'avaient à identifier que deux sous-catégories différentes, les taux d'identifications sont élevés et aucune différence entre les voyants et les non-voyants n'est apparue. On remarque cependant figure 5.10 (en bas) que le son utilisé pour le PF2 de la palette *électronique* a été bien plus confondu avec le son du PF1 que pour les autre palettes. En moyenne, les sous-catégories de PF ont été correctement identifiées à 90% pour la palette *instrumentale*, 88% pour la palette *naturelle* et 80% pour la palette *électronique*.

5.5.3 Discussion

Dans un premier temps, les résultats de ces tests d'identification montrent que les sujets sont capables d'identifier des sons sur la base d'une description morphologique, tout en faisant abstraction du contenu sémantique et de la cause du son. Ce résultat permet ainsi de valider cette nouvelle approche basée sur des palettes sonores conçues à partir d'un ensemble de morphocons.

Les résultats obtenus aux différents tests donnent des lignes directrices, pour modifier le voca-

bulaire mis en place, lorsque certains morphocons sont mal reconnus et pour modifier les palettes, lorsque certains sons ont mal été identifiés.

Pour le test 1, le fort taux d'identification montre que les formes temporelles choisies pour différencier les catégories sont bien reconnues et efficaces. Malgré tout, dans un cas : le PI de la palette *naturelle* (un son d'oiseau), l'icône a été perçu comme une séquence de deux (pour les *Point Difficiles*) ou trois sons (pour les *Point de Repères*). L'analyse du spectrogramme du son en question (figure 5.3, première colonne, troisième ligne) montre qu'il est composé de plusieurs petits évènements (quatre) et que ceci a posé problème aux sujets, même si sa durée totale était inférieure à 0.2 seconde.

Les résultats du test 2 mettent en évidence quelques problèmes pour le design des morphocons de certaines sous-catégories de POI. Le moins efficace a été le son complexe, utilisé pour le POI7 et décrit comme ne contenant pas de hauteur. Ce son a été peu reconnu et souvent confondu avec la sous-catégorie POI1 (décrit comme un son harmonique grave). Il apparait que les sujets ont eu du mal avec la notion de "son complexe", il est donc nécessaire de trouver une autre forme morphologique pour cette sous-catégorie. Au niveau de la conception des palettes sonores, il est nécessaire d'accentuer davantage les différences entre les sons graves et les sons aigus (confusion entre POI1 et POI2) et de bien marquer les hauteurs, dans le cas des séquences ascendante ou descendante (POI3 et POI4), afin de diminuer les confusions entre ces couples d'icônes. En général, malgré leur grand nombre, les sous-catégories de POI ont été bien reconnues.

Le test 3 (identification des sous-catégories de *Point de Repères*) s'est révélé être la partie la plus difficile du test et a conduit à de faibles résultats. Il semble que les rythmes choisis étaient trop similaires ou qu'une séquence de trois sons n'est pas suffisante pour créer une bonne perception du rythme sans connaissance du tempo. Une solution pourrait être de combiner le rythme avec un autre type de variation ou d'utiliser au moins quatre sons pour cette catégorie.

Le test 4, n'a posé quasiment aucune difficulté. Les résultats de ce test nous ont cependant conduit à modifier le son correspondant au PF2 de la palette *électronique*.

De manière générale, les différences de performances observées entre les voyants et non-voyants peuvent avoir deux causes. D'une part, bien que conçu pour les non-voyants (avec des raccourcis clavier pour écouter les sons de la palette *exemple*), l'interface du formulaire internet était un peu lourde et plusieurs sujets non-voyants ont rapporté avoir eu certaines difficultés pour se déplacer entre les champs de saisie et les différents boutons de lecture. Ceci pourrait expliquer pourquoi certains sujets (principalement non-voyants) n'ont pas achevé toutes les sections du test. D'autre part, la majorité des sujets voyants ayant répondu au test faisaient partie de la communauté "Auditory List" et peuvent donc être considérés comme des experts en son ce qui n'était pas le cas pour les non-voyants. Ce constat oblige à tempérer les observations relatives aux différences de performances entre voyants et non-voyants.

5.6 Conclusion

Ce chapitre a présenté une extension du concept d'earcon vers le concept de morphocon. Les morphocons sont basés sur des profils morphologiques de paramètres acoustiques permettant de créer un langage sonores pouvant être appliqués à n'importe quel type de sons. Ce nouveau type d'icône sonore est conçu pour répondre aux attentes des utilisateurs en terme de paramétrisation des messages audio. Sur la base de ce concept, un vocabulaire sonore a été conçu dans le cadre du projet NAVIG et appliquée à trois types de sons pour créer trois palettes différentes (*instrumentale*, *naturelle*, et *électronique*). Ces trois palettes ont été évaluées par un test perceptif de catégorisation disponible sur internet. Les résultats indiquent que tous les sujets sont capables de percevoir les variations temporelles de paramètres acoustiques comme des formes abstraites et de les reconnaître même si elles sont appliquées à différents types de sons.

Cette étude ne se basant que sur l'efficacité des morphocons, il serait intéressant de faire davantage de tests pour les comparer aux earcons, aux auditory icons ainsi qu'aux spearcons, en prenant en compte les performances, la facilité d'apprentissage et la satisfaction des utilisateurs pour une tâche spécifique.

Étant donné que les morphocons sont décrits en utilisant des profils morphologiques simples, ils peuvent être appliqués à tous les types de sons. La création de nombreuses palettes pourrait être envisagée de façon automatique en utilisant plusieurs types de méthodes :

- La synthèse par concaténation de petit segments audio. Ce type de synthèse utilise une base de données de sons enregistrés, et un algorithme de sélection d'unités qui choisit les segments de la base de données qui conviennent le mieux pour la séquence sonore que l'on souhaite synthétiser par concaténation [Schwarz, 2007].
- L'utilisation de descripteurs audio permettant de rechercher des profils définis dans des banques de sons. Plusieurs équipes ont mis au point des descripteurs audio permettant de rechercher dans des extraits sonores des profils morphologiques particuliers [Peeters et Deruty, 2008].
- L'application d'une série d'effets sonores (modification de l'enveloppe, du pitch, ...) à un son donné, afin de modeler le profil morphologique que l'on souhaite obtenir. Mettre en place une telle méthode pourrait, par exemple, permettre au système d'intégrer des sons que l'utilisateur a lui-même fournis au système.

Il serait aussi intéressant d'examiner, si ces formes morphologiques peuvent être transmises en utilisant d'autres modalités sensorielles, telles que l'haptique (en utilisant par exemple des *tactons* [Brewster et Brown, 2004]).

Chapitre 6

Conclusion générale

Sommaire

Dans ce manuscrit, nous avons abordé la problématique suivante : est-il possible d'utiliser le son 3D (et plus particulièrement la synthèse binaurale) et la sonification dans un dispositif d'aide à la navigation pour les non-voyants ? Les contributions apportées par cette thèse à cette problématique sont récapitulées dans la section 6.1. Si ces contributions apportent quelques réponses notamment appliquées au cas des systèmes d'aide à la navigation, elles soulèvent de nombreuses perspectives pour les différents domaines pouvant utiliser le son 3D et la sonification. La section 6.2 présentera quelques unes d'entre elles.

6.1 Contributions de la thèse

Trois contributions ont été apportées par cette thèse pour répondre à la problématique mentionnée précédemment selon les trois axes principaux définis dans l'introduction (chapitre 1).

6.1.1 Amélioration du rendu binaural avec des HRTF non-individuelles

Comment régler le problème de l'individualisation des HRTF ?

Nous avons vu au cours du manuscrit qu'un rendu sonore 3D effectué avec de la synthèse binaurale sans HRTF individuelles entraîne de nombreux problèmes de localisation pouvant être très gênants dans le cadre d'une application de navigation. La mesure des HRTF individuelles des sujets dans un cadre commercial n'étant pas envisageable et l'individualisation d'HRTF génériques par des méthodes de traitement du signal étant un problème encore très complexe et loin d'être résolu, il est nécessaire de trouver une méthode plus simple pour améliorer le rendu binaural. La méthode que nous avons mise en place dans cette thèse (chapitre 3), consiste, plutôt que d'adapter les HRTF à l'individu, à adapter l'individu aux HRTF (en utilisant les mécanismes de plasticité cérébrale). Elle permet d'adapter graduellement le sujet en utilisant son meilleur jeu d'HRTF comme point de départ. Un petit test de jugement perceptif permet au sujet de sélectionner dans une base réduite d'HRTF (moins de dix jeux d'HRTF), les HRTF les plus proches des siennes. L'ITD de ces HRTF est individualisé (à partir de la mesure de la circonférence de la tête du sujet) pour créer des HRTF hybrides. Ensuite, en utilisant un jeu sur une plateforme multimodale, un minimum de trois sessions d'adaptation permet à l'utilisateur de rejoindre les performances qu'il aurait obtenu avec des HRTF individuelles et même de les surmonter.

6.1.2 Amélioration des indices de localisation par l'utilisation de la sonification

Quelles sont les capacités de localisation et de saisie d'objets sonores dans l'espace péripersonnel ?

Deux expériences ont été mises en place pour mesurer les capacités de localisation et de mouvement de saisie vers des sons réels et virtuels dans l'espace péripersonnel (chapitre 4). La première expérience (réalisée avec des sons réels), a permis, d'une part, d'établir les performances de localisation du système auditif dans cette zone très peu étudiée dans la littérature, et d'autre part, de comparer les capacités de mouvements de saisie en fonction de la main utilisée. Les résultats ont montré que dans la zone étudiée, la précision angulaire est meilleure devant et plus faible sur les cotés et que l'erreur moyenne en azimut est de 12°. Concernant la perception de la distance, les résultats de cette étude montrent que le système auditif a tendance à compresser la distance perçue dans la zone étudiée et que cette compression est plus forte dans la zone frontale que dans les zones latérales. L'étude de la précision en fonction de la main de pointage montre qu'il n'y a pas de différence entre les mouvements de saisie effectués avec la main principale et ceux effectués avec la main secondaire. La deuxième expérience, réalisée avec des sons virtuels, a mis en évidence une grande dégradation des indices de localisation avec la synthèse binaurale. Les performances de localisation angulaire sont fortement dégradées (l'erreur moyenne en azimut est d'environ 24°), avec des performances meilleures sur les côtés que dans la zone frontale. La perception de la distance, quant à elle, est quasiment inexistance (avec toujours des performances meilleures sur les côtés que dans la zone frontale). Cette expérience a ainsi permis de mettre en

évidence l'incapacité de la synthèse binaurale, telle qu'elle est réalisée aujourd'hui, à reproduire des indices de distance.

Est-il possible d'améliorer les indices de perception de la distance en utilisant la sonification ?

Afin d'améliorer la perception de la distance d'une source virtuelle générée par de la synthèse binaurale, une nouvelle méthode de sonification a été mise en place. Inspirée de la méthode de sonification par mapping de paramètres, cette méthode consiste à utiliser des effets sonores pour ajouter au son original des informations sur la distance. Ces effets, dont un ou plusieurs paramètres évoluent en fonction de la distance de la source, permettent de modifier les caractéristiques du son virtuel et d'ajouter des indices acoustiques améliorant la perception de la distance. Trois métaphores de sonification de la distance ont été réalisées selon cette méthode. Une expérience de localisation, destinée à quantifier l'apport de ces métaphores de sonification, a permis de montrer que deux d'entre elles permettent d'augmenter significativement la perception de la distance et ce malgré un temps d'apprentissage très court. Les résultats avec ces deux métaphores dépassent les performances obtenues dans le cas idéal (i.e. avec des sons réels) montrant ainsi qu'il pourrait être possible, en choisissant correctement les effets et les paramètres à faire varier, d'atteindre des performances de localisation parfaites (pour lesquelles la distance perçue serait égale à la distance simulée).

6.1.3 Sonification personnalisable par les utilisateurs

Comment mettre en place des méthodes de sonification permettant de satisfaire les critères esthétiques de tous les utilisateurs ?

Bien que la sonification permette efficacement de transmettre tout type d'informations sous forme sonore, elle souffre bien souvent d'un manque d'esthétisme qui réduit sa capacité a être utilisée dans le cadre d'une application grand publique à usage fréquent.
Afin de résoudre ce problème, une nouvelle méthode de sonification a été mise au point pour permettre la création de vocabulaires sonores indépendantes du type de son utilisé lorsqu'il est nécessaire de transmettre plusieurs informations sous la forme de messages sonores (chapitre 5). Les *morphocons* sont des petites entités sonores héritant des propriétés des *earcons* et dont la construction se fait sur la base de descriptions de l'évolution temporelle du son. Un vocabulaire de *morphocons* a été mis en place et appliqué à trois types de sons (naturels, instrumentaux et artificiels). Un test de catégorisation a permis de montrer que les utilisateurs sont capables de percevoir les variations temporelles de paramètres acoustiques comme des formes abstraites et de reconnaître les messages transmis indépendamment du type de son utilisé. Ce résultat montre qu'il est possible de concevoir des interfaces sonores permettant à l'utilisateur de choisir le type de son sans nécessiter de nouvel apprentissage des différents sons de l'interface. Cette méthode a donc un fort potentiel à satisfaire les critères esthétiques de chacun des utilisateurs étant donné que les sons

ne sont plus choisis par le designer mais sélectionnés par l'utilisateur lui-même.

6.2 Perspectives de recherche

Nous allons donner dans cette section, quelques perspectives ouvertes par cette thèse.

6.2.1 Mise en place et test d'un dispositif de navigation en situation réelle

Cette thèse a permis de montrer que le son 3D et la sonification peuvent être utilisés dans un dispositif d'aide à la navigation pour les non-voyants. Les premiers tests sur le prototype d'aide à la navigation réalisés dans le cadre du projet NAVIG ont montré qu'il est possible de coupler différents modules (GPS, vision artificielle, moteur de sonification 3D, etc.) et de les faire fonctionner sur un même ordinateur pour guider un utilisateur le long d'un trajet ou l'aider à saisir un objet. Pour aller plus loin, il est nécessaire de tester l'efficacité de la sonification 3D en la comparant aux méthodes de guidage plus traditionnelles (telles que la description d'itinéraire sous forme vocale) et de valider les approches de sonification adoptées avec des expériences permettant d'évaluer la satisfaction des utilisateurs. L'efficacité de la méthode de guidage pourra être testée en environnement contrôlé (sur un stade ou un grand espace libre), permettant de comparer les différentes méthodes de guidage avec plusieurs scénarios de navigation ayant les mêmes caractéristiques (même longueur de trajet, même nombre de changements de trajectoire, etc.). Des expérimentations en situation réelle (en ville) seront ensuite indispensables pour évaluer la satisfaction des utilisateurs face au rendu sonore, ainsi que l'ergonomie du système dans sa globalité.

6.2.2 Généralisation de la méthode d'apprentissage des HRTF non-individuelles

Les améliorations acquises avec la méthode d'adaptation aux HRTF non-individuelles ont été évaluées avec seulement une ou trois sessions d'apprentissage de 12 minutes chacune. Étant donné la courbe d'évolution des résultats, il apparaît que l'apprentissage n'est pas forcément complet et que plus de sessions d'apprentissage pourraient mener à de meilleures performances. Afin de valider cette hypothèse, il serait intéressant de tester la méthode d'apprentissage sur une plus longue durée. De plus, il serait intéressant de tester la pérennité de cette adaptation en évaluant les performances de localisation avec les HRTF non-individuelles quelques semaines après leur apprentissage. Afin d'utiliser cette méthode pour un système de réalité augmentée (comme le projet NAVIG) avec une restitution sur casque osseux, il faut qu'il n'y ait pas de conflit entre la carte audio spatiale réelle de l'utilisateur et la carte virtuelle apprise. Il serait intéressant d'étudier ces effets potentiels.

6.2.3 Évaluation de l'ergonomie des méthodes de sonification mises en place

Les méthodes de sonifications mises en place dans les chapitres 4 et 5 ont été évaluées en terme d'efficacité ou de performance. Il a été montré que la méthode du chapitre 4 permet d'améliorer la perception de la distance indépendamment du type de son utilisé. Pour la méthode mise en place dans le chapitre 5, il a été montré que les utilisateurs sont capables de reconnaître les messages transmis indépendamment du type de son utilisé. Afin de valider l'aspect novateur de ces approches en terme d'esthétique, il est nécessaire de tester ces méthodes de sonification avec des techniques issues de l'ergonomie pour évaluer la satisfaction des utilisateurs lorsqu'ils ont le choix du type de son à utiliser.

6.3 Publications liées à la thèse

Cette section regroupe les différentes publications qui sont liées au manuscrit présenté ici.

6.3.1 Articles de revue à comité de lecture

- **[Parseihian et Katz, 2012b]** : G. PARSEIHIAN & B. F.G. KATZ
  "Rapid Head-Related Transfer Function adaptation using a virtual auditory environment", *Journal of Acoustical Society of America*, Volume 131, Issue 4, pp. 2948-2957 (2012).
- **[Parseihian et Katz, 2012a]** : G. PARSEIHIAN & B. F.G. KATZ
  "Morphocons : A new sonification concept based on morphological earcons", *Journal of the Audio Engineering Society*, Volume 60, Issue 6, pp. 409-418 (2012).
- **[Katz et Parseihian, 2012]** : B. F.G. KATZ & G. PARSEIHIAN
  "Perceptually based head-related transfer function database optimization", *Journal of Acoustical Society of America*, Volume 131, Issue 2, pp. EL99-EL105 (2012).
- **[Kammoun *et al.*, 2012]** : S. KAMMOUN, G . PARSEIHIAN, O. GUTIERREZ, A. BRILHAULT, A. SERPA, M. REYNAL, O. ORIOLA, M. MACÉ, M. AUVRAY, M. DENIS, S. THORPE, P. TRUILLET, B. F.G. KATZ & C. JOUFFRAIS
  "Navigation and space perception assistance for the visually impaired : The NAVIG project", *IRBM*, Volume 33, Issue 2, pp. 182-189 (2012).
- **[Katz *et al.*, 2012b]** : B. F.G. KATZ, S. KAMMOUN, G. PARSEIHIAN, O. GUTIERREZ, A. BRILHAULT, M. AUVRAY, P. TRUILLET, M. DENIS, S. THORPE & C. JOUFFRAIS
  "NAVIG : augmented reality guidance system for the visually impaired. Combining object localization, GNSS, and spatial audio", *Virtual Reality*, Volume 16, Issue 3 (2012).
- **[Katz *et al.*, 2012a]** : B. F.G. KATZ, F. DRAMAS, G. PARSEIHIAN, O. GUTIERREZ, S. KAMMOUN, A. BRILHAULT, L. BRUNET, M. GALLAY, B. ORIOLA, M. AUVRAY, P. TRUILLET, M. DENIS, S. THORPE & C. JOUFFRAIS
  "NAVIG : Guidance system for the visually impaired using virtual augmented reality", *Journal of Technology and Disability*, Volume 24, Issue 2 (2012).

6.3.2 Conférences avec actes

- **[Parseihian *et al.*, 2012]** : G. PARSEIHIAN, S. CONAN & B. F.G. KATZ
  "Sound effect metaphors for near field distance sonification", *Proceedings of the 18th international conference on Auditory display (ICAD 2012)*, Atlanta, USA, June 18-22, 2012.
- **[Parseihian *et al.*, 2010]** : G. PARSEIHIAN, A. BRILHAULT & F. DRAMAS
  "NAVIG : An object localization system for the blind". *Workshop Pervasive 2010 : Multimodal Location Based Techniques for Extreme Navigation*, Helsinki, May 17, 2010.

6.3.3 Conférences sans actes

- **[Parseihian et Katz, 2011]** : G. PARSEIHIAN & B. F.G. KATZ
  "Rapid auditory system adaptation using a virtual auditory environment". *International Multisensory Research Forum*, Fukuoka, Japan, October 17-20, 2011.
- **[Parseihian et Katz, 2010]** : G. PARSEIHIAN & B. F.G. KATZ
  "Recalibrating the auditory system through audio-kinesthesic training". *European Workshop on Imagery & Cognition*, Helsinki, Finland, June 16-19, 2010.
- **[Gallay *et al.*, 2010]** : M. GALLAY, M. DENIS, G. PARSEIHIAN & M. AUVRAY
  "Egocentric and allocentric reference frames in a virtual auditory environment : differences in navigation skills between blind and sighted individuals". *European Workshop on Imagery & Cognition*, Helsinki, Finland, June 16-19, 2010.

Annexe A

Besoins utilisateur et conception participative

Le système NAVIG n'est pas conçu dans le but de remplacer les chiens-guides ou la canne blanche. Il doit être considéré comme un complément à ces dispositifs et aux séances d'entrainement en mobilité et orientation que les non-voyants peuvent suivre dans les instituts spécialisés. Par conséquent, l'objectif principal du dispositif NAVIG est d'aider les non-voyants à améliorer, d'une part, leur autonomie au quotidien, et d'autre part, leur capacité à se faire une représentation mentale de leur environnement (au-delà de ce qui est possible avec les dispositifs tels que la canne et les chiens, qu'ils ont déjà l'habitude d'utiliser).

Pour concevoir un tel dispositif, il est primordial de comprendre les besoins des utilisateurs, les problèmes qu'ils rencontrent au quotidien lors de leur déplacement ou lors de l'utilisation de dispositif de suppléance. Afin de mettre en place un dispositif accessible et utilisable, le projet s'est basé sur l'intégration des utilisateurs dans un processus de conception participative.

La conception participative est une méthode de travail utilisée dans la conception de logiciel ou de dispositif interactif. Elle consiste à impliquer les utilisateurs dans l'ensemble des processus de développement. De nombreux travaux ont été réalisés dans le domaine de l'ergonomie pour développer des méthodes et des outils permettant de mettre en oeuvre de tels processus. Quelques règles de base pour la mise en oeuvre sont définies par la norme ISO 13407 [ISO, 3407]. Il est entre autre important de consulter les utilisateurs dans les différentes phases du processus qui comprennent :

- L'analyse des besoins et des activités des utilisateurs
- La production d'idées
- La conception et le prototypage
- L'évaluation des conceptions par rapport aux exigences

Plusieurs travaux ont été réalisés dans le cadre du projet pour ces différentes phases. Les travaux de [Brock *et al.*, 2010] détaillent ces quatre phases ainsi que leur application à la conception d'un dispositif de saisie et de planification d'itinéraire. L'analyse de l'activité des non-voyants en situa-

tion de préparation d'itinéraire et de déplacement a été réalisée à Paris par [Brunet, 2010]. Cette analyse, réalisée sur six sujets non-voyants, est présentée dans [Brunet *et al.*, 2012].

Concernant l'interface de sortie sonore, nous avons organisé deux groupes de travail avec des non-voyants utilisateurs de l'IJA de Toulouse pour, d'une part, analyser les besoins des utilisateurs, et d'autre part, produire des idées sur la manière de transmettre les informations aux utilisateurs. Dans cette section, nous allons présenter ces deux réunions puis faire un bilan des différentes informations à fournir ainsi que les contraintes à prendre en compte pour la transmission des informations à l'utilisateur.

A.1 Brainstorming sur les informations à donner

Depuis le début du projet, les instructeurs en mobilité et les utilisateurs potentiels ont spécifiquement insisté sur le fait qu'un dispositif d'assistance doit être hautement personnalisable. Par exemple, en ce qui concerne les itinéraires proposés par le système, les utilisateurs les plus confiants et indépendants pourront demander le plus court chemin (même si cela implique des difficultés plus grandes), tandis que les utilisateurs moins expérimentés pourront préférer des trajets plus long avec moins de difficultés.

Afin d'identifier et de classer par niveaux de difficultés les points de l'itinéraire à privilegier ou à éviter dans le calcul de l'itinéraire optimal, une séance de réflexion de deux heures a été réalisée avec six non-voyants de l'IJA. Réalisé avec Mathieu Gallay (LIMSI) et Lucie Brunet (LIMSI), ce groupe de travail a permis d'une part de définir les principaux obstacles rencontrés dans le milieu urbain ainsi que les difficultés relatives aux différents types de croisement et d'autre part de recueillir l'avis des utilisateurs sur l'intérêt de signaler des informations additionnelles telles que des points de confirmation ou des points d'intérêt.

Au niveau des critères qui guident le choix d'un itinéraire en particulier, les participants se sont accordés sur le fait qu'il faut préférer : les passages simples, les rues calmes (avec peu de trafic piéton et automobile), les passerelles et rues piétonnes (pour permettre d'accélérer la marche) et le trajet le plus court si le temps presse. En outre, un ensemble d'éléments à éviter a également été défini :

- les ronds points et les gros carrefours, du fait de la difficulté à comprendre leur géométrie ;
- les places, les grands espaces, pour la difficulté à s'y orienter et à y marcher droit ;
- les trottoirs trop larges, où sont présents de nombreux obstacles ;
- les trottoirs trop étroits et les zones encombrées de poteaux et barrières, qui obligent souvent à marcher sur la chaussée ;
- les zones partagées où les voitures et les cyclistes sont difficile à détecter.

Ces résultats sont en accord avec ceux de [Gaunet et Briffault, 2005].

Pour affiner ces résultats afin d'être en mesure de les appliquer à un processus de sélection d'itinéraire automatique, une méthode de score de difficulté pour chacun des éléments a été établie.

Obstacles	Score de difficulté
Petits trottoirs	4.67 ± 0.8
Grand espaces	4.33 ± 0.5
Obstacles hauts (non détectables par la canne)	3.33 ± 1.4
Ronds points	3.17 ± 0.4
Carrefours sans feu	2.83 ± 0.4
Trottoirs très larges	2.67 ± 0.5
Terrasses de café	2.17 ± 0.4
Escaliers	2.00 ± 1.1

TABLE A.1 – Moyennes des scores de difficulté des principaux obstacles recencés et déviation standard. 1=pas de difficulté, 5=grande difficulté.

Les participants ont d'abord été invités à citer les trois types d'obstacles qu'ils trouvent les plus problématiques lors de leurs déplacements quotidiens. À partir de leurs réponses, une liste d'obstacles a été établie. Les participants ont ensuite évalué la difficulté de chacun de ces obstacles en utilisant une échelle de Likert (1=pas de difficulté, 5=grande difficulté). Les éléments obtenus et leur difficulté moyenne sont présentés dans le tableau A.1. Ces indices peuvent être utilisés lors du calcul d'itinéraire pour évaluer un niveau de difficulté global du parcours (en additionnant les indices de chacuns des obstacles rencontrés). Les utilisateurs pourront utiliser ce niveau de difficulté de deux manières :

- Le système présente les différents itinéraires possibles avec le niveau de difficulté et le temps de parcours prévu. À l'utilisateur de choisir le trajet qu'il souhaite emprunter.
- L'utilisateur fait une requête d'itinéraire en définissant un seuil de difficulté maximal. Le système choisit alors l'itinéraire le plus court possible correspondant aux critères choisis par l'utilisateur.

Concernant les croisements, il a été demandé aux utilisateurs d'ordonner chaque type de croisement en fonction de leur difficulté. À partir de leurs réponses, un coefficient de difficulté à prendre en compte dans le calcul de l'itinéraire a été défini pour chaque type de croisement :

- Croisement en T : coefficient 1
- Croisement en X : coefficient 2
- Croisement à n branches : coefficient $n - 2$ (pour un croisement en X : 4 branches - 2 = coefficient 2 ; pour un croisement à 6 branches, coefficient 4)
- Rond point à n branches : coefficient $n - 1$
- Modificateur : il est proposé de rajouter +1 au coefficient pour chaque présence de voie de bus, tramway, piste cyclable, terre-plein central, traversée en deux temps, contre allée, ...

Au niveau des informations additionnelles, les participants sont en accord sur le fait qu'il ne faut pas ajouter trop d'informations. Pour eux, l'ajout de point de confirmation (signalisation d'objet leur permettant de vérifier qu'ils sont au bon endroit), doit être en option et est surtout utile aux utilisateurs peu expérimentés. La signalisation de point d'intérêt (POI) leur paraît intéressante mais doit pouvoir être configurable selon au moins trois niveaux :

Participants	Âge	Déficience	Aide au déplacement utilisée	Expertise en nouvelles technologies
NV1	37	NVN	Canne, Trekker	Expert
NV2	56	NVN	Canne, Kapten	Expert
NV3	24	NVN	Chien, GPS	Expert
NV4	18	NVN	Canne	Novice
NV5	18	NVT	Canne, Trekker, Kapten	Expert
NV6	23	NVT	Canne	Novice

TABLE A.2 – Récapitulatif des participants et de leurs caractéristiques. NVN : Non-voyant de naissance, NVT : Non-voyant tardif.

- Ne signaler aucun POI ;
- Ne signaler que certaines catégories de POI ;
- Signaler tous les POI sur le parcours.

En général, les utilisateurs sont très attachés au fait de n'avoir que les informations suffisantes et nécessaires. En effet, les informations étant transmises selon la modalité auditive, la prise de repères auditifs naturels dans l'environnement pourrait être gênée par une surabondance de messages. La tâche de déplacement en ville demandant un niveau de concentration important, il ne faut pas que les sons du dispositif entraînent une charge cognitive supplémentaire. L'information doit donc permettre d'aider l'utilisateur de la façon la plus efficace, tout en étant la moins intrusive possible. De plus, le fait de donner trop d'information donne l'impression aux non-voyants d'être déresponsabilisés ce qui va à l'encontre de la volonté d'autonomie que le dispositif souhaite apporter.

A.2 Séance de production d'idée : la notion de "guidage idéal"

Afin de produire avec les utilisateurs potentiels des idées de guidage avec la modalité sonore, une séance de créativité autour du concept de "guidage idéal" par un dispositif auditif utilisant le son 3D a été réalisée. Cette séance de créativité, utilisant la méthodologie du brainstorming, avait deux buts :

- Favoriser la créativité de l'utilisateur afin de définir le concept de guidage idéal ;
- Trouver des idées de métaphores sonores innovantes par rapport à des points particuliers dans la navigation.

Le fait que ces pistes proviennent de l'imagination des utilisateurs cibles est une première étape vers la notion d'acceptabilité du dispositif.

A.2.1 Méthodologie

a) Participants

6 participants non-voyants ont participé à cette séance. Leur âge moyen était de 29 ± 15 ans. Quatre étaient des non-voyants de naissance et deux des non-voyants tardifs. Quatre se considéraient comme expert dans l'utilisation de nouvelles technologies (ordinateur, système d'aide au déplacement, ...), alors que deux étaient novices. Le tableau A.2 récapitule les caractéristiques de chacun des participants.

b) Passation

D'une durée de deux heures, cette séance de créativité était basée sur la méthodologie du brainstorming.

Le brainstorming ou "remue-méninges" est une méthode de réunion de groupe visant à favoriser la créativité des participants et récolter un grand nombre d'idées originales pour l'élaboration d'un produit ou d'un concept. Le brainstorming est défini par deux grand principes : le non-jugement (juger, que ce soit en bien ou en mal, tue la créativité et ne permet pas l'enchaînement des idées) et la recherche la plus étendue possible. Ces deux principes ce traduisent par quatre règles [Rickards, 1999] :

- Toute critique est interdite ;
- Se laisser aller (les idées farfelues sont les bienvenues) ;
- Rebondir (on cherche la combinaison et l'amélioration des idées) ;
- Chercher à obtenir le plus grand nombre d'idées possibles.

Avec des participants voyants, toutes les idées produites peuvent être notées sur des papiers ou sur un tableau permettant de favoriser une interaction dynamique collective, le partage, ainsi que la structuration, la réorganisation et le choix des idées. Les supports et les méthodes d'interaction utilisés pendant les brainstormings étant essentiellement visuelles, nous avons cherché des moyens de favoriser la créativité en utilisant des supports accessibles aux non-voyants. Ainsi, les informations écrites sur le tableau étaient régulièrement relues et l'utilisation de support de type braille et audio était privilégiée.

La séance était organisée en quatre parties :

1. *Prise de contact :* le but de cette phase était de mettre les participants en confiance, d'évaluer leur niveau d'expertise en nouvelles technologies, de connaître les types d'aides au déplacement qu'ils utilisent et enfin de leur expliquer le concept du brainstorming.
2. *Production d'idée autour du concept de guidage idéal :* cette partie consistait à élaborer des concepts de guidage idéal sans aucune limitation d'idées et de définir des pistes sur les meilleures façons de transmettre les informations de guidage en amont et pendant la navigation. Nous nous sommes basés pour cela sur les techniques de brainstorming onirique (rêve éveillé volontaire [Aznar, 2007]) : l'animateur entraîne son groupe en utilisant des techniques

de décontraction, de déconnexion du quotidien, d'éloignement vers un imaginaire où chacun exprime ses "rêves".

Cette partie s'est déroulée en trois phases basées sur un scénario de navigation inconnu des participants. Ce trajet, une ballade touristique de cinq minutes dans les rues du centre de Paris, était enregistré en binaural. L'utilisation de l'enregistrement binaural permet d'immerger les participants dans l'environnement urbain parisien (pas forcement très décontractant) et ainsi de les déconnecter de la salle de réunion.

- Écoute commune (10 minutes) : Écoute du trajet enregistré en binaural avec des description orales faites par le porteur des microphones pendant la réalisation du trajet. Le but de cette écoute est d'immerger l'auditeur dans un environnement sonore et d'utiliser le concept d'audiovision pour lui permettre de s'imaginer en situation de ballade. L'écoute était accompagnée d'une plage de braille comportant les informations sur le trajet (noms de rues et directions).
- Travail en groupe (15 min) : Travail en deux groupes à partir du scénario de navigation écouté. Chaque groupe crée différents types de guidage destinés à être présentés en amont et pendant la navigation. Chaque groupe est accompagné par un animateur.
- Mise en commun (20 min) : Chaque groupe présente les résultats de ses réflexions sur le scenario établi au préalable. Puis l'ensemble des participants débatent et argumentent autour des idées qui en sont sorties afin de les améliorer et d'en proposer des nouvelles.

3. *Découverte du son 3D :* Après une partie ouverte à tous les types de propositions de guidage (sonores, tactiles, etc.), nous cherchons à recentrer l'utilisateur sur le concept de son 3D. Cette partie a pour objectif de sensibiliser les participants à l'écoute spatiale et à la technologie du son binaurale afin de poursuivre la réunion sur la création de différents types de guidages sonores réalisables avec cette technologie.
   - Écoute et description de trajectoires sonores réelles en utilisant, par exemple : une balle de ping-pong, des clochettes, etc.
   - Démonstration de la synthèse binaurale : écoute de différents sons se déplacant autour du sujet et description (par l'animateur) de la trajectoire.
4. *Production d'idées de métaphores sonores :* cette partie consiste à élaborer des messages sonores 3D sur les grandes fonctions de guidage couramment rencontrées ou pouvant poser problème aux non-voyants. Ces grandes fonctions peuvent être : la vue d'ensemble d'un déplacement, la localisation à un point précis, l'orientation, la description d'une intersection, d'une irrégularité dans le parcours ou encore la description de l'environnement. Cette partie est commune à tous les participants qui vont chacun apporter leurs idées et rebondir sur les idées des autres.

A.2.2 Résultats principaux

Nous avons rencontré de grandes difficultés à faire sortir les utilisateurs de ce qu'ils connaissent déjà et à leur faire comprendre que le concept de "guidage idéal" est placé dans l'imaginaire et

non dans ce qu'ils pensent possible de réaliser au niveau technologique ; quelques idées ont pu être extraites de ce brainstorming. Elles sont résumées et classées ci-dessous en fonction de leur place dans le dispositif : les idées sur le dispositif en général, sur la préparation d'itinéraire et celles sur le déplacement.

a) Réflexions sur le dispositif en général

Plusieurs niveaux de détails doivent pouvoir être fournis par le système en fonction du degré d'expertise en locomotion de l'utilisateur (novice, médium ou expert). Le système doit permettre de facilement faire répéter la dernière information et doit comprendre une fonction "où suis-je".

Le dispositif ne doit pas émettre trop de sons, ni trop d'informations pour ne pas saturer l'utilisateur. Les sons du système doivent être facilement différenciables des sons de l'environnement. Certains participants souhaitent avoir des sons électroniques (type jeux vidéo), d'autres des sons réels hors contexte de l'environnement (comme des sons marins par exemple).

b) La préparation d'itinéraire

La description de l'itinéraire doit pouvoir être réalisée à la manière d'une *carte* (vue du ciel, point de vue allocentré) ou d'un *trajet* (description d'itinéraire donné par un instructeur en locomotion, point de vue égocentré). Les utilisateurs souhaitent pouvoir naviguer virtuellement dans l'itinéraire de la même façon que dans un jeu vidéo. La simulation du trajet pourrait alors leur fournir plusieurs types d'informations comme la trajectoire à suivre, la description des points de repères et d'intérêts présents le long du trajet et ainsi favoriser la tâche de mémorisation de l'itinéraire.

Avec la description du type *trajet*, les utilisateurs souhaitent avoir le choix entre suivre un coach virtuel (par exemple des bruits de pas situés quelques mètres devant eux) ou se déplacer volontairement dans le trajet virtuel. Certains participants ont aussi évoqué l'avantage de pouvoir émettre des sons pour analyser l'espace dans lequel ils se trouvent (bien que très intéressante cette idée est aujourd'hui compliquée à réaliser à l'échelle d'une ville ou même d'une rue et ne sera donc pas développée par la suite).

c) Le déplacement

Trois modes de guidages ont été définis par les participants pendant la séance :

- *le mode Petit Poucet* qui consiste à se déplacer vers des balises sonores virtuelles dont l'intensité augmente avec la diminution de la distance. Lorsque l'utilisateur atteint une balise, celle-ci s'éteint et la suivante s'allume.
- *le mode Coach* qui consiste à suivre un guide virtuel situé quelques mètres devant l'utilisateur. Comme pour la préparation d'itinéraire, suivant le niveau de détail choisi par l'utilisateur, le coach peut émettre un son en continu ou par intermittence (type bruit de pas, sifflements) ou bien encore n'émettre des sons que lorsque l'utilisateur s'écarte du chemin.

– *le mode Touriste* qui doit permettre à l'utilisateur de découvrir de nouveaux endroits. Le déplacement n'est plus forcement utilitaire mais a pour but par exemple d'explorer un nouveau quartier ou une nouvelle ville. Dans ce mode, l'utilisateur ne souhaite pas arriver à un point précis (le guidage sonore n'est donc pas nécessaire) ; il souhaite seulement connaître ce qu'il y a autour de lui. La description de l'environnement pourrait, à la demande, être réalisée à 360° en faisant un scan de ce qui entoure l'utilisateur et en le décrivant avec de la synthèse vocale spatialisée.

A.3 Bilan des sessions de conception participative

Nous allons ici présenter une synthèse des informations que nous avons pu extraire des différentes sessions réalisées avec les non-voyants (les réunions décrites dans les sections A.1 et A.2 mais aussi l'analyse d'activité réalisée par [Brunet, 2010] ainsi que les entretiens informels qui ont eu lieu dans le cadre ou en dehors du projet).

Le constat général et plutôt évident est qu'il n'y a pas une seule et unique façon de penser le dispositif. Tous les non-voyants sont différents. Certains souhaitent avoir beaucoup d'informations, quand d'autres n'en veulent que le minimum nécessaire ; certains prennent le temps de préparer leur itinéraire à la maison alors que d'autres non ; certains souhaitent que le système utilise des sons de jeux vidéos alors que d'autres préfèreraient des sons musicaux. Le système NAVIG devra donc être paramétrable par l'utilisateur et pouvoir contenir plusieurs configurations qui dépendront du type de déplacement que souhaite réaliser l'utilisateur.

A.3.1 Les informations à transmettre

a) Avant la navigation

Pour la préparation de l'itinéraire (à la maison), il faut que le système puisse fournir un maximum d'informations afin que le non-voyant puisse se faire une bonne représentation spatiale de l'environnement dans lequel il va se déplacer. Ces informations pourront être présentées de façon allocentrée (carte sonore vu du dessus) ou de façon égocentrée (l'utilisateur peut se déplacer virtuellement dans le trajet). Dans les deux cas, les informations devront être présentées de façon séquentielle. Seule la position spatiale des informations sonores dépendra du type de représentation.

b) Pendant la navigation

L'information doit permettre d'aider l'utilisateur de la façon la plus efficace tout en étant la moins intrusive possible. La signalisation des erreurs de trajectoire doit être rapide et efficace.

Les informations nécessaires au déplacement de l'utilisateur sur un trajet, sont :
– Noms des rues empruntées

- Noms des rues croisées
- Sur un segment : distance à parcourir, nombre de rues à croiser avant de changer de direction, direction à emprunter une fois la distance parcourue

Pendant la navigation, en plus des informations sur la trajectoire à suivre, le système doit pouvoir répondre à deux requêtes :

- Où suis-je ?
  - Nom de la rue
  - Numéro de la rue
  - Sens dans lequel il faut marcher (pour aller vers la destination)
  - Prochaine intersection
  - Objets proches (points de repères, d'intérêts et rues adjacentes)
- Description d'un croisement
  - Type de croisement (T, X, complexe, rond point ; et nombre de voies)
  - Noms des rues qui partent de ce croisement
  - Orientation des rues (angle par rapport à la rue où est l'utilisateur)

En plus de ces informations, le système doit pouvoir, en option, signaler des points de repère (PR), des points d'intérêt (POI) ainsi que des points favoris (PF). Ces informations doivent être signalées lorsque l'utilisateur est à une distance inférieure à une distance prédéfinie qui reste à déterminer. Ces points sont divisés en sous catégories, permettant ainsi à l'utilisateur de ne connaître la présence que des sous-catégories qui l'intéressent. Pour chacune des catégories de point (PR, POI ou PF), le système peut :

- Ne rien signaler
- Ne signaler que certaines sous-catégories
- Tout signaler

Au moins deux modes de déplacement doivent être définis : le mode *touriste* et le mode *déplacement utile*. Ces modes permettent à l'utilisateur d'enregistrer plusieurs profils au niveau de la quantité d'informations que doit transmettre le système.

A.3.2 Comment transmettre les informations

Du point de vue des utilisateurs, les informations sonores doivent être les moins nombreuses possibles. En général, ils ne sont pas favorables à l'utilisation du son et préfèreraient souvent que les informations soient transmises par de la synthèse vocale (ce point de vue a été modéré après l'écoute de sons 3D virtuels). Dans tous les cas, ils souhaitent que le niveau de verbosité du système puisse être configurable et que les sons utilisés soient facilement différentiables du paysage sonore urbain. Au niveau du type de son à utiliser, aucun consensus n'a pu être établi. Si certains participants souhaitent que les sons du système soient électroniques et qu'ils ressemblent à des sons de jeux vidéo, d'autres préfèreraient des sons musicaux ou des sons naturels (hors contexte).

Annexe B

Les différents éléments du système NAVIG

Les différents objectifs du projet NAVIG seront atteints en combinant en entrée du système des informations provenant de la géolocalisation satellite et d'un système de reconnaissance d'image très rapide. Les informations de guidage seront fournies via un casque stéréophonique sous forme de sons 3D pouvant combiner informations vocales (avec du TTS) et sonification.

Le prototype du système est composé de plusieurs éléments (appelés agents) structurés autour d'un "framework" multi-agent utilisant un protocole de communication basé sur le "middleware" IVY [Buisson *et al.*, 2002]. Avec ce protocole, les agents peuvent se connecter et se déconnecter dynamiquement du bus IVY et ne reçoivent que les messages auxquels ils sont abonnés. L'architecture générale du système est représentée figure B.1. Les principaux éléments du système peuvent être divisés en trois catégories : les données en entrée, l'interface utilisateur et le système de contrôle interne. Les données en entrée sont composées d'un système de géopositionnement par satellite (GPS), de capteurs d'orientation et d'accélération, d'un système d'information géographique (GIS), d'un module de traitement d'image combiné à des caméras stéréoscopiques montées sur la tête et d'un module de fusion des données. L'interface avec l'utilisateur se fait par reconnaissance vocale en entrée et par sonification 3D en sortie. Le système de contrôle interne appelé contrôleur de dialogue permet de combiner les informations en entrée avec les préférences de l'utilisateur et d'envoyer les instructions du système au module de sonification. Dans la section qui vient, nous allons présenter les différentes briques essentielles de ce système.

B.1 Vision artificielle

Le module de vision artificielle du système NAVIG est conçu pour extraire les informations pertinentes de l'environnement de façon à compenser le déficit visuel de l'utilisateur. Pour ce faire, des cameras stéréoscopiques positionnées sur la tête de l'utilisateur permettent de capter des

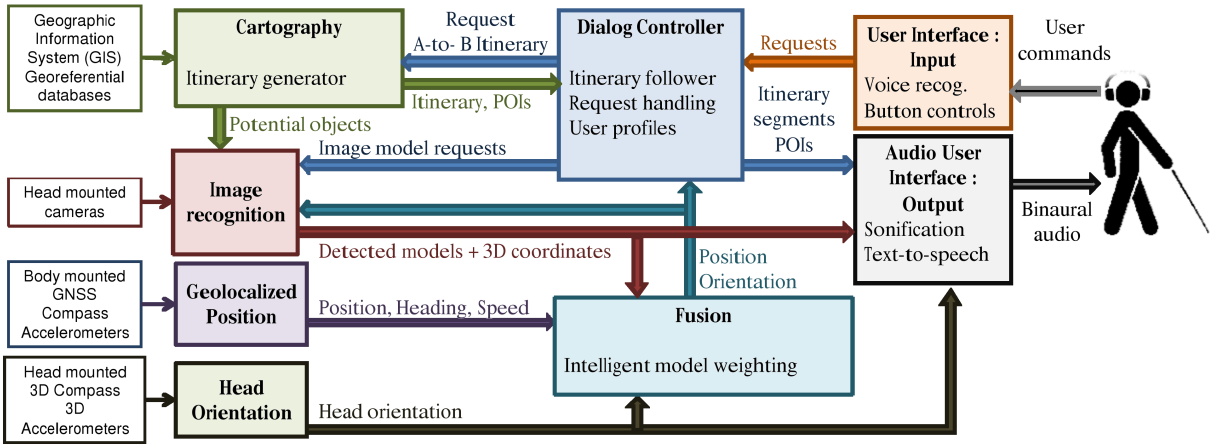


FIGURE B.1 – Vue d'ensemble de l'architecture du système NAVIG.

images dont est extraite la position d'un objet d'intérêt grâce à un algorithme en temps réel de vision artificielle. Les objets d'intérêts peuvent être des objets que l'utilisateur cherche à localiser ou des points de repères géolocalisés demandés par le système pour améliorer le positionnement de l'utilisateur. Dans les deux cas, la fonction permettant la recherche du modèle correspondant à l'objet à trouver s'appuie sur une librairie de traitement d'images inspirée par la neurophysiologie de la vision humaine (SpikeNet).

L'algorithme de reconnaissance SpikeNet se fonde sur la recherche en neurophysiologie de la vision et plus spécifiquement sur les mécanismes humains impliqués dans l'analyse extrêmement rapide des scènes visuelles [Thorpe *et al.*, 1996]. La modélisation informatique de ces mécanismes a mené au développement d'un moteur de reconnaissance permettant un traitement très rapide. Cette technologie met en oeuvre un nouveau procédé de traitement d'images basé sur la simulation de larges réseaux de neurones impulsionnelles et asynchrones. Le système d'analyse d'images SpikeNet permet de localiser et de reconnaître en temps réel des objets présentés à l'image, quels que soient leur position, leur nombre, leur taille ou leur orientation. Contrairement aux systèmes de vision artificielle conventionnels, l'algorithme de détection SpikeNet est particulièrement robuste aux variations de contraste ou de luminosité. Il est capable de fonctionner dans des environnements complexes et peut gérer des images fortement bruitées.

Lorsque l'utilisateur cherche un objet (exemple figure B.2 (à gauche)), le système charge les modèles correspondant à cet objet et les recherches dans le flux d'images provenant d'une des deux caméra. Une fois la cible détectée, sa position par rapport à l'utilisateur est calculée en utilisant des méthodes de stéréovision. L'objet est ensuite présenté à l'utilisateur en utilisant des sons 3D virtuels (section 2.2) dont la position est rafraichie toutes les 60 ms par le module de vision et interpolée entre temps grâce aux informations fournies par le capteur d'orientation positionné sur le casque. Cette fonction joue un rôle très important dans le système NAVIG car elle permet de restaurer un bouclage "visuomoteur" fonctionnel permettant à l'utilisateur de bouger son bras vers une cible d'intérêt [Dramas *et al.*, 2010].

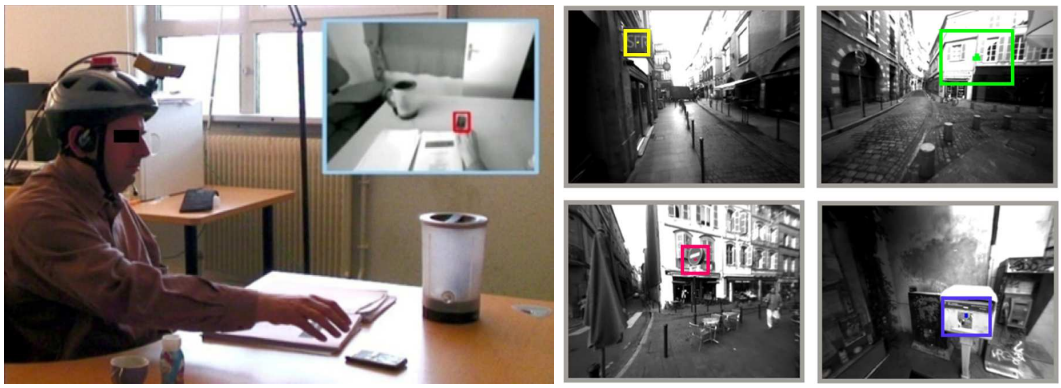


FIGURE B.2 – Gauche : Détection d'objet en champ proche ; Droite : exemples de points de repères visuels géolocalisés utilisés pour le positionnement de l'utilisateur (enseigne de magasin, façade, panneau signalisation, boîte aux lettre).

La deuxième fonction du module de vision, détaillée dans [Brilhault *et al.*, 2011], est de fournir au système des informations supplémentaires pour le positionnement précis de l'utilisateur. Lorsque l'utilisateur est guidé vers une destination choisie, le système tente de détecter des cibles visuelles géolocalisées le long de l'itinéraire. Ces cibles, si elles sont détectées, sont utilisées pour affiner la position courante du GPS. Ces points de repères visuels peuvent être des paneaux de signalisations, des statues, des boîtes aux lettres où même des façades présentant certaines particularités qui les distinguent des autres (voir figure B.2 (droite)). Ils sont stockés avec leurs coordonnées géographiques dans le SIG. Lorsqu'une cible visuelle est détectée, la position de l'utilisateur peut être estimée en utilisant la distance et l'angle de la cible par rapport à la caméra ainsi que les données des autres capteurs (accéléromètres et capteur d'orientation). Cette méthode permet de donner une estimation de la position uniquement basée sur la vision lorsque le signal GPS est perdu ou peut être intégré dans un schéma de fusion plus large (décrit dans la section B.2).

B.2 Système de géolocalisation avec précision piéton

La partie permettant la géolocalisation de l'utilisateur se base sur les informations provenant de plusieurs modules : le module GPS (augmenté de plusieurs capteurs), le module d'orientation de la tête, le module de reconnaissance d'image et le module de fusion.

L'utilisation des GPS traditionnels dans une application spécifique aux piétons est limité par la précision de ce dernier. En effet, avec une précision en général de l'ordre de 20 mètres et des données souvent peu fiables en environnement urbain (avec de grands immeubles), un GPS est difficilement utilisable dans un système d'aide à la navigation pour les non-voyants. S'il existe plusieurs algorithmes pour prédire la position des véhicules à moteur (basés sur la position et la vitesse du véhicule) et ainsi affiner la précision de leur positionnement, il n'est en revanche pas possible de les appliquer au piéton étant donné le caractère imprédictible des mouvements de celui-ci.

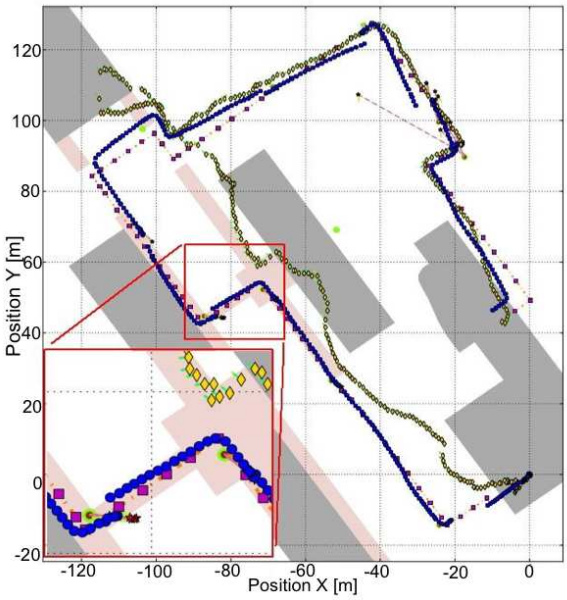


FIGURE B.3 – Trajet test réalisé sur le campus de l'Université de Toulouse. Les bâtiments sont indiqués par les polygones gris et les arcades par des polygones roses. Plusieurs trajets sont montrés : le trajet réalisé (violet), le positionnement par un GPS commercial (jaune), la position estimée par le module de fusion (rouge) et la position calculée par le module de fusion (bleu).

Un des principaux objectifs du projet NAVIG, était donc de mettre en place un système de géopositionnement piéton avec une précision inférieure à 5 mètres 95% du temps. Pour ce faire, l'idée adoptée a été de combiner un capteur GPS traditionnel avec des capteurs inertiels pour obtenir un premier positionnement avec une précision de l'ordre de 10 mètres (partie devant être réalisée par NAVOCAP), puis d'affiner cette information avec la position de points de repères détectés par le module de vision le long du trajet. La position du piéton est fournie au contrôleur de dialogue par le module de fusion qui a pour rôle de combiner les informations provenant du GPS (augmenté de capteur) avec celles du système de vision. Les modèles des cibles à détecter sont fournis au module de vision au cours du trajet par le Système d'Information Géographique (SIG). La figure B.3 représente un trajet test effectué sur le campus de l'Université de Toulouse. On remarque que les erreurs de positionnement très grandes (parfois supérieures à 20 m) du GPS (en jaune) sont pratiquement complètement corrigées par le module de fusion (en bleu). Pour plus d'informations sur l'algorithme de fusion, le lecteur pourra se reporter à [Brilhault *et al.*, 2011].

B.3 Système d'Information Géographique piéton

Le Système d'Information Géographique (SIG) a été défini par [Burrough, 1986] comme un outil permettant de capturer, manipuler, afficher, demander et analyser des données géographiques. Le SIG est un des éléments essentiel du système d'aide à la navigation pour les non-voyants [Golledge *et al.*, 1998]. S'appuyant sur une base de données numérisée du territoire et sur des outils d'analyse spécifiques, le SIG doit fournir à l'utilisateur les informations environnementales

précises pour assurer le succès de la tâche de navigation.

B.3.1 Le SIG NAVIG

Afin de permettre à l'utilisateur de se construire une carte cognitive de l'environnement dans lequel il va se déplacer, il est important d'adapter les SIG ordinaire aux informations spécifiques pour les non-voyants [Jacobson et Kitchin, 1997]. Ces informations pourront ensuite être fournies à l'utilisateur pendant la préparation de l'itinéraire ainsi que pendant la navigation afin de favoriser au maximum la construction d'une représentation mentale de l'environnement.

Dans le contexte d'un système d'aide à l'orientation pour les non-voyants, [Gaunet et Briffault, 2005] ont montré qu'un SIG adapté à la navigation pédestre devait inclure les rue, les trottoirs, les passages piétons ainsi que les intersections. Ces informations doivent donc être collectées et stockées dans le SIG avec un degré élevé de précision afin d'être prises en compte dans la procédure de planification d'itinéraire (voir section B.3.2) et fournies lors du déplacement.

Actuellement les SIG commerciaux ont été exclusivement développés dans un but de navigation véhiculé. De plus, nous avons vu dans les sections A.3 et B.2 que le SIG du système doit contenir des informations supplémentaires qui ne sont pas simple à répertorier (ex : position d'un changement de revêtement, enseigne publicitaire, ...). Plusieurs solutions ont été envisagées pour constituer automatiquement le SIG du système (récupération du SIG mis en place par le service de la voirie de la mairie de Toulouse, extraction de cibles visuelles dans des bases telles que Google Street View, ...), nous ne détaillerons pas les différentes solutions envisagées dans ce document et nous contenterons de détailler le SIG idéal qui a été imaginé de notre côté pour satisfaire au mieux les besoins des non-voyants.

Les séries de brainstorming et d'interview réalisées avec des utilisateurs potentiels du système ainsi qu'avec des instructeurs en orientation et en mobilité ont permis de définir cinq grandes classes d'objets qui doivent être incluses dans le SIG :

- Les zones piétonnes : Tous les chemins pédestres possibles tels que les trottoirs, les passages piétons, etc. (pour plus d'informations, voir [Zheng *et al.*, 2009]).
- Les points difficiles : Lieux pouvant être potentiellement dangereux ou compliqués à comprendre pour les non-voyants (voir section 5.2.1).
- Les points de repères : Lieux ou objets pouvant être détecter par l'utilisateur sans utiliser le système. Ces points doivent permettre à l'utilisateur de confirmer sa propre position dans le trajet. Ils peuvent être tactiles (changement de revêtement du sol : pavés/goudron), vestibulaires (un escalier), sonores (fontaine) ou olfactifs (boulangerie).
- Les points d'intérêts : Lieux présentant un intérêt potentiel pour l'utilisateur. Ils peuvent être utilisés comme destination finale ou juste être signalés à l'utilisateur pendant son déplacement afin de lui permettre une meilleure compréhension de l'environnement (e.g., bâtiments publiques, magasins, ...).

– Les points visuels : Ils correspondent aux cibles géolocalisées qui doivent être détectées par le système de vision artificielle pour améliorer le géopositionnement.

Chaque objet indexé dans la base peut être rattaché à plusieurs catégories. Un arrêt de bus, par exemple, peut être considéré comme un point de repère tactile (étant donné que le non-voyant peut le détecter avec sa canne), il peut aussi constituer un point d'intérêt (étant donné que c'est une destination potentielle) et peut être un point visuel (si le module de vision artificielle est capable de le détecter). L'utilisateur a aussi la possibilité d'ajouter des positions spécifiques (comme sa maison, son travail ou même une zone particulièrement dangereuse les jours de pluie) qui sont intégrées dans une classe du SIG appelée point favori. Chacun de ses points peut être enregistré dans la base avec une étiquette définie par l'utilisateur.

B.3.2 Planification d'itinéraire

La planification de l'itinéraire est une des composantes essentielle du dispositif d'aide à la navigation. Partie intégrante du SIG, elle intervient lorsque la position de l'utilisateur et sa destination ont été déterminées.

Traditionnellement, la planification d'itinéraire consiste à minimiser le chemin entre deux points tout en prenant en compte certaines contraintes. Pour un trajet en voiture, par exemple, la principale contrainte est le sens de la route (la voiture ne peut pas emprunter de sens interdit). Pour les non-voyants, nous avons vu dans l'annexe A que les contraintes sont nombreuses et qu'elles peuvent dépendre de l'utilisateur. En générale, les non-voyants peuvent préférer une route plus longue si elle leur permet d'éviter certaines difficultés ou obstacles mais il arrive que dans certaines situations, ils préfèrent prendre le chemin le plus rapide même s'il les entraîne dans des situations plus dangereuses. Le choix de l'optimisation du trajet varie donc en fonction de l'utilisateur (suivant son expérience en mobilité) mais aussi en fonction du type de déplacement (milieu connu ou inconnu, balade ou rendez-vous, ...). Une première version de l'algorithme de planification de l'itinéraire est détaillée dans [Kammoun *et al.*, 2010]. Par la suite, les sessions de conception participative ont permis de mettre en évidence de nombreux facteurs à prendre en compte dans le calcul d'itinéraire pour les non-voyants. Une partie est présentée dans l'annexe A avec une première idée des poids à leur attribuer pour le calcul de l'itinéraire. Dans un dispositif "idéal", l'utilisateur devrait pouvoir paramétrer le poids de chacun des facteurs à prendre en compte en fonction de sa propre expérience de déplacement. Il devrait aussi avoir la possibilité de définir plusieurs jeux de paramètres en fonction du type de déplacement qu'il souhaite réaliser (déplacement avec un objectif particulier ou pour visiter).

L'itinéraire, une fois calculé, est envoyé au contrôleur de dialogue sous la forme d'une succession de points d'itinéraire, de points difficiles, ainsi que des points de repères, d'intérêts, visuels et favoris présents à proximité du parcours.

B.4 Contrôleur de dialogue

Placé au coeur du système, le contrôleur de dialogue est en quelque sorte la partie “intelligente” du système. Ce module qui communique avec la quasi totalité des modules doit prendre les décisions générales, gérer les requêtes en entrée et les messages à envoyer en sortie. De plus, il a pour charge de stocker les différents niveaux de préférences des utilisateurs (pour la planification de l'itinéraire, le choix des points d'intérêts et de repères à présenter ainsi que la verbosité et le type de sons à utiliser en sortie du système) et de les transmettre en fonction de la situation aux différents modules concernés.

En navigation, le contrôleur de dialogue a pour rôle d'assurer le suivi d'itinéraire et de fournir à l'interface de sortie les instructions à transmettre à l'utilisateur. Le suivie d'itinéraire est réalisé à partir des données transmises par le SIG et la géolocalisation. Si l'itinéraire est respecté, le contrôleur de dialogue envoie les intructions de guidage au fur et à mesure de l'avancement. S'il n'est pas respecté, le contrôleur de dialogue doit prendre la décision de déclencher un nouveau calcul d'itinéraire.

B.5 Interaction homme-machine

L'interaction homme-machine est divisée en deux module : l'interface en entrée (permettre aux utilisateurs de donner des consignes au système) et l'interface en sortie (transmettre des messages à l'utilisateur). Pour ces deux parties, il est important de prendre en compte l'avis des utilisateurs et de se baser sur leur retour d'expérience afin de concevoir un système ergonomique.

B.5.1 Interface en entrée

L'interface en entrée a fait l'objet de plusieurs stages qui se sont déroulés à l'IRIT. Elle a pour rôle de capturer les consignes émanant de l'utilisateur. Ces consignes pouvant être la saisie d'une adresse vers laquelle aller, un objet à trouver, la saisie des réglages utilisateurs, etc ...

La principale modalité utilisée pour l'interaction en entrée est la reconnaissance vocale (via un microphone et le logiciel Dragon Naturally Speaking [1]. Un boitier de commande (figure B.4) a été ensuite ajouté au dispositif pour gérer des interactions simples telles que la gestion du volume des informations sonores, le démarrage, la mise en pause ou l'arrêt de la navigation.

B.5.2 Interface en sortie

L'interface en sortie consiste à émettre les messages vers l'utilisateur. Elle utilise la modalité auditive et est constituée de messages vocaux (synthèse vocale) et de messages sonores (sonification 3D) présentés sur un casque stéréophonique à conduction osseuse (afin de ne pas masquer les

1. url : `http://www.nuance.com/dragon/index.htm`

FIGURE B.4 – Photo du boitier utilisé pour gérer les interactions basiques du système NAVIG.

oreilles). Elle a fait l'objet de cette thèse. L'étude de l'adaptation de la synthèse binaurale (utilisée pour la génération de son 3D) à un contexte commercial est détaillée dans le chapitre 3. La sonification d'objet dans l'espace péripersonnel est abordée dans le chapitre 4 et enfin, la mise en place de palettes d'indicateurs sonores permettant d'aider l'utilisateur à se déplacer en champs lointain est traitée dans le chapitre 5.

Bibliographie

[Absar et Guastavino, 2008] Absar, R. et Guastavino, C. (2008). Usability of non-speech sounds in user interfaces. *In Proceedings of the 14th International Conference on Auditory Display (ICAD2008)*, Paris, France.

[Adams et Beaton, 2000] Adams, C. J. et Beaton, R. J. (2000). An investigation of navigation processes in human locomotor behavior. *In Proceedings of the Human Factors and Ergonomics Society Annual Meeting.*

[Alba *et al.*, 2008] Alba, A., Zubieta, C. et Arce-Santana, E. (2008). Binaural sonification of disparity maps. *In Image Processing Workshop (PI08).*

[Algazi *et al.*, 2001a] Algazi, V. R., Avendano, C. et Duda, R. O. (2001a). Elevation localization and head-related transfer function analysis at low frequencies. *J. Acoust. Soc. Am.*, 109(3):1110–1122.

[Algazi *et al.*, 2001b] Algazi, V. R., Duda, R. O., Thompson, D. M. et Avendano, C. (2001b). The cipic HRTF database. *In IEEE WASPAA01*, pages 99–102.

[Allen et Berkley, 1979] Allen, J. et Berkley, D. (1979). Image method for efficiently simulating small-room acoustics. *Acoustical Society of America Journal*, 65:943–950.

[Amemiya *et al.*, 2004] Amemiya, T., Yamashita, J., Hirota, K. et Hirose, M. (2004). Virtual leading blocks for the deaf-blind : A real-time way- nder by verbal-nonverbal hybrid interface and high-density r d tag space. *In VR '04 : Proceedings of the IEEE Virtual Reality 2004, IEEE Computer Society, p. 165.*

[Aussal *et al.*, 2012] Aussal, M., Alouges, F. et Katz, B. (2012). Hrtf interpolation and itd personalization for binaural synthesis using spherical harmonics. *In 4th International Symposium on Ambisonics and Spherical Acoustics.*

[Auvray, 2004] Auvray, M. (2004). *Immersion et perception spatiale. L'exemple des dispositifs de substitution sensorielle.* Thèse de doctorat, EHESS.

[Auvray *et al.*, 2007] Auvray, M., Hanneton, S. et O'Regan, J. K. (2007). Learning to perceive with a visuo-auditory substitution system : localisation and object recognition with 'the vOICe'. *Perception*, 36(3):416–430.

[Aytekin *et al.*, 2008] Aytekin, M., Moss, C. et Simon, J. (2008). A sensorimotor approach to sound localization. *Neural Computation*, 20(3):603–635.

[Aznar, 2007] Aznar, G. (2007). *100 techniques de créativité pour les produire et les gérer*. Editions d'Organisation.

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

[Bach-y Rita *et al.*, 1969] Bach-y Rita, P., Collins, C. C., Saunders, F. A., White, B. et Scadden, L. (1969). Vision substitution by tactile image projection. *Transactions of the Pacific Coast OtoOphthalmological Society annual meeting*, 221(5184):83–91.

[Bach-y Rita *et al.*, 1998] Bach-y Rita, P., Kaczmarek, K. A., Tyler, M. E. et Garcia-Lara, J. (1998). Form perception with a 49- point electrotactile stimulus array on the tongue : a technical note. *Journal of Rehabilitation Research & Development*, 35(4):427.

[Baier *et al.*, 2007] Baier, G., Hermann, T. et Stephani, U. (2007). Event-based sonification of eeg rhythms in real time. *Clinical Neurophysiology*, 118(6):1377–1386.

[Balakrishnan *et al.*, 2004] Balakrishnan, G., Sainarayanan, G., Nagarajan, R. et Yaacob, S. (2004). Object discrimination using stereo vision for blind through stereo sonification. *In ICVGIP*, pages 234–239.

[Ballora *et al.*, 2000] Ballora, M., Pennycook, B. W. et Glass, L. (2000). Audification of heart rhythms in csound. *In* Belanger, R., éditeur : *The Csound book : perspectives in software synthesis, sound design, signal processing, and programming*. MIT Press.

[Barrass *et al.*, 2010] Barrass, S., Schaffert, N. et Barrass, T. (2010). Probing preferences between six interactive sonifications designed for recreational sports, health and fitness. *In Proc the 3rd Interactive Sonification Workshop (ISon'2010)*, Stockholm, Sweden. KTH.

[Batteau, 1967] Batteau, D. (1967). The role of the pinna in human localization. *In Proc. Roy. Soc*, volume B168, pages 158–180.

[Batteau, 1968] Batteau, D. (1968). Listening with naked ear. *In The neuropsychology of spatially oriented behaviour*, pages 109–133. Dorsy Press.

[Begault, 1992] Begault, D. (1992). Perceptual effects of synthetic reverberation on three-dimensional audio systems. *J. Audio Eng. Soc*, 40(11):895–904.

[Begault, 1994] Begault, D. (1994). *3-D Sound for Virtual Reality and Multimedia*. Academic Press, Cambridge.

[Begault et Wenzel, 1993] Begault, D. et Wenzel, E. (1993). Headphone localization of speech. *Human Factors*, 35(2):361–376.

[Bensaoula *et al.*, 2006] Bensaoula, S., Boulebtateche, B. et Bedda, M. (2006). Electronic device for blind mobility aid. *J of Engineering and Applied Sciences*, 1(4):514–522.

[Berkhout *et al.*, 1993] Berkhout, A. J., de Vries, D. et Vogel, P. (1993). Acoustic control by wave field synthesis. *J. Acoust. Soc. Am.*, 93(5):2764–2778.

[Blattner *et al.*, 1989] Blattner, M., Sumikawa, D. et Greenberg, R. (1989). Earcons and icons : Their structure and common design principles (abstract only). *SIGCHI Bull.*, 21:123–124.

[Blauert, 1997] Blauert, J. (1997). *Spatial Hearing, The Psychophysics of Human Sound Localization.* MIT Press, Cambridge.

[Blenkhorn et Evans, 1997] Blenkhorn, P. et Evans, D. (1997). A system for enabling blind people to identify landmarks - the sound buoy. *In IEEE Transactions on Rehabilitation Engineering*, volume 5, pages 276–278.

[Blum *et al.*, 2004] Blum, A., Katz, B. F. et Warusfel, O. (2004). Eliciting adaptation to non-individual HRTF spectral cues with multi-modal training. *In Proc. CFA/DAGA.*

[Borenstein et Ulrich, 1997] Borenstein, J. et Ulrich, I. (1997). The guidecane - a computerized travel aid for the active guidance of blind pedestrians. *In IEEE International Conference on Robotics and Automation.*

[Bovermann *et al.*, 2005] Bovermann, T., Hermann, T. et Ritter, H. (2005). The local heat exploration model for interactive sonification. *In Proceedings of the International Conference on Auditory Display (ICAD2005)*, pages 85–91, Limerick, Ireland.

[Bregman, 1994] Bregman, A. S. (1994). *Auditory Scene Analysis : The Perceptual Organization of Sound.* Cambridge, MA : MIT Press.

[Brewster, 1997] Brewster, S. (1997). Using non-speech sound to overcome information overload. *Display*, 17:179–189.

[Brewster et Brown, 2004] Brewster, S. et Brown, L. (2004). Tactons : structured tactile messages for non-visual information display. *In Proceedings of the fifth conference on Australasian user interface*, volume 28 de *AUIC '04*, pages 15–23, Darlinghurst, Australia, Australia. Australian Computer Society, Inc.

[Brewster *et al.*, 1993] Brewster, S., Wright, P. et Edwards, A. (1993). An evaluation of earcons for use in auditory human-computer interfaces. *In Proceedings of the INTERACT '93 and CHI '93 conference on Human factors in computing systems*, CHI '93, pages 222–227, New York, NY, USA. ACM.

[Brewster et Murray, 2000] Brewster, S. A. et Murray, R. (2000). Presenting dynamic information on mobile computers. *Personal Ubiquitous Comput.*, 4(4):209–212.

[Brilhault *et al.*, 2011] Brilhault, A., Kammoun, S., Gutierrez, O., Truillet, P. et Jouffrais, C. (2011). Fusion of artificial vision and gps to improve blind pedestrian positioning. *In Intl. Conf. on New Technologies, Mobility and Security, IEEE, France.*

[Brock *et al.*, 2010] Brock, A., Vinot, J.-L., Oriola, B., Kammoun, S., Truillet, P. et Jouffrais, C. (2010). Méthodes et outils de conception participative avec des utilisateurs non-voyants. *In Conference Internationale Francophone sur I'Interaction Homme-Machine*, IHM '10, pages 65–72. ACM.

[Bronkhorst, 2002] Bronkhorst, A. (2002). Modeling auditory distance perception in rooms. *Human Factors*, 397(6719):517–520.

[Bronkhorst, 1995] Bronkhorst, A. W. (1995). Localization of real and virtual sound sources. *Acoustical Society of America Journal*, 98:2542–2553.

[Brown *et al.*, 2003] Brown, L., Brewster, S., Ramloll, S., Burton, R., et Riedel, B. (2003). Design guidelines for audio presentation of graphs and tables. *In Proceedings of the 9th International Conference on Auditory Display.*

[Brown *et al.*, 1989] Brown, M. L., Newsome, S. L. et Glinert, E. P. (1989). An experiment into the use of auditory cues to reduce visual workload. *SIGCHI Bull.*, 20:339–346.

[Brunet, 2010] Brunet, L. (2010). Étude des besoins et des stratégies des personnes non-voyantes lors de la navigation pour la conception d'un dispositif d'aide performant et accepté (Needs and strategy study of blind people during navigation for the design of a functional and accepted aid device). Mémoire de D.E.A., Departement of Ergonomics, Universitée Paris-Sud, Orsay, France.

[Brunet *et al.*, 2012] Brunet, L., Gallay, M., Darses, F. et Auvray, M. (2012). Strategies and needs of blind pedestrians during urban navigation. *J Applied Ergonomics*, Submitted.

[Brungart, 1993] Brungart, D. (1993). Distance simulation in virtual audio displays. *In Aerospace and Electronics Conference, 1993. NAECON 1993., Proceedings of the IEEE 1993 National*, pages 612–617. IEEE.

[Brungart *et al.*, 1999] Brungart, D., Durlach, N. et Rabinowitz, W. (1999). Auditory localization of nearby sources. II. localization of a broadband source. *J. Acoust. Soc. Am.*, 106(4):1956–1968.

[Brungart et Rabinowitz, 1999] Brungart, D. et Rabinowitz, W. (1999). Auditory localization of nearby sources. head-related transfer functions. *J. Acoust. Soc. Am.*, 106(3):1465–1479.

[Brungart *et al.*, 2000] Brungart, D., Rabinowitz, W. et Durlach, N. (2000). Evaluation of response methods for the localization of nearby objects. *Attention, Perception, & Psychophysics*, 62(1):48–65.

[Brungart *et al.*, 2004] Brungart, D., Simpson, B., McKinley, R., Kordik, A., Dallman, R. et Ovenshire, D. (2004). The interaction between head-tracker latency, sound duration, and response time in the localization of virtual sound sources. *In Proceedings of the 10th international conference on Auditory display (ICAD 2004).*

[Brungart, 1999] Brungart, D. S. (1999). Auditory parallax effects in the hrtf for nearby sources. *In Proceedings of 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics.*

[Brungart et Simpson, 2008] Brungart, D. S. et Simpson, B. (2008). Design, validation, and in-flight evaluation of an auditory attitude indicator based on pilot-selected music. *In Proceedings of the International Conference on Auditory Display (ICAD2008)*, Paris, France.

[Brungart et Simpson, 2009] Brungart, D. S. et Simpson, B. D. (2009). Effects of bandwidth on auditory localization with a noise masker. *Acoustical Society of America Journal*, 126(6):3199.

[Buisson *et al.*, 2002] Buisson, M., Bustico, A., Chatty, S., Colin, F.-R., Jestin, Y., Maury, S., Mertz, C. et Truillet, P. (2002). Ivy : un bus logiciel au service du développement de prototypes de systèmes interactifs. *In 14th French-speaking conference on Human-computer interaction (IHM '02).*

[Bujacz, 2010] Bujacz, M. (2010). *Representing 3D scenes through spatial audio in an electronic travel for the blind.* Thèse de doctorat, Technical University of Lodz. Faculty of Electrical, Electronic, Computer and Control Engineering.

[Bujacz *et al.*, 2011] Bujacz, M., Pec, M., Skulimowski, P., Strumillo, P. et Materka, A. (2011). Sonification of 3d scenes in an electronic travel aid for the blind. *Advances in Sound Localization.*

[Bujacz et Strumillo, 2006] Bujacz, M. et Strumillo, P. (2006). Stereophonic representation of virtual 3d scenes - a simulated mobility aid for the blind. *In New Trends in Audio and Video.*

[Burrough, 1986] Burrough, P. (1986). *Principles of Geographical Information Systems for Land Resources, Assessment.* Oxford University Press, USA.

[Busson, 2006] Busson, S. (2006). *Individualisation d'indices acoustiques pour la synthèse binaurale.* Thèse de doctorat, Université de la Méditerranée, Aix-Marseille II.

[Buxton, 1989] Buxton, W. (1989). Introduction to this special issue on nonspeech audio. *Human-computer Interaction*, 4:1–9.

[Carlile *et al.*, 2007] Carlile, S., Blackman, T. et Cooper, J. (2007). The plastic ear : Coping with a life time of change. *In 19th International Congress on Acoustics (ICA).*

[Carlile *et al.*, 2000] Carlile, S., Jin, C. et Van Raad, V. (2000). Continuous virtual auditory space using hrtf interpolation : acoustic and psychophysical errors. *In IEEE PacificRim Conference on Multimedia (2000)*, pages 220–223.

[Carlile *et al.*, 1997] Carlile, S., Leong, P. et Hyams, S. (1997). The nature and distribution of errors in sound localization by human listeners. *Hearing Research*, 114(1-2):179–196.

[Chateau, 1996] Chateau, N. (1996). *Localisation de sources sonores multiples dans l'hémisphère supérieur.* Thèse de doctorat, Université de la Méditerranée Aix Marseille II, laboratoire de mécanique et d'acoustique.

[Chion, 1983] Chion, M. (1983). *Guide To Sound Objects. Pierre Schaeffer and Musical Research (English translation by Dack, J. and North, C.).* Buchet/Chastel.

[Chion, 1994] Chion, M. (1994). *Le son au cinéma.* Edition de l'Etoile / Cahier du cinéma, Collection Essais, Paris.

[Cohen, 1993] Cohen, J. (1993). `kirk here :' : using genre sounds to monitor background activity. *In INTERACT '93 and CHI '93 conference companion on Human factors in computing systems*, pages 63–64.

[Coleman, 1962] Coleman, P. (1962). Failure to localize the source distance of an unfamiliar sound. *J. Acoust. Soc. Am.*, 34:345–346.

[Coleman, 1963] Coleman, P. (1963). An analysis of cues to auditory depth perception in free space. *Psychological Bulletin*, 60(3):302–315.

[Coleman, 1968] Coleman, P. (1968). Dual role of frequency spectrum in determination of auditory distance. *The Journal of the Acoustical Society of America*, 44:631.

[Collins, 1985] Collins, C. C. (1985). On mobility aids for the blind. *In* Strelow, D. H. W. . E. R., éditeur : *Electronic spatial sensing for the blind*, pages 35–64. Boston : Martinus Nijhoff.

[Costa *et al.*, 2008] Costa, G., Gusberti, A., Graffigna, J. P., Guzzo, M. et Nasisi, O. (2008). Mobility and orientation aid for blind persons using artificial vision. *Journal of Physics - Conference Series*, 90.

[Damaske et Wagener, 1969] Damaske, P. et Wagener, B. (1969). Richtungshorversuche uber einen nachgebildeten Kopf. *Acustica*, 21(30):118.

[Delalande *et al.*, 1996] Delalande, F., Formosa, M., Fremiot, M., Gobin, P., Malbosc, P., Mandelbrojt, J. et Pelder, E. (1996). *Les Unités Sémiotiques Temporelles - Éléments nouveaux d'analyse musicale.*

[Devallez *et al.*, 2008] Devallez, D., Fontana, F. et Rocchesso, D. (2008). Linearizing auditory distance estimates by means of virtual acoustics. *Acta Acustica united with Acustica*, 94(6):813–824.

[Dingler *et al.*, 2008] Dingler, T., Lindsay, J. et Walker, B. (2008). Learnabiltiy of sound cues for environmental features : Auditory icons, earcons, spearcons, and speech. *Methods*, pages 1–6.

[Djelani *et al.*, 2000] Djelani, T., Pörschmann, C., Sahrhage, J. et Blauert, J. (2000). An interactive virtual-environment generator for psychoacoustic research ii : Collection of head-related impulse responses and evaluation of auditory localization. *Acta Acustica united with Acustica*, 86(6):1046–1053.

[Dobrucki *et al.*, 2010] Dobrucki, A., Plaskota, P., Pruchnicki, P., Pec, M., Bujacz, M. et Strumillo, P. (2010). Measurement system for personalized head-related transfer functions and its verification by virtual source localization trials with visually impaired and sighted individuals. *J. Audio Eng. Soc*, 58(9):724–738.

[Dombois, 2002] Dombois, F. (2002). Auditory seismology – on free oscillations, focal mechanisms, explosions, and synthetic seismograms. *In Proceedings of the 8th International Conference on Auditory Display (ICAD2002)*, Kyoto, Japan.

[Dombois et Eckel, 2011] Dombois, F. et Eckel, G. (2011). Audification. *In* Hermann, T., Hunt, A. et Neuhoff, J. G., éditeurs : *The Sonification Handbook*. Logos Publishing House.

[Doucet *et al.*, 2005] Doucet, M.-E., Guillemot, J.-P., Lassonde, M., Gagné, J.-P., Leclerc, C. et Lepore, F. (2005). Blind subjects process auditory spectral cues more efficiently than sighted individuals. *Experimental brain research*, 160(2):194–202.

[Dramas, 2010] Dramas, F. (2010). *Localisation d'objets pour les non-voyants : augmentation sensorielle et neuroprothèse.* Thèse de doctorat, Université de Tolouse.

[Dramas *et al.*, 2008a] Dramas, F., Katz, B. F. et Jouffrais, C. (2008a). Auditory-guided reaching movements in the peripersonal frontal space (poster). *In Acoustics 08, Paris, 29/06/2008-04/07/2008*, volume 123, page 3723. J Acoust Soc Am.

[Dramas *et al.*, 2008b] Dramas, F., Katz, B. F. et Jouffrais, C. (2008b). Auditory-guided reaching movements in the peripersonal frontal space (poster). *In Acoustics 08, Paris, 29/06/2008-04/07/2008*, volume 123, page 3723. J. Acoust. Soc. Am.

[Dramas *et al.*, 2010] Dramas, F., Thorpe, S. et Jouffrais, C. (2010). Artificial vision for the blind : A bio-inspired algorithm for objects and obstacles detection. *International Journal of Image and Graphics*, 10(4):531–544.

[Durette *et al.*, 2008] Durette, B., Louveton, N., Alleysson, D. et Hérault, J. (2008). Visuo-auditory sensory substitution for mobility assistance : testing TheVIBE. *In Workshop on Computer Vision Applications for the Visually Impaired*. Département Images et Signal Département Images et Signal.

[Durlach *et al.*, 1993] Durlach, N., Shinn-Cunningham, B. et Held, R. (1993). Supernormal auditory localization. i - general background. *Presence : Teleoperators and Virtual Environments*, 2(2):89–103.

[Edworthy, 1998] Edworthy, J. (1998). Does sound help us to work better with machines ? *Interact. Comput.*, 10:401–409.

[Edworthy *et al.*, 1991] Edworthy, J., Loxley, S. et Dennis, I. (1991). Improving auditory warning design : Relationship between warning sound parameters and perceived urgency. *Human Factors*, 33:205–231.

[Farcy et Damaschini, 1997] Farcy, R. et Damaschini, R. (1997). Triangulating laser profilometer as a three-dimensional space perception system for the blind. *Applied Optics*, 36:8227–8232.

[Farcy *et al.*, 2003] Farcy, R., Leroux, R., Damaschini, R., Legras, R., Bellik, Y., Jacquet, C., Greene, J. et Pardo, P. (2003). Laser telemetry to improve the mobility of blind people : Report of the 6 month training course. *In Proc. of the International Conference On Smart homes and health Telematics, ICOST 2003*.

[Farcy *et al.*, 2006] Farcy, R., Leroux, R. et Jucha, A. (2006). Electronic travel aids and electronic orientation aids for blind people : technical, rehabilitation and everyday life points of view. *In* Hersh, M. A., éditeur : *CVHI 2006 : Conference on assistive technology for vision and hearing impairment*.

[Fitch et Kramer, 1992] Fitch, W. T. et Kramer, G. (1992). Sonifying the body electric : Superiority of an auditory over a visual display in a complex multivariate system. *In Proceedings of the International Conference on Auditory Display (ICAD1992)*. Addison Wesley Longman".

[Flowers et Hauer, 1993] Flowers, J. H. et Hauer, T. A. (1993). "sound" alternatives to visual graphics for exploratory data analysis. *Behavior Research Methods, Instruments & Computers*, 25(2):242–249.

[Fontana *et al.*, 2002a] Fontana, F., Fusiello, A., Gobbi, M., Murino, V., Rocchesso, D., Sartor, L. et Panuccio, A. (2002a). A cross-modal electronic travel aid device. *In Proceedings of the 4th International Symposium on Mobile Human-Computer Interaction*, Mobile HCI '02, pages 393–397. Springer-Verlag.

[Fontana et Rocchesso, 2003] Fontana, F. et Rocchesso, D. (2003). A physics-based approach to the presentation of acoustic depth. *In Proceedings of the International Conference on Auditory Display (ICAD2003)*, pages 79–82.

[Fontana et Rocchesso, 2008] Fontana, F. et Rocchesso, D. (2008). Auditory distance perception in an acoustic pipe. *ACM Trans. Appl. Percept.*, 5(3):16 :1–16 :15.

[Fontana *et al.*, 2002b] Fontana, F., Rocchesso, D. et Ottaviani, L. (2002b). A structural approach to distance rendering in personal auditory displays. *In IEEE International Conference on Multimodal Interfaces (ICMI 2002)*.

[Frauenberger *et al.*, 2007] Frauenberger, C., de Campo, A. et Eckel, G. (2007). Analysing time series data. *In Proceedings of the 13th International Conference on Auditory Display*, Montréal, Canada.

[Frémiot, 1999] Frémiot, M. (1999). De l'objet musical aux unités sémiotiques temporelles. *In Ouïr, entendre, écouter, comprendre après Schaeffer*, pages 227–246. INA/Buchet-Chastel.

[Frey *et al.*, 2009a] Frey, A., Daquet, A., Poitrenaud, S., Tijus, C., Frémiot, M., Prod'Homme, L., Mandelbrojt, J., Timsit-Berthier, M., Bootz, P., Hautbois, X. et Besson, M. (2009a). Pertinence cognitive des unités sémiotiques temporelles. *MusicæScientiæ*, XIII(2):415–440.

[Frey *et al.*, 2009b] Frey, A., Marie, C., Prod'Homme, L., Timsit-Berthier, M., Schön, D. et Besson, M. (2009b). Temporal semiotic units as minimal meaningful units in music ? an electrophysiological approach. *Music Perception*, 26(3):247–256.

[Friberg et Gärdenfors, 2004] Friberg, J. et Gärdenfors, D. (2004). Audio games : new perspectives on game audio. *In Proceedings of the 2004 ACM SIGCHI International Conference on Advances in computer entertainment technology*, pages 148–154.

[Gallay *et al.*, 2012] Gallay, M., Denis, M. et Auvray, M. ((in press) 2012). Navigation assistance for blind pedestrians : Guidelines for the design of devices and implications for spatial cognition. *In* Thora Tenbrink, J. et Claramunt, C., éditeurs : *Representating space in cognition : Interrelations of behaviour, language, and formal models.* Oxford University Press (UK).

[Gallay *et al.*, 2010] Gallay, M., Denis, M., Parseihian, G. et Auvray, M. (2010). Egocentric and allocentric reference frames in a virtual auditory environment : differences in navigation skills between blind and sighted individuals. *In European Workshop on Imagery & Cognition*, Helsinki, Finland.

[Gardner, 1968] Gardner, M. B. (1968). Proximity image effect in sound localization. *J. Acoust. Soc. Am.*, 43(1):163–163.

[Garzonis *et al.*, 2009] Garzonis, S., Jones, S., Jay, T. et O'Neill, E. (2009). Auditory icon and earcon mobile service notifications : intuitiveness, learnability, memorability and preference. *In Proceedings of the 27th international conference on Human factors in computing systems*, CHI '09, pages 1513–1522, New York, NY, USA. ACM.

[Gaudy *et al.*, 2006] Gaudy, T., Natkin, S. et Archambault, D. (2006). Playing audiogames without instructions for uses : To do without instruction leaflet or without language itself ? *In CGAMES'06, Int. Conf. on Computer Games, Dublin, Ireland*, pages 263–268.

[Gaunet et Briffault, 2005] Gaunet, F. et Briffault, X. (2005). Exploring the functional specifications of a localized wayfinding verbal aid for blind pedestrians : Simple and structured urban areas. *Human Computer Interaction*, 20:267–314.

[Gaver, 1986] Gaver, W. (1986). Auditory icons : using sound in computer interfaces. *Hum.-Comput. Interact.*, 2:167–177.

[Gaver, 1997] Gaver, W. (1997). *Handbook of Human-Computer Interaction*, chapitre Auditory Interfaces. Helander, M.G. and Landauer, T.K. and Prabhu, P.

[Gerzon, 1985] Gerzon, M. A. (1985). Ambisonics in multichannel broadcasting and video. *J. Audio Eng. Soc*, 33(1):859–871.

[Gifford *et al.*, 2006] Gifford, S., Knox, J., James, J. et Prakash, A. (2006). Introduction to the talking points project. *In Assets '06 : Proceedings of the 8th international ACM SIGACCESS conference on Computers and accessibility*, pages 271–272.

[Godbout et Boyd, 2010] Godbout, A. et Boyd, J. E. (2010). Corrective sonic feedback for speed skating : A case study. *In Proceedings of the 16th International Conference on Auditory Display.*

[Golledge *et al.*, 1998] Golledge, R., Klatzky, R., Loomis, J., Speigle, J. et Tietz, J. (1998). A geographical information system for a gps based personal guidance system.

[Golledge *et al.*, 2004] Golledge, R. G., Marston, J. R., Loomis, J. M. et Klatzky, R. L. (2004). Stated preferences for components of a personnal guidance system for nonvisual navigation. *J of Visual Impairment & Blindness*, 98(3):135–147.

[Gonzalez-Mora *et al.*, 2006] Gonzalez-Mora, J., Rodriguez-Hernandez, A., Burunat, E., Martin, F. et Castellano, M. (2006). Seeing the world by hearing : Virtual acoustic space (vas) a new space perception system for blind people. *In Information and Communication Technologies, 2006. ICTTA '06. 2nd.*

[Grond et Berger, 2011] Grond, F. et Berger, J. (2011). Parameter mapping sonification. *In* Hermann, T., Hunt, A. et Neuhoff, J. G., éditeurs : *The Sonification Handbook*. Logos Publishing House.

[Guillon, 2009] Guillon, P. (2009). *Individualisation des indices spectraux pour la synthèse binaurale : recherche et exploitation des similarités inter-individuelles pour l'adaptation ou la reconstruction de HRTF*. Thèse de doctorat, Université du Maine.

[Haas et Edworthy, 2006] Haas, E. et Edworthy, J. (2006). An introduction to auditory warnings and alarms. *In* Wogalter, M. S., éditeur : *Handbook of warnings*. Mahwah, NJ : Lawrence Erlbaum.

[Haber *et al.*, 1993] Haber, L., Haber, R. N., Penningroth, S., Novak, K. et Radgowski, H. (1993). Comparison of nine methods of indicating the direction to objects : data from blind adults. *Perception*, 22(1):35–47.

[Hanneton *et al.*, 2010] Hanneton, S., Auvray, M. et Durette, B. (2010). The Vibe : a versatile vision-to-audition sensory substitution device. *Applied Bionics and Biomechanics*, 7(4):269–276.

[Hartmann, 1989] Hartmann, W. M. (1989). On the minimum audible angle - a decision theory approach. *J. Acoust. Soc. Am.*, 85(5):2031–2041.

[Haustein, 1969] Haustein, B. (1969). Entfernungswahrnehmung des menschlichen Gehörs. *Hochfrequenztech. und Elektroakustik*, 78:46–57.

[Haustein et Schirmer, 1970] Haustein, B. G. et Schirmer, W. (1970). Messeinrichtung zur Untersuchung des Richtungslokalisationsvermogens. *Hochfrequenztech. und Elektroakustik*, 79: 96–101.

[Helal *et al.*, 2001] Helal, A., Moore, S. et Ramachandran, B. (2001). Drishti : An integrated navigation system for visually impaired and disabled. *In Proceedings of the 5th IEEE International Symposium on Wearable Computers*, ISWC '01, page 149, Washington, DC, USA. IEEE Computer Society.

[Helle *et al.*, 2001] Helle, S., Leplâtre, G., Marila, J. et Laine, P. (2001). Menu sonification in a mobile phone - a prototype study. *In Proceedings of the 7th International Conference on Auditory Display (ICAD2001)*.

[Henze *et al.*, 2006] Henze, N., Heuten, W. et Boll, S. (2006). Non-intrusive somatosensory navigation support for blind pedestrians. *In Proc. of EuroHaptics06*.

[Hermann, 2011] Hermann, T. (2011). Model-based sonification. *In* Hermann, T., Hunt, A. et Neuhoff, J. G., éditeurs : *The Sonification Handbook*. Logos Publishing House.

[Hermann *et al.*, 2001] Hermann, T., Hansen, M. H. et Ritter, H. (2001). Sonification of markov chain monte carlo simulations. *In Proceedings of 7th International Conference on Auditory Display*, pages 208–216, Helsinki, Finland.

[Hermann *et al.*, 2011] Hermann, T., Hunt, A. et Neuhoff, J., éditeurs (2011). *The Sonification Handbook*. Logos Publishing House, Berlin, Germany.

[Hermann *et al.*, 2002] Hermann, T., Meinicke, P., Bekel, H., Ritter, H., Müller, H. M. et Weiss, S. (2002). Sonification for eeg data analysis. *In Proceedings of the 8th International Conference on Auditory Display (ICAD2002)*.

[Hermann et Ritter, 1999] Hermann, T. et Ritter, H. (1999). Listen to your data : Model-based sonification for data analysis. *In* GE, L., éditeur : *Advances in intelligent computing and multimedia systems*, Baden-Baden, Germany.

[Hershkowitz et Durlach, 1969] Hershkowitz, R. et Durlach, N. (1969). Interaural time and amplitude jnds for a 500-hz tone. *J. Acoust. Soc. Am.*, 46(6B):1464–1467.

[Heuten *et al.*, 2008] Heuten, W., Henze, N., Boll, S. et Pielot, M. (2008). Tactile wayfinder : a non-visual support system for wayfinding. *In Proceedings of the 5th Nordic conference on Human-computer interaction : building bridges*, NordiCHI '08, pages 172–181, New York, NY, USA. ACM.

[Hofman *et al.*, 1998] Hofman, P. M., Van Riswick, J. G. et Van Opstal, A. J. (1998). Relearning sound localization with new ears. *Nature neuroscience*, 1(5):417–421.

[Holland et Morse, 2001] Holland, S. et Morse, D. (2001). Audiogps : spatial audio in a minimal attention interface. *In Third International Workshop on Human Computer Interaction with Mobile Devices*.

[Holland *et al.*, 2002] Holland, S., Morse, D. et Gedenryd, H. (2002). Audio gps : Spatial audio navigation with a minimal attention interface. *Personal and Ubiquitous Comput.*, 6(4):253–259.

[Honda *et al.*, 2007] Honda, A., Shibata, H., Gyoba, J., Saitou, K., Iwaya, Y. et Suzuki, Y. (2007). Transfer effects on sound localization performances from playing a virtual three-dimensional auditory game. *Applied Acoustics*, 68(8):885–896.

[Hub *et al.*, 2006] Hub, A., Hartter, T. et Ertl, T. (2006). Interactive tracking of movable objects for the blind on the basis of environment models and perception-oriented object recognition methods. *In Proceedings of the 8th international ACM SIGACCESS conference on Computers and accessibility*, Assets '06, pages 111–118, New York, NY, USA. ACM.

[ISO, 3407] ISO (13407). Processus de conception centrée sur l'opérateur humain pour les systèmes interactifs. Septembre 1999.

[Iwaya, 2006] Iwaya, Y. (2006). Individualization of head-related transfer functions with tournament-style listening test : Listening with other's ears. *Acoustical Science and Technology*, 27(6):340–343.

[Jacobson et Kitchin, 1997] Jacobson, R. et Kitchin, R. (1997). Gis and people with visual impairments or blindness : Exploring the potential for education, orientation, and navigation. *Transactions in Geographic Information Systems*, 2(4):315–332.

[Jacquet *et al.*, 2006] Jacquet, C., Bellik, Y. et Bourda, Y. (2006). Electronic locomotion aids for the blind : Towards mode assistive systems. *In Studies in Computational Intelligence, Intelligent Paradigms in Assistive and Preventive Healthcare*.

[James, 1996] James, F. (1996). Presenting html structure in audio : User satisfaction with audio hypertext. *In Proceedings of the 3rd International Conference on Auditory Display (ICAD96)*.

[Javer et Schwarz, 1995] Javer, A. et Schwarz, D. (1995). Plasticity in human directional hearing. *J. Otolaryngol.*, 24(2):111 – 117.

[Jin *et al.*, 2000] Jin, C., Leong, P., Leung, J., Corderoy, A. et Carlile, S. (2000). Enabling individualized virtual auditory space using morphological measurements. *In Proceedings of the First IEEE Pacific-Rim Conference on Multimedia (2000 International Symposium on Multimedia Information Processing*, pages 235–238.

[Kaczmarek, 2000] Kaczmarek, K. A. (2000). Sensory augmentation and substitution. *In* Bronzinov, J. D., éditeur : *The Biomedical Engineering Handboo*. Boca Raton : CRC Press LLC.

[Kahana et Nelson, 2007] Kahana, Y. et Nelson, P. A. (2007). Boundary element simulations of the transfer function of human heads and baffled pinnae using accurate geometric models. *J Sound and Vibration*, 300(3-5):552–579.

[Kammoun *et al.*, 2010] Kammoun, S., Dramas, F., B., O. et Jouffrais, C. (2010). Route selection algorithm for blind pedestrian. *In Intl. Conf. on Control, Automation and Systems, IEEE, KINTEX, Gyeonggi-do, Korea.*

[Kammoun *et al.*, 2012] Kammoun, S., Parseihian, G., Gutierrez, O., Brilhault, A., Serpa, A., Raynal, M., Oriola, B., Macé, M., Auvray, M., Denis, M., Thorpe, S., Truillet, P., Katz, B. et Jouffrais, C. (2012). Navigation and space perception assistance for the visually impaired : The navig project. *Ingénierie et Recherche Biomédicale*, 33:182–189.

[Katz et Parseihian, 2012] Katz, B. et Parseihian, G. (2012). Perceptually based head-related transfert function database optimization. *J. Acoust. Soc. Am.*, 131(2):EL99–EL105.

[Katz *et al.*, 2011] Katz, B., Rio, E. et Piccinali, L. (2011). LIMSI Spatialization Engine. Inter Deposit Digital Number : F.001.340014.000.S.P.2010.000.31235.

[Katz, 2001] Katz, B. F. (2001). Boundary element method calculation of individual head-related transfer function. II. Impedance effects and comparisons to real measurements. *J. Acoust. Soc. Am.*, 110(5):2440–2448.

[Katz *et al.*, 2012a] Katz, B. F., Dramas, F., Parseihian, G., Gutierrez, O., Kammoun, S., Brilhault, A., Brunet, L., Auvray, M., Truillet, P., Denis, M., Thorpe, S. et Jouffrais, C. (2012a). Navig : Guidance system for the visually impaired using virtual augmented reality. *J of Technology and Disability*, 24(2).

[Katz *et al.*, 2012b] Katz, B. F., Kammoun, S., Parseihian, G., Gutierrez, O., Brilhault, A., Auvray, M., Truillet, P., Denis, M., Thorpe, S. et Jouffrais, C. (2012b). Navig : Augmented reality guidance system for the visually impaired. combining object localization, gnss, and spatial audio. *Virtual Reality*, 16(3).

[Kawai *et al.*, 2000] Kawai, Y., Kobayashi, M., Minagawa, H., Miyakawa, M. et Tomita, F. (2000). A support system for visually impaired persons using three-dimensional virtual sound. *In int. Conf. Computers Helping People with special Needs (ICCHP 2000).*

[Kearney *et al.*, 2012] Kearney, G., Gorzel, M., Rice, H. et Boland, F. (2012). Distance perception in interactive virtual acoustic environments using first and higher order ambisonic sound fields. *Acta Acustica united with Acustica*, 98(1):61–71.

[Keuroghlian et Knudsen, 2007] Keuroghlian, A. S. et Knudsen, E. I. (2007). Adaptive auditory plasticity in developing and adult animals. *Progress in Neurobiology*, 82(3):109–121.

[Kim *et al.*, 2001] Kim, H., Suzuki, Y., Takane, S. et Sone, T. (2001). Control of auditory distance perception based on the auditory parallax model. *Applied Acoustics*, 62(3):245–270.

[Kim et Choi, 2005] Kim, S.-M. et Choi, W. (2005). On the externalization of virtual sound images in headphone reproduction : a wiener filter approach. *J. Acoust. Soc. Am.*, 117(6):3657–3665.

[King, 2009] King, A. (2009). Visual influences on auditory spatial learning. *Philosophical Transactions of the Royal Society of London*, 364(1515):331 – 339.

[Klatzky *et al.*, 2006] Klatzky, R., Marston, J., Giudice, N., Golledge, R. et Loomis, J. (2006). Cognitive load of navigating without vision when guided by virtual sound versus spatial language. *J of Exp Psychol Appl*, 12(4):223–232.

[Knudsen, 1984] Knudsen, E. I. (1984). The role of auditory experience in the development and maintenance of sound localization. *Trends in Neurosciences*, 7(9):326–330.

[Kopčo *et al.*, 2008] Kopčo, N., Schoolmaster, M. et Shinn-Cunningham, B. (2008). Learning to judge distance of nearby sounds in reverberant and anechoic environments.

[Kopčo et Shinn-Cunningham, 2011] Kopčo, N. et Shinn-Cunningham, B. (2011). Effect of stimulus spectrum on distance perception for nearby sources. *J. Acoust. Soc. Am.*, 130(3):1530–1541.

[Kraig, 2010] Kraig, F. (2010). The usability metric for user experience. *Interact. Comput.*, 22(5):323–327.

[Kramer, 1993] Kramer, G. (1993). *Auditory Display : Sonification, Audification and Auditory Interfaces*. Perseus Publishing.

[Kramer *et al.*, 1999] Kramer, K., Walker, B., Bonebright, T., Cook, P., Flowers, J., Miner, N. et Neuhoff, J. (1999). Sonification report : Status of the field and research agenda. Faculty Publications, Department of Psychology.

[Kuhn, 1977] Kuhn, G. F. (1977). Model for the interaural time differences in the azimuthal plane. *J. Acoust. Soc. Am.*, 62:157–167.

[Kulkarni et Colburn, 2000] Kulkarni, A. et Colburn, H. (2000). Variability in the characterization of the headphone transfer-function. *J. Acoust. Soc. Am.*, 107(2):1071–1074.

[Kupers et Ptito, 2004] Kupers, R. et Ptito, M. (2004). "Seeing" through the tongue : cross-modal plasticity in the congenitally blind. *International Congress Series*, 1270:79–84.

[Langlois *et al.*, 2010] Langlois, S., Loiseau, S., Tardieu, J., Cera, A. et Misdariis, N. (2010). Evaluation de la sonification d'un système multimédia automobile. *In 22ème conférence francophone sur l'Interaction Homme-Machine*.

[Larcher, 2001] Larcher, V. (2001). *Techniques de spatialisation des sons pour la réalité virtuelle*. Thèse de doctorat, Université Pierre & Marie Curie, Paris.

[Larcher et Jot, 1997] Larcher, V. et Jot, J. (1997). Techniques d'interpolation de filtres audio-numériques, application à la reproduction spatiale des sons sur écouteurs. *Proc. CFA :Congrès Français d'Acoustique*.

[Lee *et al.*, 2004] Lee, S. L., Kim, L. H. et Sung, K. M. (2004). Reduction of sound localization error for non-individualized HRTF by directional weighting function. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, 87(6):1531–1536.

[Leplâtre, 2002] Leplâtre, G. (2002). *The Design and Evaluation of non-Speech Sounds to Support Navigation in Restricted Display Devices*. Thèse de doctorat, Departement of Computing Science. University of Glasgow, Glasgow, UK.

[Leplâtre et Mcgregor, 2004] Leplâtre, G. et Mcgregor, I. (2004). How to tackle auditory interface aesthetics ? discussion and case study. *In Proceedings of the International Conference on Auditory Display (ICAD2004)*.

[Lessard *et al.*, 1998] Lessard, N., Paré, M., Lepore, F. et Lassonde, M. (1998). Early-blind human subjects localize sound sources better than sighted subjects. *Nature*, 395(6699):278–280.

[Lewald, 2002] Lewald, J. (2002). Vertical sound localization in blind humans. *Neuropsychologia*, 40(12):1868–1872.

[LISTEN, 2003] LISTEN (2003). IRCAM LISTEN HRTF database. http ://recherche.ircam.fr/equipes/salles/listen/.

[Lokki et Grohn, 2005] Lokki, T. et Grohn, M. (2005). Navigation with auditory cues in a virtual environment. *IEEE MultiMedia*, 12(2):80–86.

[Loomis, 1985] Loomis, J. (1985). Digital map and navigation system for the visually impaired. Unpublished manuscript, University of California, Santa Barbara.

[Loomis *et al.*, 1998a] Loomis, J., Golledge, R. et Klatzky, R. (1998a). Navigation system for the blind : auditory display modes and guidance. *Presence : Teleoper. Virtual Environ.*, 7:193–203.

[Loomis *et al.*, 1998b] Loomis, J., Klatzky, R., Philbeck, J. et Golledge, R. (1998b). Assessing auditory distance perception using perceptually directed action. *Attention, Perception, & Psychophysics*, 60(6):966–980.

[Loomis *et al.*, 2005] Loomis, J., Marston, J., Golledge, R. et Klatzky, R. L. (2005). Personal guidance system for people with visual impairment : A comparison of spatial displays for route guidance. *Journal of Visual Impairment & Blindness*, 99:219–232.

[Loomis *et al.*, 2000] Loomis, J. M., Golledge, R. G. et Klatzky, R. L. (2000). Gps-based navigation systems for the visually impaired. *In Symposium on Low-Vision at the Meeting of the American Academy of Optometry, Orlando, FL.*

[Loomis *et al.*, 1994] Loomis, J. M., Golledge, R. G., Klatzky, R. L., Speigle, J. M. et Tietz, J. (1994). Personal guidance system for the visually impaired. *In Proceedings of the first annual ACM conference on Assistive technologies*, Assets '94, pages 85–91. ACM.

[Lucas, 1994] Lucas, P. (1994). An evaluation of the communicative ability of auditory icons and earcons. *In Proceedings of the Second International Conference on Auditory Display (ICAD1994)*, pages 121–128, Santa Fe, NM, U.S. International Community for Auditory Display.

[Majdak *et al.*, 2010] Majdak, P., Goupell, M. J. et Laback, B. (2010). 3-d localization of virtual sound sources : Effects of visual environment , pointing method , and training. *Attention, Perception, & Psychophysics*, 72(2):454–469.

[Majdak *et al.*, 2008] Majdak, P., Laback, B., Goupell, M. et Mihocic, M. (2008). The accuracy of localizing virtual sound sources : Effects of pointing method and visual environment. *In Proc. of the 124th AES Convention, NL–Amsterdam, May 17–20.*

[Makous, 1990] Makous, J. C. (1990). Two-dimensional sound localization by human listeners. *J. Acoust. Soc. Am.*, 87(5):2188–2200.

[Marston *et al.*, 2006] Marston, J., Loomis, J., Klatzky, R., Golledge, R. G. et Smith, E. (2006). Evaluation of spatial displays for navigation without sight. *ACM Transactions on Applied Perception*, 3.

[Marston *et al.*, 2007] Marston, J. R., Loomis, J. M., Klatzky, R. L. et Golledge, R. G. (2007). Nonvisual route following with guidance from a simple haptic or auditory display. *Journal of Visual Impairment Blindness*, 101(4):203–211.

[Martin *et al.*, 2001] Martin, R. L., McAnally, K. I. et Senova, M. A. (2001). Free-field equivalent localization of virtual audio. *J. Audio Eng. Soc*, 49(1/2):14–22.

[McAdams et Bigand, 1993] McAdams, S. et Bigand, E. (1993). *Thinking in sound : the cognitive psychology of human audition*. Clarendon Press/Oxford University Press, New York, NY, USA.

[McAnally et Martin, 2002] McAnally, K. I. et Martin, R. L. (2002). Variability in the headphone-to-ear-canal transfer function. *J. Audio Eng. Soc*, 50(4):263–266.

[McKeag et McGrath, 1996] McKeag, A. et McGrath, D. S. (1996). Sound field format to binaural decoder with head tracking. *In Audio Engineering Society Convention 6r*.

[Meier et Saranti, 2008] Meier, M. et Saranti, A. (2008). Sonic explorations with earthquake data. *In Proceedings of the 14th International Conference on Auditory Display*, Paris, France.

[Meijer, 1992] Meijer, P. B. (1992). An experimental system for auditory image representations. *IEEE Transactions on Biomedical Engineering*, 39(2):112–121.

[Mershon et King, 1975] Mershon, D. et King, L. (1975). Intensity and reverberation as factors in the auditory perception of egocentric distance. *Attention, Perception, & Psychophysics*, 18(6): 409–415.

[Middlebrooks *et al.*, 2000] Middlebrooks, J., Macpherson, E. et Onsan, Z. (2000). Psychophysical customization of directional transfer functions for virtual sound localization. *J. Acoust. Soc. Am.*, 108(6):3088 – 3091.

[Middlebrooks, 1999a] Middlebrooks, J. C. (1999a). Individual differences in external-ear transfer functions reduced by scaling in frequency. *J. Acoust. Soc. Am.*, 106(3):1480–1492.

[Middlebrooks, 1999b] Middlebrooks, J. C. (1999b). Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency. *J. Acoust. Soc. Am.*, 106(3):1493–1510.

[Mills, 1958] Mills, A. W. (1958). On the minimum audible angle. *J. Acoust. Soc. Am.*, 30(4):237–246.

[Minard *et al.*, 2010] Minard, A., Misdariis, N., Houix, O. et Susini, P. (2010). Catégorisation de sons environnementaux sur la base de profils morphologiques (Environmental sounds categorization on the basis of morphological profiles). *In 10ème Congrès Français d'Acoustique*, Lyon, France.

[Møller *et al.*, 1995] Møller, H., Hammershøi, D., Jensen, C. et Sørensen, M. F. (1995). Transfer characteristics of headphones measured on human ears. *J. Audio Eng. Soc*, 43(4):203–217.

[Moore, 1997] Moore, B. C. (1997). *An introduction to the psychology of hearing (4th ed.)*. Academic Press, San Diego, Calif.

[Moore *et al.*, 2003] Moore, D., Amitay, S. et Hawkey, D. (2003). Auditory perceptual learning. *Learn. Mem.*, 10:83 – 85.

[Morimoto et Aokata, 1984] Morimoto, M. et Aokata, H. (1984). Localization cues of sound sources in the upper hemisphere. *J Acoust Soc Jpn*, 5(3):165–173.

[Nicol, 2010] Nicol, R. (2010). *Binaural technology.* Audio Engeneering Society.

[Nielsen, 1992] Nielsen, S. (1992). Auditory distance perception in different rooms. *In Audio Engineering Society Convention 92.*

[Noisternig *et al.*, 2003] Noisternig, M., Sontacchi, A., Musil, T. et Holdrich, R. (2003). A 3d ambisonic based binaural sound reproduction system. *In Audio Engineering Society Conference : 24th International Conference : Multichannel Audio, The New Reality.*

[Oldfield et Parker, 1984] Oldfield, S. R. et Parker, S. P. (1984). Acuity of sound localisation : A topography of auditory space : I. normal hearing conditions. *Perception*, 13(5):581–600.

[Olivan *et al.*, 2004] Olivan, J., Kemp, B. et Roessen, M. (2004). Easy listening to sleep recordings : tools and examples. *Sleep Medecine*, 5(6):601–603.

[Ortega-González *et al.*, 2010a] Ortega-González, V., Garbaya, S. et Merienne, F. (2010a). Reducing reversal errors in localizing the source of sound in virtual environment without head tracking. *In Proceedings of the 5th international conference on Haptic and audio interaction design*, HAID'10, pages 85–96, Berlin, Heidelberg. Springer-Verlag.

[Ortega-González *et al.*, 2010b] Ortega-González, V., Garbaya, S. et Merienne, F. (2010b). Using 3d sound for providing 3d interaction in virtual environment. *In ASME Conference Proceedings*, pages 311–321.

[Ortiz et Wright, 2009] Ortiz, J. et Wright, B. (2009). Contributions of procedure and stimulus learning to early, rapid perceptual improvements. *J. Experimental Psychology : Human Perception and Performance*, 35(1):188 – 194.

[Paquier *et al.*, 2011] Paquier, M., Koehl, V. et Jantzem, B. (2011). Effects of headphone transfer function scattering on sound perception. *In IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA).*

[Park *et al.*, 2005] Park, M., Choi, S., Kim, S. et Bae, K. (2005). Improvement of front-back sound localization characteristics in headphone-based 3d sound generation. *In Proc. Internet and Multimedia Systems, and Applications.*

[Parseihian *et al.*, 2010] Parseihian, G., Brilhault, A. et Dramas, F. (2010). Navig : An object localization system for the blind. *In Workshop Pervasive 2010 : Multimodal Location Based Techniques for Extreme Navigation*, Helsinki.

[Parseihian *et al.*, 2012] Parseihian, G., Conan, S. et Katz, B. (2012). Sound effect metaphors for near field distance sonification. *In Proceedings of the 18th international conference on Auditory display (ICAD 2012).*

[Parseihian et Katz, 2010] Parseihian, G. et Katz, B. (2010). Recalibrating the auditory system through audio-kinesthesic training. *In European Workshop on Imagery & Cognition*, Helsinki, Finland.

[Parseihian et Katz, 2011] Parseihian, G. et Katz, B. (2011). Rapid auditory system adaptation using a virtual auditory environment. *In International Multisensory Research Forum*, Fukuoka, Japan.

[Parseihian et Katz, 2012a] Parseihian, G. et Katz, B. (2012a). Morphocons : A new sonification concept based on morphological earcons. *J. Audio Eng. Soc*, 60(6):409–418.

[Parseihian et Katz, 2012b] Parseihian, G. et Katz, B. (2012b). Rapid head-related transfer function adaptation using a virtual auditory environment. *J. Acoust. Soc. Am.*, 131(4).

[Pauletto et Hunt, 2009] Pauletto, S. et Hunt, A. (2009). Interactive sonification of complex data. *International Journal of Human-Computer Studies*, 67(11):923 – 933.

[Peeters et Deruty, 2008] Peeters, G. et Deruty, E. (2008). Automatic morphological description of sounds. *In Proc of Acoustics'08 Paris.*

[Pelczynski *et al.*, 2006] Pelczynski, P., Strumillo, P. et Bujacz, M. (2006). Formant-based speech synthesis in auditory presentation of 3d scene elements to the blind. *In ACOUSTICS High Tatras 06 - 33rd International Acoustical Conference - EAA Symposium.*

[Peres *et al.*, 2007] Peres, S., Kortum, P. et Stallmann, K. (2007). Auditory progress bars preference, performance, and aesthetics. *In Proceedings of the International Conference on Auditory Display (ICAD2007).*

[Pereverzev *et al.*, 1997] Pereverzev, S. V., Loshak, A., Backhaus, S., Davis, J. C. et Packard, R. E. (1997). Quantum oscillations between two weakly coupled reservoirs of superfluid he-3. *Nature*, 388:449–451.

[Pernaux, 2003] Pernaux, J.-M. (2003). *Spatialisation du son par les techniques binaurales : application aux services de télécommunications.* Thèse de doctorat, INPG, France Telecom RD, Lanion (France).

[Picinali *et al.*, 2010] Picinali, L., Menelas, B., Katz, B. F. et Bourdot, P. (2010). Evaluation of a haptic/audio system for 3-d targeting tasks. *In Audio Engineering Society Convention 128.*

[Pralong et Carlile, 1996] Pralong, D. et Carlile, S. (1996). The role of individualized headphone calibration for the generation of high fidelity virtual auditory space. *J. Acoust. Soc. Am.*, 100(6):3785–3793.

[Preibisch-Effenberger, 1966] Preibisch-Effenberger, R. (1966). *Die Schallokalisationsfahigkeit des Menschen und ihre audiometrische Verwendbarkeit zur klinischen Diagnostik.*

[Ran *et al.*, 2004] Ran, L., Helal, S. et Moore, S. (2004). Drishti : An integrated indoor/outdoor blind navigation system and service. *In Proceedings of the Second IEEE International Conference on Pervasive Computing and Communications (PerCom'04)*, PERCOM '04, page 23, Washington, DC, USA. IEEE Computer Society.

[Rayleigh, 1907] Rayleigh, L. (1907). On our perception of sound direction. *Philosophical Magazine*, 13:214–232.

[Rébillat, 2011] Rébillat, M. (2011). *Vibrations de plaques multi-excitateurs de grandes dimensions pour la création d'environnements virtuels audio-visuels. Approches acoustique, mécanique et perceptive.* Thèse de doctorat, École polytechnique.

[Recanzone, 1998] Recanzone, G. H. (1998). Rapidly induced auditory plasticity : the ventriloquism aftereffect. *Natl Acad Sci U S A*, 95(3):869–875.

[Rickards, 1999] Rickards, T. (1999). Brainstorming. *In* Runco, M. et Pritzker, S., éditeurs : *Encyclopedia of creativity*, volume 1, pages 219–227. Academic Press.

[Röber *et al.*, 2006] Röber, N., Deutschmann, E. C. et Masuch, M. (2006). Authoring of 3d virtual auditory environments. *In Proceedings of the Audio Mosty Conference - a Conference on Sound in Games*.

[Robinson et Summerfield, 1996] Robinson, K. et Summerfield, A. (1996). Adult auditory learning and training. *Ear & Hearing*, 17:51–65.

[Roentgen *et al.*, 2008] Roentgen, U., Gelderblom, G., Soede, M. et Witte, L. (2008). Inventory of electronic mobility aids for persons with visual impairments : a literature review. *J of Visual Impairment & Blindness*, 102(11):702–724.

[Rosenblum *et al.*, 1987] Rosenblum, L., Carello, C. et Pastore, R. (1987). Relative effectiveness of three stimulus variables for locating a moving sound source. *Perception*, 16(2):175–186.

[Sanders et McCormick, 1993] Sanders, M. S. et McCormick, E. J. (1993). *Human factors in engineering and design (7th ed.)*. New York, NY, England : Mcgraw-Hill Book Company.

[Sandvad, 1996] Sandvad, J. (1996). Dynamic aspects of auditory virtual environments. *In Poc. 1000th Convention of the Audio Eng. Soc.*, Copenhagen, Denmark.

[Savel *et al.*, 2009] Savel, S., Drake, C. et Rabau, G. (2009). Human auditory localisation in a distorted environment : water. *Acta Acoustica United with Acoustica*, 95(1):128 – 141.

[Schaeffer, 1977] Schaeffer, P. (1977). *Traité des objets musicaux (Treatise on Musical Objects)*. Seuil.

[Schaffert *et al.*, 2009] Schaffert, N., Mattes, K., Barrass, S. et Effenberg, A. (2009). Exploring function and aesthetics in sonifications for elite sports. *In Proceedings of the Second International Conference on Musicc Communication Science*.

[Schärer et Lindau, 2009] Schärer, Z. et Lindau, A. (2009). Evaluation of equalization methods for binaural signals. *In Audio Engineering Society Convention 126*.

[Schonstein *et al.*, 2008] Schonstein, D., Ferré, L. et Katz, B. (2008). Comparison of headphones and equalization for virtual auditory source localization. *In Proceedings of the Acoustics'08 Conference. Paris (France) : European Acoustics Association*.

[Schönstein et Katz, 2010] Schönstein, D. et Katz, B. F. G. (2010). Sélection de HRTF dans une base de données en utilisant des paramètres morphologiques pour la synthèse binaurale. *In 10ème Congrès Français d'Acoustique*, Lyon, France.

[Schwarz, 2007] Schwarz, D. (2007). Corpus-based concatenative synthesis. *Signal Processing Magazine, IEEE*, 24(2):92–104.

[Seeber et Fasti, 2003] Seeber, B. U. et Fasti, H. (2003). Subjective selection of non-individual head related transfer function. *In Int. Conf. on Auditory Display, ICAD*, pages 259–262, Boston, MA, USA.

[Shinn-Cunningham, 2000] Shinn-Cunningham, B. (2000). Learning reverberation : Considerations for spatial auditory displays. *In Proceedings of the International Conference on Auditory Displays*, pages 126–134. Citeseer.

[Shinn-Cunningham *et al.*, 1998a] Shinn-Cunningham, B. G., Durlach, N. I. et Held, R. M. (1998a). Adapting to supernormal auditory localization cues. I. Bias and resolution. *J. Acoust. Soc. Am.*, 103(6):3656–3666.

[Shinn-Cunningham *et al.*, 1998b] Shinn-Cunningham, B. G., Durlach, N. I. et Held, R. M. (1998b). Adapting to supernormal auditory localization cues. II. Constraints on adaptation of mean response. *J. Acoust. Soc. Am.*, 103(6):3667–3676.

[Shinn-Cunningham *et al.*, 2000] Shinn-Cunningham, B. G., Santarelli, S. et Kopčo, N. (2000). Tori of confusion : Binaural localization cues for sources within reach of a listener. *J. Acoust. Soc. Am.*, 107(3):1627–1636.

[Shoval *et al.*, 1998] Shoval, S., Borenstein, J. et Koren, Y. (1998). The navbelt - a computerized travel aid for the blind based on mobile robotics technology. *IEEE Transactions on Biomedical Engineering*, 45(11):1376–1386.

[Sikora *et al.*, 1995] Sikora, C., Roberts, L. et Murray, L. (1995). Musical vs. real world feedback signals. *In Conference companion on Human factors in computing systems*, CHI '95, pages 220–221, New York, NY, USA. ACM.

[Simonov *et al.*, 2008] Simonov, M., Sciarappa, A. et Manca, S. (2008). Personal augmenting device and assistive living model for visually impaired. *In 5th International Workshop on Wearable Micro and Nanosystems for Personalised Health - pHealth 2008 (Valencia – Spain, May 21-23)*.

[Smalley, 1986] Smalley, D. (1986). Spectro-morphology and structuring processes. *In* Emmerson, S., éditeur : *The Language of Electroacoustic Music*. London : Macmillan.

[Smith, 1992] Smith, J. O. (1992). Physical modeling using digital waveguides. *Computer Music Journal*, 16(4):74–91.

[Soechting et Flanders, 1989] Soechting, J. et Flanders, M. (1989). Sensorimotor representations for pointing to targets in three-dimensional space. *J Neurophysiology*, 62(2):582–594.

[Sorkin, 1987] Sorkin, R. D. (1987). Design of auditory and tactile display. *In* Salvendy, G., éditeur : *Handbook of Human Factors*, pages 549–576. New York : Wiley & Sons.

[Spence et Driver, 1997] Spence, C. et Driver, J. (1997). Audiovisual links in attention : Implications for interface design. *In* Harris, D., éditeur : *Engineering Psychology and Cognitive Ergonomics Vol. 2 : Job Design and Product Design*. Hampshire : Ashgate Publishing.

[Steele et Chon, 2007] Steele, D. et Chon, S. (2007). A perceptual study of sound annoyance. *In Proceedings of Audio Mostly 2007*.

[Stockman *et al.*, 2007] Stockman, T., Rajgor, N., Metatla, O. et Harrar, L. (2007). The design of interactive audio soccer. *In Proceedings of the 13th International Conference on Auditory Display*.

[Stockman *et al.*, 2005] Stockman, T., Valgerour Nickerson, L. et Hind, G. (2005). Auditory graphs : A summary of current experience and towards a research agenda. *In Proceedings of the International Conference on Auditory Display (ICAD2005).*

[Stratton, 1896] Stratton, G. (1896). Some preliminary experiments on vision without inversion of the retinal image. *Psychological Review*, 3(6):611 – 617.

[Strothotte *et al.*, 1995] Strothotte, T., Petrie, H., Johnson, V. et Reichert, L. (1995). Mobic : user needs and preliminary design for a mobility aid for blind travellers. *In Proceedings of the 2nd Tide congress.*

[Suied *et al.*, 2008] Suied, C., Susini, P. et McAdams, S. (2008). Evaluating warning sound urgency with reaction times. *J Experimental Psychology : Applied*, 14(3):201–212.

[Tang et Beebe, 2003] Tang, H. et Beebe, D. J. (2003). Design and microfabrication of a flexible oral electrotactile display. *Journal Of Microelectromechanical Systems*, 12(1):29–36.

[Thoret *et al.*, 2012] Thoret, E., Aramaki, M., Kronland-Martinet, R., Velay, J. L. et Ystad, S. (2012). From shape to sound : sonification of two dimensional curves by reenaction of biological movements. *In 9th International Symposium on Computer Music Modeling and Retrieval*, London.

[Thorpe *et al.*, 1996] Thorpe, S., Fize, D. et Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, 381(6582):520–522.

[Tissot, 2010] Tissot, G. (2010). La notion de morphologie sonore et le développement des technologies en musique électroacoustique : deux éléments complémentaires d'une unique esthétique ? *In actes des Journées d'Informatique Musicales (JIM 2010).*

[Tohoku, 2001] Tohoku (2001). Tohoku HRTF database. http ://www.ais.riec.tohoku.ac.jp/lab/db-hrtf/.

[Tran *et al.*, 2000] Tran, T., Letowski, T. et Abouchacra, K. (2000). Evaluation of acoustic beacon characteristics for navigation tasks. *Ergonomics*, 43(6):807–827.

[Tünnermann et Hermann, 2009] Tünnermann, R. et Hermann, T. (2009). Multi-touch interactions for model-based sonification. *In Proceedings of the 15th International Conference on Auditory Display (ICAD2009)*, Copenhagen, Denmark.

[Van Wanrooij et Van Opstal, 2005] Van Wanrooij, M. et Van Opstal, A. (2005). Relearning sound localization with a new ear. *J. Neurosci.*, 25(22):5413 – 5424.

[Verron *et al.*, 2010] Verron, C., Aramaki, M., Kronland-Martinet, R. et Pallone, G. (2010). A 3D Immersive Synthesizer for Environmental Sounds. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(6):1550–1561.

[Vézien *et al.*, 2009] Vézien, J. M., Ménélas, B., Nelson, J., Picinali, L., Bourdot, P., Ammi, M., Katz, B., Burkhardt, J., Pastur, L. et Lusseyran, F. (2009). Multisensory vr exploration for computer fluid dynamics in the corsaire project. *Virtual Reality*, 13(4):257–271.

[Völk *et al.*, 2008] Völk, F., Heinemann, F. et Fastl, H. (2008). Externalization in binaural synthesis : effects of recording environment and measurement procedure. *J. Acoust. Soc. Am.*, 123(5):3935–3935.

[von Békésy, 1960] von Békésy, G. (1960). *Experiments in hearing*. McGraw-Hill Book Compagny.

[Walker et Kogan, 2009] Walker, B. et Kogan, A. (2009). Spearcon performance and preference for auditory menus on a mobile phone. *In* Stephanidis, C., éditeur : *Universal Access in Human-Computer Interaction. Intelligent and Ubiquitous Interaction Environments*, volume 5615, pages 445–454. Springer Berlin / Heidelberg.

[Walker et Kramer, 2004] Walker, B. et Kramer, G. (2004). Ecological psychoacoustics and auditory displays : Hearing, grouping, and meaning making. *In* Neuhoff, J., éditeur : *Ecological psychoacoustics*, pages 150–175. New York : Academic Press.

[Walker et Lindsay, 2003] Walker, B. et Lindsay, J. (2003). Effect of beacon sounds on navigation performance in a virtual reality environment. *In Proceedings of the Ninth International Conference on Auditory Display (ICAD2003)*, pages 204–207, Boston, USA. Boston, MA.

[Walker et Lindsay, 2005] Walker, B. et Lindsay, J. (2005). Navigation performance in a virtual environment with bonephones. *In Proc. of the Int'l Conf. on Auditory Display (ICAD2005)*, pages 260–263.

[Walker et Lindsay, 2006] Walker, B. et Lindsay, J. (2006). Navigation performance with a virtual auditory display : Effects of beacon sound, capture radius, and practice. *Human Factors : The Journal of the Human Factors and Ergonomics Society Summer*, 48(2):265–278.

[Walker *et al.*, 2006] Walker, B., Nance, A. et Lindsay, J. (2006). Spearcons : speech-based earcons improve navigation performance in auditory menus. *In Proceedings of the International Conference on Auditory Display (ICAD2006)*, pages 95–98.

[Walker et Kramer, 1996] Walker, B. N. et Kramer, G. (1996). Mappings and metaphors in auditory displays : An experimental assessment. *In* Frysinger, S. P. et Kramer, G., éditeurs : *Proceedings of the 3rd International Conference on Auditory Display (ICAD96)*.

[Walker et Nees, 2011] Walker, B. N. et Nees, M. A. (2011). Theory of sonification. *In* Hermann, T., Hunt, A. et Neuhoff, J. G., éditeurs : *The Sonification Handbook*, chapitre 2, pages 9–39. Logos Publishing House,, Berlin, Germany.

[Wenzel, 1999] Wenzel, E. (1999). Effect of increasing system latency on localization of virtual sounds. *In 16th International Conference : Spatial Sound Reproduction (March 1999)*.

[Wenzel *et al.*, 1988] Wenzel, E., Wightman, F. et Foster, S. (1988). A virtual display system for conveying three-dimensional acoustic information. *In Proc. Human Factos Soc. 32nd Ann. Meeting*, volume 1, pages 86–90.

[Wenzel *et al.*, 1993] Wenzel, E. M., Arruda, M., Kistler, D. J. et Wightman, F. L. (1993). Localization using nonindividualized head-related transfer functions. *J. Acoust. Soc. Am.*, 94(1): 111–123.

[Wenzel *et al.*, 1991] Wenzel, E. M., Wightman, F. L. et Kistler, D. J. (1991). Localization with non-individualized virtual acoustic display cues. *In CHI '91 : Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 351–359, New York, NY, USA. ACM.

[Wightman et Kistler, 2005] Wightman, F. et Kistler, D. J. (2005). Measurement and validation of human hrtfs for use in hearing research. *Acta Acustica united with Acustica*, 91(3):429–439.

[Wightman et Kistler, 1989a] Wightman, F. L. et Kistler, D. J. (1989a). Headphone simulation of free-field listening. I : Stimulus synthesis. *J. Acoust. Soc. Am.*, 85(2):858–867.

[Wightman et Kistler, 1989b] Wightman, F. L. et Kistler, D. J. (1989b). Headphone simulation of free-field listening. II : Psychophysical validation. *J. Acoust. Soc. Am.*, 85(2):868–878.

[Wightman et Kistler, 1993] Wightman, F. L. et Kistler, D. J. (1993). Multidimensional scaling analysis of head-related transfer functions. *In IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*.

[Wightman et Kistler, 1999] Wightman, F. L. et Kistler, D. J. (1999). Resolution of front-back ambiguity in spatial hearing by listener and source movement. *J. Acoust. Soc. Am.*, 105(5):2841–2853.

[Willis et Helal, 2005] Willis, S. et Helal, S. (2005). Rfid information grid for blind navigation and wayfinding. *Wearable Computers, IEEE International Symposium*, pages 34–37.

[Wilson *et al.*, 2007] Wilson, J., Walker, B., Lindsay, J., Cambias, C. et Dellaert, F. (2007). Swan : System for wearable audio navigation. *In Proceedings of the 2007 11th IEEE International Symposium on Wearable Computers*, pages 1–8, Washington, DC, USA. IEEE Computer Society.

[Woodworth et Schlosberg, 1954] Woodworth, R. et Schlosberg, H. (1954). Experimental psychology.

[Xenakis, 1979] Xenakis, I. (1979). *Arts/Sciences*. Alliages, Tournai, Casterman.

[Xu *et al.*, 2009] Xu, S., Li, Z. et Salvendy, G. (2009). Identification of anthropometric measurements for individualization of head-related transfer functions. *Acta Acustica united with Acustica*, 95(1):168 – 177.

[Yamagishi et Ozawa, 2011] Yamagishi, D. et Ozawa, K. (2011). Effects of timbre on learning to remediate sound localization in the horizontal plane. *In Principles and applications of spatial hearing*. World Scientific.

[Young, 1928] Young, P. T. (1928). Auditory localization with acoustical transposition of the ears. *J Experimental Psychology*, 11(6):399–429.

[Zahorik, 2002] Zahorik, P. (2002). Assessing auditory distance perception using virtual acoustics. *J. Acoust. Soc. Am.*, 111(4):1832–1846.

[Zahorik *et al.*, 2006] Zahorik, P., Bangayan, P., Sundareswaran, V., Wang, K. et Tam, C. (2006). Perceptual recalibration in human sound localization : learning to remediate front-back reversals. *J. Acoust. Soc. Am.*, 120(1):343–359.

[Zahorik *et al.*, 2005] Zahorik, P., Brungart, D. et Bronkhorst, A. (2005). Auditory distance perception in humans : A summary of past and present research. *Acta Acoustica United with Acoustica*, 91(February 2003):409 – 420.

[Zhang *et al.*, 1998] Zhang, M., Tan, K.-C. et Er, M. (1998). Three-dimensional sound synthesis based on head-related transfert functions. *J. Audio Eng. Soc*, 46(10):836 – 844.

[Zheng *et al.*, 2009] Zheng, J., Winstanley, A., Pan, Z. et Coveney, S. (2009). Spatial characteristics of walking areas for pedestrian navigation. *Third International Conference on Multimedia and Ubiquitous Engineering*, 0:452–458.

[Zwiers *et al.*, 2001a] Zwiers, M. P., Van Opstal, A. J. et Cruysberg, J. R. (2001a). A spatial hearing deficit in early-blind humans. *J Neurosci*, 21(9):RC142 : 1–5.

[Zwiers *et al.*, 2001b] Zwiers, M. P., Van Opstal, A. J. et Cruysberg, J. R. (2001b). Two-dimensional sound-localization behavior of early-blind humans. *Experimental Brain Research*, 140(2):206–222.

[Zwiers *et al.*, 2003] Zwiers, M. P., Van Opstal, A. J. et Paige, G. D. (2003). Plasticity in human sound localization induced by compressed spatial vision. *Nature Neuroscience*, 6(2):175–181.