



HAL
open science

Étude de son 3D pour une interaction audio-visuelle en environnement virtuel enrichi

Erik Vladimir Ortega González

► **To cite this version:**

Erik Vladimir Ortega González. Étude de son 3D pour une interaction audio-visuelle en environnement virtuel enrichi. Synthèse d'image et réalité virtuelle [cs.GR]. Arts et Métiers ParisTech, 2011. Français. NNT : 2011ENAM003 . pastel-00602192

HAL Id: pastel-00602192

<https://pastel.hal.science/pastel-00602192>

Submitted on 21 Jun 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

2011-ENAM-003

École doctorale n°432 : Sciences des Métiers de l'Ingénieur

Doctorat ParisTech

THÈSE

pour obtenir le grade de docteur délivré par

l'École Nationale Supérieure d'Arts et Métiers

Spécialité " Informatique et traitement de signal "

présentée et soutenue publiquement par

Erik Vladimir ORTEGA GONZÁLEZ

le 21 février 2011

Etude de son 3D pour une interaction audio-visuelle en environnement virtuel enrichi

Directeur de thèse : **Frédéric MERIENNE**

Co-encadrement de la thèse : **Samir GARBAYA**

Jury

Mme. Nicol ROZENN , Chercheuse, France Télécom R&D, Orange Labs	Rapporteur
M. Jean-Louis VERCHER , Professeur, Institut des Sciences du Mouvement, Université de la Méditerranée	Rapporteur
M. Stephen BREWSTER , Professeur, Department of Computing Science, University of Glasgow	Examineur
M. Luis PASTOR , Professeur, Grupo de Modelado y Realidad Virtual, Universidad Rey Juan Carlos	Examineur
M. James OLIVER , Professeur, Virtual Reality Applications Center, Iowa State University	Examineur
M. Jean SRENG , Chercheur, Laboratoire de Simulation Interactive, Commission d'Energie Atomique	Examineur
M. Frédéric MERIENNE , Professeur, Institut Image, Arts et Métiers ParisTech	Directeur
M. Samir GARBAYA , Maître de Conférences, Institut Image, Arts et Métiers ParisTech	Examineur

**Arts et Métiers ParisTech – Institut Image de Chalon-Sur-Saône
Laboratoire Electronique Informatique et Image (Le2i – UMR CNRS 5158)**

Résumé

La plupart des applications qui intègrent le son 3D en environnement virtuel sont limitées à la simulation acoustique. L'objectif de cette thèse est d'étudier l'apport du son 3D dans l'interaction en environnement virtuel. Nous avons développé une technique de stimulation sonore appelée : Sensation Artificielle Spatialisée Auditive (SASA), basée sur la création d'effets sonores. Cette technique restitue des sensations de spatialisation de son permettant la localisation précise (en azimut et en élévation) de sources sonores. Afin d'améliorer le temps de localisation de sources sonores, nous avons développé un modèle qui intègre la fonction HRTF (*Head-Related Transfer Function*) avec la technique SASA. Pour simuler la profondeur de sources sonores, nous avons développé une technique basée sur la combinaison entre l'indice d'intensité et les différences interaurales de temps. Les techniques développées dans le cadre de cette thèse ont été utilisées pour le guidage auditif afin de fournir des informations spatiales en azimut, en élévation et en profondeur. Nous avons appliqué cette technique dans la simulation chirurgicale pour assister l'opérateur dans la phase de triangulation. Nous avons validé cette approche en termes de performance de manipulation et du retour d'informations de l'utilisateur. Les résultats obtenus des travaux de cette thèse sont prometteurs pour l'utilisation de son 3D dans l'interaction en environnement virtuel.

Mots-clés : Son 3D, HRTF, Localisation de sources sonores, Interfaces sonores, Environnement Virtuel

Abstract

Most of the applications integrating 3D sound in virtual environment are limited to acoustic simulation. The objective of this thesis is to study the added value of 3D sound in the interaction with virtual environment. We developed an audio stimulation technique called "Artificial Spatial Audio Sensation: (ASAS)" based on the creation of audio effects. This technique conveys the sensations of spatialized sound that allows accurate sound source localization in azimuth and in elevation. In order to improve the localization time we also developed a model that integrates Head-Related Transfer Function (HRTF) with the ASAS technique. For the simulation of the depth of spatial sound sources, we developed a technique based on a combined model that integrates the intensity cue and the interaural time differences. The techniques developed in this thesis were applied for audio cueing in order to convey accurate spatial information in azimuth, elevation and depth. The combined model was applied in surgical simulation to assist the operator in the execution of the triangulation gesture. We conducted an experimental validation of this approach in terms of the criteria of manipulation performance and user's appreciation. The results obtained in this research are promising for the use of 3D sound to enrich the interaction in virtual environment.

Keywords: · 3D sound, HRTF, Sound Source Localization, Sound Interfaces, Virtual Environment

Dédicace

à Aurélie pour son soutien et pour nos projets à venir
y a mis papás Rocío y César por su cariño y su ejemplo

Remerciements

Je tiens à remercier d'abord mes encadrants Samir Garbaya et Frédéric Merienne pour leur soutien et leur accompagnement durant la réalisation de ma thèse. Ces trois années ont été riches en apprentissage académique et personnel, et ceci, grâce à l'opportunité que vous m'avez donné de pouvoir réaliser ma thèse en France.

Mes parents, Rocío et César m'ont toujours encouragé dans la poursuite de mes études et ils m'ont appris l'importance de la culture de l'effort. La réalisation de ma thèse est aussi votre réussite. Merci pour être toujours avec moi, pour vos conseils, pour votre exemple mais au dessus de tout pour votre amour. Mon frère et sa famille sont aussi toujours dans mon cœur et je vous dis merci pour toute votre affection.

Aurélié, ton soutien a été sans doute ma force dans ces deux dernières années. C'est grâce à toi que j'ai pu trouver le courage pour réaliser ce défi qu'a représenté pour moi cette thèse. Merci d'être toujours là pour moi. Te amo.

Je suis également reconnaissant envers tous mes collègues de l'Institut Image de Chalon-sur-Saône pour leur collaboration. Je souhaite remercier spécialement Hanh, Jeong, Claudia, Simon, Florent, Baris, Mihail, Justin, Loïc, Ravidranath, Amine, Hamid et Diego pour leur amitié et pour tous les bons moments que nous avons partagés aussi bien à l'intérieur qu'à l'extérieur de l'Institut. Merci aussi à Géraldine pour son soutien et sa patience lors de toutes les démarches administratives. Et sans oublier bien sur d'exprimer ma reconnaissance à Laura, Julien, Ulises et Amparo qui m'ont énormément soutenu pendant mon séjour en France.

Il me semble également important de remercier Luis Pastor Pérez pour l'opportunité qu'il m'a donnée en me permettant de réaliser un stage doctoral à l'Université Rey Juan Carlos à Madrid en Espagne. De la même façon, je remercie tous les membres du Groupe de Modélisation et Réalité Virtuelle pour leur accueil et pour les échanges toujours intéressants et encourageants que nous avons eus.

Je suis très reconnaissant que le Grand Chalon et Arts et Métiers ParisTech aient financé mes travaux de recherches pendant ma thèse. J'ai également pu compter sur le soutien financier du Conseil National de Science et Technologie (CONACyT) au Mexique. Merci à cette grande institution pour rendre possible les rêves de jeunes chercheurs comme moi dans la poursuite de leur formation au delà de notre pays.

Mon niveau de français ne m'aurait pas permis de rédiger cette thèse avec autant de justesse sans l'aide précieuse d'Aurélié, de Frédéric et très spécialement de Samir pour la correction du texte à la virgule près. J'ai également beaucoup appris grâce à vous et je vous remercie chaleureusement de votre grande patience.

Je souhaite remercier finalement tous mes amis, collègues, anciens collègues, anciens professeurs, et membres de ma famille qui m'ont chacun appris beaucoup de choses dans la vie. Je vous apprécie et je vous remercie pour tous les moments partagés.

Sommaire

1. Introduction	1
1.1. Contexte.....	1
1.2. Problématique.....	2
1.3. Objectifs du projet de thèse	3
1.3.1. Objectif principal	3
1.3.2. Objectifs secondaires	4
1.4. Approche proposée.....	4
1.5. Contributions.....	5
1.6. Structure du document.....	6
2. Son 3D: aspects acoustiques et applications en réalité virtuelle	7
2.1. Écoute spatiale	7
2.1.1. Écoute spatiale	9
2.1.2. Localisation de sources sonores spatiales	9
2.2. Technologies de son 3D.....	10
2.2.1. Spatialisation sonore et HRTF	12
2.2.2. Acoustique virtuelle	14
2.3. Retour sonore et son 3D pour l'interaction en environnement virtuel	15
2.3.1. Retour sonore non spatial en environnement virtuel.....	15
2.3.2. Interaction en environnement virtuel	17
2.3.3. Son 3D pour l'interaction.....	17
2.3.4. Approches existantes pour améliorer la localisation de sources sonores en environnement virtuel	20
2.3.5. Application potentielle du son pour la manipulation	21
3. Techniques et outils employés.....	23
3.1. Concepts et techniques de traitement de signal.....	23
3.2. La fonction HRTF.....	28
3.2.1. Les fondements de la fonction HRTF	28
3.2.2. Individualisation et généralisation des fonctions HRTF	30
3.2.3. Base de données HRTF.....	31

3.2.4.	Modèle de la fonction HRTF	31
3.3.	<i>Méthode d'analyse statistique</i>	33
3.3.1.	Statistiques descriptives	33
3.3.2.	Représentations graphiques	34
3.3.3.	Distribution fréquentielle	35
3.3.4.	Expérimentation	36
3.3.5.	Méthode d'Equations Généralisées d'Estimation	38
3.3.6.	Méthode d'Estimation Généralisée de Moyennes Marginales	39
3.4.	<i>Plateforme expérimentale</i>	40
3.4.1.	Architecture logicielle	40
3.4.2.	Architecture logicielle	41
4.	Approche proposée : Sensations Auditives Spatiales Artificielles (SASA)	45
4.1.	<i>L'approche SASA</i>	45
4.1.1.	Description générale de l'approche SASA	45
4.1.2.	Structure d'un groupe SASA	46
4.1.3.	Méthode itérative de génération d'indices sensoriels SASA	50
4.1.4.	Intégration successive des indices SASA	52
4.2.	<i>Proposition préliminaire : groupe d'indices SASA pour la localisation sonore précise</i>	53
4.2.1.	Contexte	53
4.2.2.	Stimulus sonore	54
4.2.3.	Indices de groupe SASA pour une localisation précise	55
4.2.4.	Modèle de directionnalisation utilisé	57
4.2.5.	Protocole de validation expérimentale	59
4.2.6.	Résultats et analyse des données	62
4.2.7.	Distribution fréquentielle des données	66
4.2.8.	Analyse statistique	66
4.2.9.	Conclusions	69
5.	Étude comparative et intégration des techniques SASA et HRTF	71
5.1.	<i>Motivation de la recherche d'un modèle performant</i>	71
5.2.	<i>Modèle combiné HRTF - SASA</i>	71
5.3.	<i>Étude expérimentale</i>	73
5.3.1.	Essais et stratégie de localisation de source sonore	74
5.3.2.	Évaluation expérimentale	74
5.3.3.	Groupe de participants et hypothèses	75

5.4.	<i>Résultats expérimentaux et analyse de données</i>	75
5.4.1.	Évaluation de la performance	75
5.4.2.	Retour d'informations de l'utilisateur	81
5.4.3.	Comparaison de notre approche avec les techniques de spatialisation sonore existantes .	82
5.5.	<i>Conclusions</i>	83
6.	Approche CITD pour améliorer la simulation de la profondeur	85
6.1.	<i>Profondeur de sources sonores spatiales</i>	85
6.2.	<i>Simulation de la profondeur des sources sonores spatiales</i>	87
6.3.	<i>Approche proposée : méthode de Différences Interaurales de Temps Compensées</i>	88
6.3.1.	Méthode de Différences Interaurales de Temps Compensées (CITD).....	88
6.3.2.	Simulation de la profondeur basée sur l'indice d'intensité	93
6.4.	<i>Étude expérimentale</i>	93
6.4.1.	Évaluation de la performance de l'utilisateur	93
6.4.2.	Retour d'informations de l'utilisateur	96
6.5.	<i>Résultats expérimentaux et analyse de données</i>	98
6.5.1.	Évaluation de performance de l'utilisateur	98
6.6.	<i>Analyse du retour d'informations de l'utilisateur</i>	104
6.7.	<i>Conclusions</i>	105
7.	Application du Son 3D pour la simulation de chirurgie arthroscopique	107
7.1.	<i>La simulation de chirurgie arthroscopique</i>	107
7.1.1.	La simulation chirurgicale	107
7.1.2.	Le geste de triangulation et la manipulation des instruments	108
7.2.	<i>Le simulateur de chirurgie arthroscopique utilisé</i>	109
7.3.	<i>Approche : le Son 3D Enrichi pour guider le geste de triangulation</i>	111
7.3.1.	Le Son 3D Enrichi	112
7.3.2.	Méthode d'intégration du Son 3D Enrichi en environnement virtuel	112
7.3.3.	Application du Son 3D Enrichi pour guider le geste de triangulation dans la simulation chirurgicale	113
7.3.4.	Groupe d'indices SASA utilisé	119
7.4.	<i>Étude expérimentale</i>	121
7.4.1.	Évaluation de la performance	121
7.4.2.	Retour d'informations de l'utilisateur	125
7.5.	<i>Résultats et analyse de données</i>	126

7.5.1.	Résultats de l'évaluation de performance	126
7.5.2.	Analyse du retour d'informations de l'utilisateur	133
7.6.	<i>Conclusions</i>	134
8.	Conclusions et perspectives de recherche	137
8.1.	<i>Conclusions</i>	137
8.1.1.	Problématique scientifique traitée.....	137
8.1.2.	Solutions développées.....	138
8.1.3.	Résultats obtenus	138
8.2.	<i>Perspectives de recherche</i>	139
8.3.	<i>Publications réalisées</i>	139
8.3.1.	Articles soumis pour la publication dans des revues internationales	139
8.3.2.	Articles publiés dans des conférences internationales avec comité de lecture.....	140
8.3.3.	Poster	140
9.	Références	141
10.	Annexe : Questionnaires utilisés.....	147
10.1.	<i>Analyse du retour d'informations de l'utilisateur : Chapitre 5</i>	147
10.2.	<i>Analyse du retour d'informations de l'utilisateur : Chapitre 6</i>	149
10.3.	<i>Analyse du retour d'informations de l'utilisateur : Chapitre 7</i>	152
11.	Glossaire.....	153

Liste de figures

Figure 2.1 Les propriétés spatiales d'une source sonore.....	8
Figure 2.2. Les étapes de traitement du son 3D.	10
Figure 2.3. Différentes applications du son 3D pour l'interaction 3D.	18
Figure 3.1. La convolution obtenue par l'application de la réponse impulsionnelle d'un système.....	25
Figure 3.2. La convolution discrète obtenu par la Transformée Rapide de Fourier.	27
Figure 3.3. Diagramme de la méthode complète de superposition et d'addition pour l'application d'une réponse impulsionnelle à un signal discret long.....	29
Figure 3.4. Fonction HRTF d'une source sonore spatiale.....	32
Figure 3.5. Processus d'application de la HRTF en temps réel.....	33
Figure 3.6. Exemple d'un graphique de type boîte à moustaches avec deux échantillons.....	34
Figure 3.7. Exemple d'un graphique de moyennes avec deux échantillons.....	35
Figure 3.8. Exemple de protocole expérimental.....	38
Figure 3.9. Schéma de l'architecture matérielle de la plateforme expérimentale utilisée.	41
Figure 3.10. Schéma de l'architecture logicielle de la plateforme expérimentale développée.....	42
Figure 4.1. Structure générale d'un groupe d'indices SASA.	46
Figure 4.2. Propriétés spatiales d'une source sonore.	47
Figure 4.3. Exemple d'une fonction d'association d'un indice SASA et ses caractéristiques.	50
Figure 4.4. Méthode de génération d'un indice sensoriel SASA.	52
Figure 4.5. Méthode d'intégration d'un groupe d'indices SASA.....	53
Figure 4.6. (a) La forme d'onde du stimulus sonore sélectionné. (b) Le spectre fréquentiel du stimulus sonore obtenu par la méthode TFR (Transformée de Fourier Rapide). La fréquence indiquée à 1.7 kHz représente la valeur de la fréquence de coupure utilisée comme valeur critique pour les indices SASA.....	55
Figure 4.7. Fonctions d'association des indices sensoriels SASA. Les axes verticaux des figures (b) et (d) sont tracés en échelle logarithmique.....	58
Figure 4.8. (a) Interface graphique affichée pour l'expérimentateur uniquement ; (b) Photo d'un participant en train de réaliser l'expérience.	60
Figure 4.9. Erreurs angulaires mesurées (en rouge).....	61
Figure 4.10. (a) Temps d'accomplissement de la tâche et (b) erreur angulaire globale selon le sexe des utilisateurs.	64
Figure 4.11. (a) Erreur absolue en azimut et (b) erreur absolue en élévation selon le sexe des utilisateurs.....	64
Figure 4.12. Performance individuelle du (a) temps d'accomplissement de la tâche et (b) l'erreur angulaire globale.....	65
Figure 4.13. Distribution de la fréquence pour : (a) temps d'accomplissement de la tâche, (b) erreur angulaire globale, (c) erreur absolue en azimut et (d) erreur absolue en élévation.	67

Figure 5.1. Les différentes méthodes implantées pour la stimulation auditive spatiale : (a) SASA, (b) HRTF et (c) HRTF+SASA.....	72
Figure 5.2. (a) Temps d'accomplissement et (b) Erreur angulaire globale pour les conditions expérimentales évaluées.....	76
Figure 5.3. (a) Erreur en azimuth et (b) Erreur en élévation pour les conditions expérimentales évaluées.	77
Figure 5.4. Performance individuelle pour le (a) temps d'accomplissement de la tâche et pour (b) l'erreur angulaire globale. A, B et C représentent les conditions expérimentales de SASA, HRTF et HRTF+SASA respectivement.	78
Figure 5.5. Distribution de fréquence pour : (a) le temps d'accomplissement de la tâche et (b) l'erreur angulaire globale.....	79
Figure 5.6. L'appréciation des utilisateurs par rapport à (a) la facilité d'utilisation et (b) au réalisme de l'interaction.	81
Figure 6.1. Les différences interaurales de temps (ITD) et d'intensité (ILD).....	87
Figure 6.2. Représentation des distances comprises entre les oreillers de l'auditeur et la HRTF utilisée et des distances comprises entre les oreilles de l'auditeur et la source sonore.....	89
Figure 6.3. Applications du retard CITD aux HRIRs en fonction de son signe : (a) positif et (b) négatif.	91
Figure 6.4. Exemples des valeurs de CITD pour différentes profondeurs et localisations de sources sonores.	91
Figure 6.5. Méthodes de simulation de la profondeur de sources sonores spatiales basées sur (a) l'indice d'intensité et sur (b) la combinaison de l'indice d'intensité et le retard CITD.	92
Figure 6.6. Courbe de la perte de puissance de la source sonore en fonction de sa profondeur pour la méthode de l'indice d'intensité. La ligne rouge représente la distance de 5 mètres à partir de laquelle la technique de CITD n'a aucun effet.....	93
Figure 6.7. Interface graphique pendant (a) les essais préliminaires et (b) les essais d'évaluation.....	95
Figure 6.8. Expérimentations de (a) l'évaluation de la performance de l'utilisateur et (b) de l'évaluation du retour d'informations de l'utilisateur.....	96
Figure 6.9. Les trajectoires suivies par les sources sonores dans l'expérience de la perception de la profondeur.....	97
Figure 6.10. Temps d'accomplissement de la tâche pour les conditions expérimentales évaluées.....	99
Figure 6.11. Erreur absolue de localisation en profondeur pour les conditions expérimentales évaluées.....	100
Figure 6.12. La distribution fréquentielle de l'erreur non absolue de localisation en profondeur.	100
Figure 6.13. Performance individuelle : (a) temps d'accomplissement de la tâche et (b) erreur absolue de localisation. Les conditions expérimentales indiquées pour 1 et 2 représentent les modalités Normale et CITD respectivement.	101
Figure 6.14. La distribution de fréquence (a) du temps d'accomplissement et (b) de l'erreur absolue de localisation en profondeur.....	102
Figure 6.15. La distribution de fréquence de l'erreur absolue pour les échantillons correspondants aux conditions expérimentales (a) Normale et (b) CITD.	103
Figure 6.16. L'appréciation des sujets pour la perception de la profondeur dans les deux conditions expérimentales étudiées.	105
Figure 7.1. Mouvements de base du geste de triangulation : (a) latéral (b) vertical et (c) pénétration.....	109
Figure 7.2. Simulateur de chirurgie arthroscopique utilisé.	110

Figure 7.3. Interfaces graphiques du simulateur chirurgical.	110
Figure 7.4. Modèles graphiques de l'épaule.....	111
Figure 7.5. Schéma du traitement d'un stimulus sonore avec la combinaison de la HRTF, l'approche SASA et la technique CITD.....	112
Figure 7.6. Schéma de la procédure d'intégration de l'approche de Son 3D Enrichi pour une tâche d'interaction 3D en environnement virtuel.....	114
Figure 7.7. Procédure pour toucher la cible virtuelle dans la simulation chirurgicale.	115
Figure 7.8. Position du point de référence de la caméra.	116
Figure 7.9. Association des propriétés spatiales de la source sonore avec les degrés de liberté de l'outil d'interaction.	117
Figure 7.10. Procédure pour toucher une cible avec le guidage auditif fourni apporté par l'approche S3E.	119
Figure 7.11. Fonction d'association du nouvel indice SASA de verticalité. L'axe vertical est tracé en échelle logarithmique.	120
Figure 7.12. Les conditions expérimentales développées.	122
Figure 7.13. Différentes localisations de la cible.	123
Figure 7.14. Erreur instantanée de la position de l'instrument par rapport la trajectoire optimale.	125
Figure 7.15. Temps d'accomplissement de la tâche pour les conditions expérimentales évaluées.	127
Figure 7.16. Longueur de la trajectoire suivie par l'instrument pour les conditions expérimentales évaluées. ..	128
Figure 7.17. Indicateur de la distance moyenne pour les conditions expérimentales évaluées.	128
Figure 7.18. Performance individuelle pour (a) le temps d'accomplissement de la tâche, pour (b) la longueur de trajectoire et pour (c) l'indicateur de la distance moyenne .Les modalités d'opération du simulateur UM et DM sont représentées par 1 et 2 respectivement. Les modalités de stimulation auditive SS et S3E sont représentées par 1 et 2 respectivement.	129
Figure 7.19. Distribution de fréquence pour : (a) le temps d'accomplissement de la tâche, (b) la longueur de trajectoire et (c) l'indicateur de la distance moyenne.....	131
Figure 7.20. Appréciation des sujets de (a) la facilité d'utilisation et de (b) l'intuitivité.....	134

Liste de tableaux

Tableau 3.1. Architecture matérielle	41
Tableau 3.2. Architecture logicielle	42
Tableau 4.1. Indices sensoriels du groupe SASA pour la localisation précise	56
Tableau 4.2. Paramètres des indices SASA du groupe pour la localisation précise	56
Tableau 4.3. Essais expérimentaux : position de la source sonore à chercher	60
Tableau 4.4. Mesures statistiques des données obtenues	63
Tableau 4.5. Tests de qualité d'ajustement.	66
Tableau 4.6. Analyse de l'effet des facteurs expérimentaux par l'Analyse de Mesures Répétées avec le modèle d'Equations Généralisées d'Estimation	68
Tableau 5.1. Mesures statistiques des données obtenues	76
Tableau 5.2. Tests de qualité d'ajustement	78
Tableau 5.3. Estimation Généralisée de Moyennes Marginales par l'Analyse de Mesures Répétées avec le modèle d'Equations Généralisées d'Estimation	80
Tableau 5.4. Comparaison du modèle combiné avec d'autres approches publiées	83
Tableau 6.1. Essais expérimentales : Localisation et profondeur des sources sonores à présenter à l'utilisateur	94
Tableau 6.2. Mesures statistiques des données obtenues	98
Tableau 6.3. Tests de qualité d'ajustement	102
Tableau 6.4. Analyse de l'effet des facteurs expérimentaux par l'Analyse de Mesures Répétées avec le modèle d'Equations Généralisées d'Estimation	104
Tableau 7.1. Propriétés spatiales du groupe des indices SASA utilisé	119
Tableau 7.2. Indices sensoriels du groupe SASA pour la geste de triangulation	120
Tableau 7.3. Conditions expérimentales	121
Tableau 7.4. Mesures statistiques des données obtenues	127
Tableau 7.5. Tests de qualité d'ajustement	130
Tableau 7.6. Analyse de l'effet des facteurs expérimentaux par l'Analyse de Mesures Répétées avec le modèle d'Equations Généralisées d'Estimation	132
Tableau 7.7. Estimation Généralisée de Moyennes Marginales par l'Analyse de Mesures Répétées avec le modèle d'Equations Généralisées d'Estimation	132

1. Introduction

1.1. Contexte

Le son 3D est un domaine qui traite de la simulation des propriétés spatiales de sources sonores telles que la directionnalité, la profondeur et la réverbération. Pour une position donnée de la source sonore, les techniques existantes de son 3D permettent de lui attribuer des propriétés spatiales correspondantes à cette position. Les techniques courantes de spatialisation de son permettent de localiser une source sonore avec une précision similaire à celle obtenue dans le monde réel.

Le son 3D est une technologie mature car elle permet de trouver des solutions appropriées à la plupart des problèmes rencontrés et le coût du matériel utilisé est de moins en moins élevé. En outre, parmi les retours sensoriels communément utilisés en réalité virtuelle, l'audio spatial présente des avantages par rapport à ses équivalents (vision et retour haptique) pour ce qui concerne les coûts et la complexité technique. Le matériel de restitution du son 3D est généralement moins coûteux que le rendu visuel immersif ou encore que le retour haptique. Le niveau de maturité des techniques de son 3D et la faible résolution de l'écoute humaine comparée à la vision et au toucher permettent un rendu moins complexe avec des résultats assez réalistes pour la restitution d'environnements non échoïques.

La contribution du son 3D dans le réalisme et l'immersion de l'utilisateur a été étudiée dans différents travaux de recherches qui confirment que le son 3D a un effet important sur ces aspects de la perception humaine. Bien que le son 3D contribue efficacement aux aspects perceptuels dans la simulation acoustique, son utilisation comme un outil pour améliorer la qualité de l'interaction a été très peu étudiée. Dans ce contexte, nous envisageons d'étudier le son 3D au delà de son rôle traditionnel dans la simulation acoustique. Nous considérons qu'il peut constituer un retour sensoriel capable de donner des informations utiles dans l'interaction en environnement virtuel.

Le retour sonore a été étudié dans de nombreux travaux de recherches comme modalité sensorielle permettant de transmettre des informations significatives dans les systèmes interactifs. Les techniques d'interaction existantes basées sur le retour auditif ne permettent cependant pas de transmettre des informations spatiales. Pour cette raison, l'utilisation de ces techniques n'est pas complètement appropriée à l'interaction en espace tridimensionnel.

Les techniques d'interaction en environnement tridimensionnel peuvent se décomposer en quatre groupes principaux de tâches : la sélection, la manipulation d'objets, la navigation et le système de commande.

La plupart de ces techniques se basent sur le retour visuel. En revanche, l'utilisation du retour sonore pour l'interaction en environnement virtuel tridimensionnel reste peu étudiée.

Nous considérons que le son 3D a un potentiel important dans l'interaction en environnement virtuel car il permet de transmettre des informations en trois dimensions : azimut, élévation et profondeur. Ces informations peuvent être utilisées pour guider ou assister l'utilisateur dans la réalisation d'une tâche interactive en espace tridimensionnel telle que la navigation, la sélection ou la manipulation. Nous envisageons donc d'étudier le potentiel du son 3D et sa contribution à l'interaction en environnement virtuel.

1.2. Problématique

Pour que le son 3D puisse contribuer efficacement à l'interaction en environnement virtuel, il est nécessaire que l'information spatiale transmise à l'utilisateur soit précise, rapide à identifier et pertinente. Les techniques existantes sont limitées : elles ne permettent pas une localisation sonore précise, les meilleures solutions existantes nécessitent l'individualisation et il n'existe aucune méthode de référence pour intégrer ces informations dans une application en environnement virtuel. Dans les paragraphes suivants, nous allons développer ces derniers points.

Précision de la localisation sonore

La précision dans la localisation sonore des techniques existantes de spatialisation ne permet pas de fournir des informations auditives avec précision. En effet, l'erreur de localisation globale pour la direction de provenance varie selon les différents auteurs entre 10 et 25 degrés. De plus, les écoutes spatiales naturelles et simulées présentent des problèmes d'ambiguïtés dans la localisation sonore. Ces problèmes, appelés aussi erreurs inverses, proviennent des difficultés observées quant à la distinction de la provenance hémisphérique des sources sonores. Plus les informations auditives transmises sont imprécises et ambiguës et moins la technique contribue efficacement à l'interaction. De ce fait, le manque de précision dans la localisation sonore et la présence d'ambiguïté représente un obstacle à l'utilisation des techniques existantes de son 3D pour assister les utilisateurs dans l'exécution de tâches interactives 3D.

Afin d'utiliser le retour sonore spatial pour assister l'interaction, il est nécessaire de développer une solution pour améliorer la précision de la localisation sonore en environnement virtuel et pour éliminer ou au moins réduire significativement la présence d'ambiguïtés. Ceci permettra donc de transmettre des informations auditives spatiales précises et sans ambiguïtés utiles pour assister l'utilisateur dans la réalisation d'une tâche virtuelle.

Individualisation des techniques de restitution sonore

Les techniques existantes présentent également le problème suivant : celles offrant les meilleurs niveaux de précision sont souvent basées sur l'individualisation. C'est-à-dire que les techniques doivent être adaptées à chaque utilisateur afin d'assurer une meilleure restitution des conditions d'écoute spatiale pour que l'utilisateur perçoive les sons comme dans les conditions réelles équivalentes. L'individualisation nécessite du temps de

traitement et du matériel supplémentaire par rapport aux techniques dites généralisées. Afin d'éviter ceci, nous envisageons de développer une solution indépendante des caractéristiques propres à chaque utilisateur.

Restitution de la profondeur de sources sonores

Les techniques existantes pour simuler la profondeur de sources sonores présentent aussi des limites. La première de ces limites réside dans la difficulté à restituer tous les indices sensoriels présents en conditions réelles : indice d'intensité, différences binaurales, indice de réverbération et filtrage fréquentiel. Par ailleurs, la plupart des techniques existantes se limitent à fournir seulement l'indice d'intensité. Cet indice se réfère à la perte de puissance du son en fonction de la distance qu'il a parcourue depuis sa source. Les solutions bénéficiant des meilleurs résultats sont aussi basées sur l'individualisation. Nous envisageons donc de développer une technique généralisée permettant de fournir une sensation de profondeur proche des conditions réelles basée sur la combinaison de l'indice d'intensité et des différences binaurales. L'intégration de cette technique dans la solution développée permettra de transmettre plus efficacement des informations auditives en trois dimensions : azimut, élévation et profondeur.

Intégration et évaluation

Lorsque la solution est validée, nous envisageons de développer une méthode pour l'intégrer dans des applications interactives en espace virtuel afin de s'assurer que les informations fournies sont efficaces pour assister l'utilisateur dans la réalisation de la tâche.

Il sera ensuite nécessaire de déterminer la contribution de la solution proposée dans le cadre d'une application finale de réalité virtuelle. Nous développons un protocole expérimental qui permettra de déterminer l'effet de la technique proposée sur un ensemble de critères objectifs de performance ainsi que sur le retour d'informations de l'utilisateur obtenu par une évaluation subjective.

1.3. Objectifs du projet de thèse

1.3.1. Objectif principal

L'objectif principal de nos travaux de recherche est l'étude de la pertinence du son 3D dans l'interaction en environnement virtuel. Les jalons suivants ont été repérés pour mener à bien cette étude :

- identification des limites des techniques de son 3D existantes pour transmettre efficacement des informations spatiales,
- proposition et validation d'une nouvelle technique de spatialisation sonore,
- application de cette technique pour assister l'utilisateur dans la réalisation d'une tâche interactive.

1.3.2. Objectifs secondaires

En plus de l'objectif principal de la thèse, nous souhaitons adresser les objectifs scientifiques et technologiques secondaires suivants :

- développement d'une plateforme expérimentale pour l'étude de l'effet des techniques de son 3D sur l'interaction en environnement virtuel,
- développement, implantation et validation d'une solution pour améliorer la précision de la localisation sonore et pour réduire l'ambiguïté de la localisation sonore,
- développement d'une méthode pour intégrer la solution proposée dans des applications interactives en environnement virtuel,
- intégration et validation de la solution proposée dans le cadre d'une application finale de réalité virtuelle.

1.4. Approche proposée

L'approche proposée dans le cadre de cette thèse s'appuie sur la combinaison des différentes solutions développées séparément en fonction des principaux problèmes traités dans les travaux de recherche réalisés. Le principe général est d'utiliser des techniques de son 3D existantes et d'y ajouter des informations supplémentaires afin d'améliorer l'information à transmettre pour les critères de précision et de rapidité pour les trois dimensions : azimut, élévation et profondeur.

La première solution développée concerne le problème de manque de précision dans la localisation sonore pour les techniques existantes de son 3D. Nous avons proposé de générer des indices sensoriels artificiels qui complètent le stimulus sonore donné. Ces indices sensoriels ont pour but de fournir des informations pour guider l'utilisateur dans l'identification de la direction de la source sonore. Cette technique étant généralisée, elle permet donc d'éviter l'individualisation. Nous avons proposé également une méthode pour définir ces indices artificiels et pour les intégrer afin de s'assurer qu'ils sont efficaces pour obtenir une localisation sonore précise.

La première solution proposée permet une localisation plus précise mais significativement plus lente que les techniques existantes de son 3D. Nous avons donc mis en place une étude afin de comparer la première solution avec la technique de spatialisation de référence de la HRTF et d'évaluer leur éventuelle intégration. Nous avons constaté que l'intégration de ces deux techniques permet une localisation précise et rapide des sources sonores. Ce modèle combiné constitue la deuxième solution proposée. Elle permet de transmettre des informations auditives précises et rapides en azimut et en élévation, par contre, la dimension de la profondeur reste à étudier.

La troisième solution consiste à agréger l'information de profondeur au modèle de la HRTF. Cette solution combine l'indice d'intensité avec les différences binaurales. L'étude développée montre que cette solution permet de restituer une sensation de profondeur sonore plus précise et plus réaliste que l'approche basée sur l'indice d'intensité uniquement.

Finalement la solution retenue consiste en la combinaison de la deuxième et de la troisième approche. Nous avons développé une méthode pour intégrer la solution combinée dans une application de réalité virtuelle. Ensuite, nous avons mené une étude sur la contribution de la solution combinée dans le cadre d'une application de simulation chirurgicale.

Dans chaque étape, nous avons suivi une méthodologie de validation expérimentale basée sur l'évaluation de la performance et sur le retour d'informations de l'utilisateur. Le design expérimental consiste en l'observation d'un groupe d'utilisateurs qui exécutent une tâche dans différentes conditions. Les résultats obtenus sont analysés avec des méthodes statistiques selon les objectifs suivants : la caractérisation complète des données et la détermination de l'effet des facteurs expérimentaux observés sur les variables définies dans chaque expérience.

1.5. Contributions

Les contributions majeures de la thèse peuvent se classer selon les aspects suivants : techniques, théoriques, méthodologiques, pratiques (applications) et de communication scientifique.

La plateforme expérimentale développée nous a permis d'étudier l'effet des solutions proposées sur l'interaction en environnement virtuel. La plateforme permet de reproduire la plupart des fonctionnalités des autres systèmes similaires existants. Cette plateforme permettra également de poursuivre la recherche sur le son 3D dans l'interaction en environnement virtuel dans les travaux à venir.

Concernant l'aspect théorique, nous avons principalement proposé et mis en place une solution originale pour la transmission des informations auditives spatiales de manière efficace. Cette solution permet pour la première fois d'obtenir simultanément des informations auditives précises et rapides pour les dimensions d'azimut, d'élévation et de profondeur.

Les méthodes proposées permettent d'évaluer les effets des solutions développées quant à la performance mais aussi pour ce qui concerne la perception sur la base de critères objectifs et subjectifs (retour d'informations de l'utilisateur). La combinaison des aspects objectifs et subjectifs pour l'évaluation représente une contribution originale. L'analyse statistique développée permet de déterminer l'effet des facteurs expérimentaux observés sur les variables définies. De plus et à la différence de la plupart des travaux de recherche publiés dans le domaine, l'analyse réalisée permet aussi d'obtenir une caractérisation statistique complète des données obtenues qui tient compte par exemple de la nature de leur distribution.

En ce qui concerne l'aspect pratique, les techniques développées constituent une nouvelle forme de guidage auditif pour améliorer l'interaction en environnement virtuel tridimensionnel. De cette manière, notre approche permet d'intégrer efficacement l'utilisation du retour sonore pour assister l'interaction en espace tridimensionnel. Les techniques proposées représentent également une nouvelle méthode d'utilisation du son 3D au delà de l'application typique de la simulation acoustique. Cette thèse constitue donc une approche originale et efficace pour l'utilisation du son 3D pour l'interaction 3D.

Dans le cadre de cette thèse, nous avons réalisé différentes publications dans des conférences internationales et deux articles soumis dans des revues internationales. Toutes les publications élaborées dans le cadre de cette thèse sont citées dans le chapitre 8.

1.6. Structure du document

Le Chapitre 2 porte sur la description et la discussion de l'état de l'art des techniques existantes de spatialisation sonore et des approches existantes pour assister l'interaction basée sur le retour auditif. Le Chapitre 3 décrit les concepts et les aspects les plus importants des techniques utilisées ainsi que la plateforme développée dans le cadre de ces travaux. Le Chapitre 4 porte sur l'approche que nous avons développée et sa validation expérimentale. Le Chapitre 5 présente l'étude expérimentale réalisée sur la comparaison et l'intégration de l'approche proposée dans le chapitre 4 et la technique de spatialisation sonore de référence : la HRTF. Dans le Chapitre 6, nous présentons et validons une technique consistant à modifier la HRTF pour améliorer la sensation de profondeur par rapport à celle obtenue par la technique basée seulement sur l'indice de perte d'intensité en fonction de la distance parcourue par le son. Le Chapitre 7 décrit une étude sur l'utilisation des solutions combinées pour assister l'opérateur dans la réalisation d'une tâche de simulation chirurgicale. Les résultats obtenus montrent une amélioration importante avec l'utilisation du son 3D. Le Chapitre 8 est dédié à la conclusion et les perspectives de recherche.

2. Son 3D: aspects acoustiques et applications en réalité virtuelle

Le son spatialisé (ou son 3D) est une technologie relativement récente qui s'intéresse à la diffusion de sources sonores avec des propriétés spatiales. Son développement fut inspiré du phénomène de l'écoute spatiale. Les travaux scientifiques publiés dans le domaine de l'écoute spatiale permettent d'avoir une bonne compréhension du sujet.

Actuellement, différentes applications du son 3D existent dans les systèmes interactifs de Réalité Virtuelle (RV) et d'Interfaces Homme Machine (IHM). Dans la plupart de ces applications, le son est utilisé comme un moyen de simulation des conditions acoustiques de l'environnement et vise à augmenter le sens de présence en environnement virtuel de l'utilisateur. Cependant, l'utilisation du son 3D pour améliorer l'interaction en environnement virtuel a été très peu étudié.

Le travail de recherche de cette thèse présente une méthode et des outils pour utiliser les techniques de son 3D pour l'interaction en environnement virtuel. Deux facteurs principaux garantissent selon nous la pertinence de l'approche de sonification de l'environnement virtuel (EV): la maturité de la technologie de son 3D et l'existence de différents modèles efficaces de retour sonore pour l'interaction.

L'interaction 3D avec l'audio nécessite une bonne précision et la capacité de fournir des informations spatiales. Actuellement, les techniques de son 3D existantes manquent de précision, tout comme l'écoute spatiale réelle. Par ailleurs, les techniques de guidage auditif existantes ne sont pas capables de fournir des informations spatiales. Ce sont les raisons qui nous ont amenées à proposer l'utilisation d'une combinaison des techniques existantes de son 3D et de guidage audio.

Dans ce chapitre, différents aspects de l'écoute spatiale, du son 3D et de l'application du retour sonore en environnement virtuel sont présentés. Pour conserver un texte concis et le plus cohérent possible, nous nous sommes limités à traiter uniquement les parties des travaux référencés qui nous apparaissent les plus importants pour notre sujet de recherche.

2.1. Écoute spatiale

L'écoute spatiale est un terme qui se réfère à la capacité d'un être humain à distinguer les propriétés spatiales d'une source sonore. Ces propriétés permettent à l'homme de localiser la source du son dans un espace.

Chez l'être humain comme chez la plupart des espèces vertébrées, cette localisation est possible grâce à l'utilisation d'un système d'écoute à deux entrées positionnées à la même hauteur que celle de la tête (Blauert [1997]). Le terme d'écoute binaurale fait allusion littéralement au phénomène d'écoute à l'aide des deux oreilles. Une partie des études existantes utilise indifféremment les termes écoute spatiale et écoute binaurale.

La psychoacoustique est une branche de la psychophysique qui s'intéresse à la compréhension des mécanismes principalement sensoriels impliqués dans l'écoute des sons. Elle étudie divers aspects comme les limites et la résolution du système auditif pour les différentes propriétés physiques du son. La psychoacoustique est un domaine de recherche proche de la physiologie, la neurologie et plus particulièrement la psychologie expérimentale. Ainsi, l'étude de l'écoute spatiale par la psychoacoustique a permis, à l'aide de différentes expérimentations, de mieux connaître les limites et la résolution des propriétés spatiales d'écoute par l'homme. Dans cette section, nous allons décrire quelques travaux scientifiques en rapport avec ce thème pour présenter le contexte dans lequel notre approche a été proposée.

Un concept important de l'écoute spatiale est celui de source sonore ou source sonore spatiale. Une source sonore spatiale peut avoir trois propriétés essentielles : la directivité, la profondeur et la réverbération. Ces propriétés sont illustrées dans la Figure 2.1. Il est nécessaire que toute source sonore, pour être considérée comme telle, possède au moins la propriété de directivité. La directivité est la propriété principale d'une source sonore spatiale.

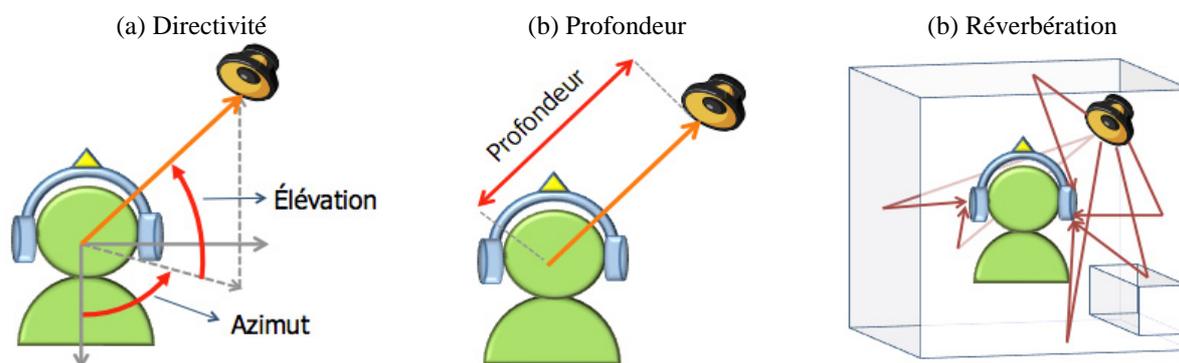


Figure 2.1 Les propriétés spatiales d'une source sonore.

Dans cette thèse, nous considérons que la notion de directivité d'une source sonore représente la direction de sa provenance. Ce concept est issu du domaine de l'électromagnétisme pour indiquer la direction d'un champ magnétique. Elle se compose de l'azimut et de l'élévation. L'azimut est l'angle constitué entre la position de la source sonore et celle de l'auditeur dans le plan horizontal. L'élévation est l'angle formé par la position de la source sonore et le plan horizontal. Ainsi, l'azimut correspond à la direction latérale de la source et l'élévation correspond à sa hauteur.

La profondeur se réfère à la distance entre la source et l'auditeur. L'information de profondeur est déterminée principalement par l'auditeur en fonction de l'intensité captée et des différences interaurales.

La réverbération se réfère à la résonance acoustique due aux surfaces environnantes. Cette résonance dépend de plusieurs facteurs tels que les propriétés acoustiques de l'environnement et ses composants, la position

de la source sonore et celle de l'auditeur. En fonction du contexte, cette information peut être importante pour la localisation de sources sonores et pour la détermination de la profondeur.

La spatialisation du son consiste à lui associer des propriétés spatiales de directivité, de profondeur et de réverbération. Par contre, la directionalisation consiste à lui associer des propriétés spatiales de directivité uniquement.

2.1.1. Écoute spatiale

Depuis les années soixante-dix, la recherche concernant l'écoute spatiale a sensiblement augmenté. Le nombre de travaux de recherche dans les techniques de spatialisation artificielle s'est accru dès le début des années quatre-vingt-dix. L'intégration des technologies de son 3D dans les systèmes d'interaction homme machine est un centre d'intérêt scientifique relativement récent, peu d'études ont été réalisées ces dix dernières années.

Blauert [1997] a publié un livre de référence sur le phénomène de l'écoute spatiale. Dans cet ouvrage, dont sa première version fut publiée en 1974, les auteurs décrivent des expérimentations sur la perception et les aspects physiologiques de l'écoute spatiale.

2.1.2. Localisation de sources sonores spatiales

La localisation de sources sonores se réfère le plus souvent au processus de détection de la direction de provenance d'une source sonore spatiale par des sujets humains. Il est également possible d'utiliser le terme de localisation sonore. La localisation sonore peut se référer aussi à la détermination de la profondeur d'une source sonore, c'est-à-dire la distance entre l'auditeur et la source. Néanmoins, nous allons utiliser dans ce texte le terme « localisation » ou « localisation de la direction de provenance » pour faire référence à la détermination de la direction de provenance de la source sonore. Pour la distance entre l'auditeur et le source sonore, nous utiliserons le terme de localisation en profondeur. Dans les paragraphes suivants, nous allons décrire un ensemble de travaux de recherche publiés sur la localisation sonore.

Knudsen et Konishi [1979] ont effectué des recherches sur le système de détection sonore de la chouette pour déterminer l'azimut et l'élévation d'une source sonore. Les auteurs ont observé la réponse aux différentes fréquences et diverses combinaisons d'obstruction physiques dans les oreilles. Ils ont conclu que la chouette utilise l'information spectrale (ou fréquentielle) pour déterminer l'élévation. La méthodologie développée et les résultats obtenus par Knudsen et Konishi [1979] représentent une référence très importante dans le domaine de la recherche en écoute spatiale.

Middlebrooks et Green [1991] ont publié un état de l'art exhaustif sur la recherche existante dans le domaine de l'écoute spatiale. Les auteurs affirment que l'être humain utilise principalement les différences interaurales pour déterminer l'azimut, et l'information spectrale pour déterminer l'élévation. D'après ces auteurs, le mouvement de la tête contribue à la localisation de sources sonores uniquement quand l'utilisateur a suffisamment de temps pour s'en servir.

Jin et al. [2004] ont étudié le rôle des indices fréquentielles monaurales et interaurales. Les auteurs considèrent que ces informations n'étaient pas suffisantes pour obtenir une localisation précise en élévation. Ces résultats montrent la nécessité d'ajouter plus d'informations pour améliorer la précision en élévation.

Simon Carlile et al. [1997] ont étudié la capacité à localiser la provenance d'une source de bruit. Leur principal objectif fut de déterminer la nature et la distribution de l'erreur dans la localisation de sources sonores. Ils avaient aussi pour objectif d'essayer de minimiser l'erreur due aux conditions d'observation et à la plateforme expérimentale. Les sources sonores furent signalées par les utilisateurs avec la tête et furent positionnées à l'aide d'un bras robot. Le suivi de la tête était effectué avec des dispositifs magnétiques. Les erreurs obtenues sont de l'ordre de 3 degrés pour l'azimut et entre 4 et 13 degrés pour l'élévation. Cette étude est importante car elle permet de mesurer la performance humaine de localisation sonore en milieu naturel.

La localisation automatique de sources sonores spatiales est un sujet de recherche différent qui risque d'être confondu avec la localisation de sources sonores. La localisation automatique de sources sonores se réfère à l'étude et au développement de mécanismes qui permettent d'identifier la provenance de sources sonores dans l'espace. Il existe un nombre important de publications sur ce sujet dont la majeure partie sont orientés vers des applications robotiques (John et Weng [1998] et Valin et al. [2003]).

2.2. Technologies de son 3D

Les technologies de son 3D servent à simuler l'écoute spatiale. Ces technologies peuvent être classées selon les différentes étapes de traitement du son 3D : la capture, la spatialisation et la restitution (voir Figure 2.2).



Figure 2.2. Les étapes de traitement du son 3D.

La première étape consiste en la capture des aspects suivants: le son diffusé, ses propriétés spatiales et son interaction avec l'environnement. Ces aspects définissent un champ sonore. Les techniques de capture de son spatial nécessitent l'utilisation d'équipements spécifiques tels que les chambres anéchoïques (ou sourdes) et les mécanismes d'enregistrement avancés. La détermination des propriétés spatiales d'une source sonore du monde réel est utile pour son association avec une source sonore équivalente dans le monde virtuel.

La spatialisation du son consiste à doter un stimulus sonore monocanal de propriétés spatiales en fonction d'une position indiquée. Le terme de spatialisation est souvent utilisé pour restituer uniquement la

propriété de directivité sans prendre en compte les autres propriétés spatiales telles que la profondeur et la réverbération.

L'étape de spatialisation est normalement effectuée à l'aide de processus de traitement de signal. Ces processus sont pour la plupart implantés sur des composants de logiciels mais il existe aussi des solutions matérielles pour la spatialisation (exemple : le dispositif Convolvotron, Foster et Elizabeth M. Wenzel [1992]). Le transfert implicite dans la spatialisation peut être basé sur des modèles mathématiques ou bien sur des données contenant les propriétés spatiales obtenues dans une étape de capture.

Il existe différents mécanismes pour la restitution de sources sonores en environnement virtuel. Ces méthodes utilisent souvent des casques ou des systèmes multicanaux. Les casques sont utilisés avec les techniques de rendu binaural (deux canaux).

Les systèmes multicanaux sont une solution courante utilisée dans les systèmes immersifs comme le CAVE. La reproduction de son 3D avec des enceintes présente au moins trois inconvénients majeurs (Begault [2005]) : elle dépend fortement de l'acoustique de l'environnement réel, elle n'est pas adaptée à des systèmes multi utilisateurs et elle nécessite une étape supplémentaire de traitement pour éliminer la diaphonie¹. Cependant, l'utilisation de techniques de rendu binaural et un casque isolant permet une restitution indépendante du contexte réel.

Il existe différentes techniques pour simuler la directivité des sources sonores. Les principales techniques de restitution de la directivité peuvent être classées en trois groupes : les techniques basées sur le positionnement géométrique, celles basées sur les différences interaurales² et celles basées sur l'anatomie humaine.

Le positionnement géométrique se base sur les systèmes de restitution sonore multicanaux. Son principe est la variation de l'intensité de la source sonore de chaque canal afin de restituer l'effet de directivité en fonction de deux facteurs : la position des enceintes et la direction envisagée. Ces techniques sont souvent basées sur des modèles trigonométriques.

Les différences interaurales se réfèrent aux différences entre les sons entendus par les deux oreilles pour une seule source sonore spatiale. Les deux différences principales sont la différence interaurale de temps (ITD : *Interaural Time Difference*) et la différence interaurale d'intensité (IID : *Interaural Intensity Difference*). Par exemple, une source sonore qui se trouve à gauche de l'auditeur arrivera plus rapidement et plus intensément à l'oreille gauche par rapport au son entendu par l'oreille droite. La perte d'intensité est due à la perte de puissance que subit toute onde sonore en se propageant dans un milieu (dans ce cas l'air).

Les techniques de directivité basées sur les différences interaurales et de positionnement géométrique ne sont pas suffisamment réalistes. En général, même si elles fournissent l'information sur les variations en azimut, elles ne donnent pas d'information sur l'élévation.

Les modèles d'écoute binaurale représentent la façon dont une source sonore est affectée par l'anatomie de l'auditeur en fonction de sa position (exemple : la HRTF). Ces modèles sont décrits dans la section suivante.

¹ La diaphonie (*crosstalk*) se réfère à l'interférence entre deux signaux simultanés.

² Les différences interaurales se réfèrent aux différences entre les sons entendus par les deux oreilles pour une seule source sonore spatiale.

Il est important de distinguer, dans la spatialisation sonore, les techniques binaurales des techniques transaurales. La restitution du son spatiale directement aux oreilles (normalement à l'aide d'un casque) et en utilisant uniquement deux canaux est appelée restitution binaurale ou rendu binaural. Les systèmes multicanaux distribués autour de l'auditeur ont pour but de restituer l'écoute binaurale aux oreilles à l'aide de différents processus de filtrage acoustique sont connus sous le nom de techniques transaurales (Jot et al. [1995]). Dans le cadre de cette thèse, nous avons adopté la technique de spatialisation binaurale en utilisant un casque pour la restitution sonore.

Le terme de son spatial se réfère à la fois aux sons spatiaux naturels et artificiels. Le terme de son spatialisé se réfère à un stimulus sonore doté de propriétés spatiales. Ce processus, dit de spatialisation, peut s'effectuer principalement de deux façons : soit à l'aide des mécanismes physiques, soit en utilisant des processus de traitement acoustique. Le son 3D est finalement un terme qui peut se référer soit à un son spatialisé ou bien à l'ensemble de théories et techniques relatives à la capture, le traitement et la diffusion du son spatial.

Tout comme pour d'autres technologies interactives, l'industrie des loisirs et notamment les jeux vidéo et les films ont eu un impact important sur le développement du son 3D et des technologies associées. Actuellement, une partie importante des composants utilisés pour la spatialisation et pour la restitution de son spatial sont relativement accessibles en termes de coût et de complexité technique.

2.2.1. Spatialisation sonore et HRTF

Différents auteurs ont étudié les aspects théoriques et techniques relatifs au rendu de son 3D et son intégration en environnement virtuel et pour les interfaces homme machine. Burgess [1992] a développé une technique de spatialisation sonore à faible coût. Cette publication est importante car elle montre la possibilité d'intégrer la spatialisation sonore dans les systèmes interactifs d'une manière efficace. D'autres auteurs ont également présenté des analyses sur les aspects théoriques et techniques relatifs au rendu sonore spatial: W. Gardner [2001], Kyriakakis et al. [1999] et Lauri Savioja et al. [1999]. Ces travaux ont présenté différentes méthodes pour la capture, la simulation de la directivité et de la réverbération ainsi que pour la restitution sonore spatiale. D'autres publications ont apporté des solutions à certains problèmes précis comme Naef et al. [2002] qui ont proposé une méthode pour la restitution du son spatial en environnement immersif. Une révision plus récente sur les aspects de son spatial fut publiée par Kapralos et al. [2008]. Etant donné qu'il existe plusieurs solutions (y compris celles qui sont à faibles coûts) pour la plupart des problèmes de spatialisation du son, nous pouvons considérer que le son 3D est une technologie mature.

La Fonction de Transfert de l'Effet de l'Anatomie (HRTF : *Head-Related Transfer Function*) est un double filtre qui exprime pour chaque canal auditif, la façon dont le son émis par une source spatiale dans une position donnée est affecté par l'anatomie d'un auditeur en particulier. Begault [2005] a établi que la HRTF (ou la fonction HRTF) représente le filtrage fréquentiel qui se produit avant l'arrivée du son au tympan dans l'oreille interne. Ces filtres restituent l'effet du filtrage fréquentiel directif dû à la morphologie des pavillons sur les sons écoutés. Ils restituent également les phénomènes de diffraction et de réflexion dus principalement à la tête et les épaules de l'auditeur. La fonction HRTF qui correspond à une personne particulière et à une position donnée d'une source sonore, peut être obtenue à l'aide de mécanismes de diffusion, d'enregistrement et de traitement acoustique permettant de capturer les effets décrits.

La diffusion du rendu sonore spatialisé réalisé à l'aide de la fonction HRTF est réalisée le plus souvent à l'aide d'un casque. L'utilisation de casques permet d'éliminer pratiquement tous les effets éventuels dus à l'écoute spatiale naturelle (HRTF réelle) pour laisser la place à la HRTF appliquée. De cette manière, les ondes sonores émises par le casque ne sont que faiblement affectées par la plupart des phénomènes associés à la fonction HRTF tels que les diffractions et les réflexions dues à la tête, les différences interaurales dues à la trajectoire parcourue par le son depuis la source et le filtrage fréquentiel directif dû au pavillon. L'utilisation de casques extraauriculaires et intraauriculaires représente alors une solution acceptable pour le rendu de la fonction HRTF.

Naturellement, ces effets peuvent varier d'un sujet à l'autre obligeant les concepteurs de systèmes audio à choisir entre deux options : les fonctions HRTF individualisées ou les fonctions HRTFs généralisées. Les fonctions HRTF individualisées permettent de fournir les fonctionnalités nécessaires pour adapter le filtrage à chaque utilisateur. Cela nécessite la mesure d'une série complète de fonctions HRTF pour chaque utilisateur. Bien que l'individualisation permette une spatialisation plus fidèle, cette option n'est pas très pratique car ses exigences techniques sont relativement complexes. La deuxième option consiste à proposer un mécanisme de spatialisation généralisé qui fonctionne d'une manière acceptable pour la plupart des utilisateurs. Cette option entraîne un manque de fidélité qui peut être à l'origine de la dégradation de la spatialisation.

Les différences interaurales de temps et d'intensité sont ambiguës dans certaines conditions. Ainsi, ce modèle est affecté par un nombre important d'erreurs inverses et de cônes de confusion. L'erreur inverse se produit lorsque l'utilisateur perçoit que la source sonore se trouve dans un hémisphère opposé à celui de sa position réelle. Les erreurs inverses sont les suivantes : haut/bas, devant/derrière et gauche/droite. Un cône de confusion se réfère aux paires de sources sonores ayant les mêmes différences interaurales et une certaine symétrie en localisation. Begault [2005] a établi que la HRTF permet de réduire la présence d'erreurs inverses haut/bas et devant/derrière ainsi que les cônes de confusion.

Jiann-Rong Wu et al. [1997] ont mis en œuvre deux expérimentations pour étudier le mouvement de la tête et la latence induite pour la localisation des sources sonores en azimut. Les auteurs ont conclu que le suivi de la tête contribue à la précision de localisation et que la latence induite contribue de son côté à l'amélioration du temps d'accomplissement de la tâche. Les erreurs angulaires sont de l'ordre de 9.45 degrés environ.

E. M. Wenzel et al. [1993] ont effectué une série d'expérimentations pour évaluer la pertinence des HRTFs individualisées. Ce travail a été inspiré par les études précédentes de Frederic L Wightman et Doris J Kistler [1989a] et Frederic L Wightman et Doris J Kistler [1989b]. Les participants devaient estimer verbalement l'origine d'un son dans deux conditions distinctes : écoute naturelle en champ libre et écoute avec spatialisation sonore en utilisant un casque. Les auteurs en ont déduit que la HRTF généralisée provoque une augmentation du taux d'erreurs inverses mais permet d'obtenir une précision relativement proche de celle obtenue avec les conditions d'écoute naturelle.

Begault et al. [2001] ont développé une expérience pour comparer les effets des aspects suivants : le suivi magnétique de la tête, la réverbération, les fonctions HRTF généralisées et les fonctions HRTF individualisées. D'après ces auteurs, le suivi de la tête contribue significativement à la réduction de l'erreur de localisation et permet d'éviter la présence d'erreurs inverses. L'utilisation des fonctions HRTF généralisées induit une petite perturbation de la performance par rapport aux fonctions HRTF individualisées. Les erreurs angulaires obtenues varient entre 10 et 25 degrés environ.

Doerr et al. [2007] ont présenté un système pour la spatialisation sonore à faible coût avec retour visuel. Les auteurs ont ajouté un cône de diffusion comme indicateur supplémentaire (*supplementary cue*). Les erreurs angulaires obtenues sont comprises entre 4 et 7 degrés en azimut et entre 11 et 20 degrés en élévation. Les temps de localisation varient entre 9 et 12 secondes.

La localisation sonore peut être également étudiée en utilisant d'autres mécanismes de spatialisation différents de la HRTF. Par exemple, M. Grohn et al. [2002] ont étudié la localisation de source sonore avec une méthode de spatialisation basée sur le positionnement géométrique. Néanmoins, la HRTF est généralement préférée pour deux raisons principales : la première raison est que la HRTF est plus fidèle à l'écoute spatiale humaine comparée aux autres techniques qui existent ; la deuxième raison est que la HRTF permet de restituer les indices spectraux qui aident à déterminer l'élévation alors que la plupart des autres techniques ne fournissent pas cette information. En outre, les résultats disponibles dans la littérature actuelle confirment que la HRTF est le modèle qui permet la meilleure précision de localisation.

La spatialisation sonore peut être effectuée à l'aide de nombreuses techniques dont la HRTF. Cette dernière permet la spatialisation sonore en fournissant des informations sur l'azimut et sur l'élévation. Néanmoins, le niveau de précision de localisation de source sonore en utilisant la fonction HRTF est faible.

2.2.2. Acoustique virtuelle

Le terme clé de l'acoustique virtuelle est l'auralisation. En général, elle se réfère aux techniques de rendu audio à partir de données numériques qui peuvent être obtenues par la simulation, la capture ou la synthèse du son (Vorlander [2007]). En général, nous pouvons dire que le son 3D fait partie des techniques d'auralisation. De cette façon, l'auralisation comprend toutes les techniques pour la capture, la spatialisation et la restitution sonore de source spatiales. Cependant, l'auralisation comprend également d'autres thématiques liées à ce sujet telles que la modélisation et la synthèse des stimuli sonores, la perception sonore, les différents types d'interaction entre les ondes sonores et les matériaux et finalement la simulation des environnements acoustiques.

La recherche dans le domaine de la simulation de l'écoute spatiale en environnement virtuel nécessite des outils spécialisés. Ces outils sont développés pour être combinés avec des mécanismes traditionnels de retours visuels et haptiques. Nous citons ci-après quelques outils existants.

Le système DIVA a été conçu pour l'acoustique virtuelle. Une description complète de ce système est disponible dans J. Huopaniemi et al. [1996] et dans Takala et al. [1996]. Jot et Warusfel [1995] ont présenté le « Spatialisateur » qui est un système pour la restitution du son 3D basé sur la perception de l'auditeur plutôt que sur la description physique de l'environnement. Nelson et al. (1996) ont présenté de leur côté une validation expérimentale du système VAP (Virtual Acoustics Project) pour l'acoustique virtuelle. Slizle et al. (2004) ont présenté le système IKA-SIM qui permet de simuler l'écoute spatiale en environnement virtuel acoustique. Les caractéristiques principales de ces systèmes sont la modélisation des environnements acoustiques (simplifiés ou complexes), la spatialisation auditive (binaurale ou multicanaux) et le retour visuel.

La plateforme développée dans le cadre de cette thèse peut être considérée comme une version simplifiée des systèmes existants (IKA-SIM, DIVA, VAP, etc.). Cependant, cet outil se focalise sur les aspects

spatiaux des sources sonores. Notre plateforme permet également d'améliorer l'interaction en combinant l'audio spatialisé et le guidage auditif.

2.3. Retour sonore et son 3D pour l'interaction en environnement virtuel

Le développement de techniques d'interaction en environnement 3D a été traditionnellement concentré sur le retour visuel. Cependant, les applications qui intègrent le retour haptique en espace virtuel commencent à se développer d'une manière importante. Quant à lui, le retour sonore non spatialisé a été utilisé avec succès pour des tâches interactives dans des espaces non tridimensionnels. Ainsi, le retour sonore n'a pas été suffisamment exploité pour l'interaction 3D, comparé au retour visuel et au retour haptique.

En outre, la multimodalité émerge comme un centre d'intérêt scientifique en RV et en IHM. Elle se réfère à l'utilisation d'au moins deux retours sensoriels simultanés. L'intérêt de la communauté scientifique pour la multimodalité a permis des avancées importantes dans le développement de techniques qui intègrent en même temps la vision, l'haptique et l'audio. Ces travaux mettent en évidence l'apport de la multimodalité pour la perception de l'environnement virtuel encourageant le développement de nouvelles techniques d'interaction basées sur le son.

Cette section est consacrée à la description du contexte de l'interaction en environnement virtuel. On présente les applications du retour sonore existantes pour l'interaction non tridimensionnel. Ensuite, on définit les caractéristiques et les besoins de l'interaction dans un environnement virtuel tridimensionnel. Enfin, on décrit les différents aspects de l'utilisation des techniques de son 3D pour l'interaction en espace tridimensionnel virtuel.

2.3.1. Retour sonore non spatial en environnement virtuel

L'intégration du retour sonore non spatial dans les interfaces homme machine a été le sujet de différentes recherches scientifiques. Le retour non spatial et non parlé (non-speech) est aussi une technique relativement courante dans les systèmes interactifs. Dans cette section, nous décrivons quelques travaux portant sur ces techniques.

Kramer [1992] a introduit le concept d'affichage auditif (*auditory display*) qui se réfère à l'utilisation du son pour transmettre des informations significatives aux utilisateurs des systèmes interactifs. Pour cet auteur, les bénéfices de ce type de retour sont divers. Les signaux acoustiques sont plus rapides à détecter que les signaux visuels. Ils permettent aussi d'utiliser des stimuli sonores en parallèle et de transmettre des informations avec des mécanismes non visuels, évitant ainsi de surcharger le retour visuel.

Kramer [1992] a également identifié les principales limites des affichages auditifs. Pour certaines tâches, la faible résolution des différentes variables sonores ne permet pas de les utiliser pour transmettre des informations avec la résolution nécessaire. Une des limites qui présente un intérêt particulier pour nos travaux de recherche est le manque de précision spatiale du retour sonore. Selon l'auteur, ce manque de précision ne permet

ni de représenter efficacement des informations volumétriques ni de signaler exactement les points d'intérêt dans l'espace.

Gaver [1986] a présenté l'approche des icônes auditives qui se réfère à l'utilisation de sons dits naturels pour fournir des informations. Le principe de cette approche est que l'origine des stimuli doit être facilement reconnue et l'association avec l'information à transmettre doit être intuitive. Par exemple, la progression d'une tâche peut être représentée par le son produit par un liquide qui est versé dans un verre jusqu'à son remplissage total. De cette manière, l'utilisateur est capable de comprendre facilement le message. Un inconvénient majeur de cette approche est qu'il n'est pas toujours facile de trouver ce genre d'association. L'auteur a présenté dans un autre travail le SonicFinder (Gaver [1989]) qui est une référence importante, consistant en une interface auditive basée sur les icônes auditives.

Le concept d'Earcons fut présenté par Blattner et al. [1989]. Cette approche est basée sur des séquences musicales qui sont utilisées pour passer des messages représentant des entités, des propriétés ou des événements. Ce modèle eut un impact important sur la conception des interfaces sonores. La conception d'Earcons demande des connaissances en composition musicale. Les utilisateurs doivent apprendre la signification de chaque séquence. Le nombre de messages qui peuvent être efficacement interprétés et la période d'apprentissage peuvent varier entre différents utilisateurs.

Rocchesso et al. [2003] présentèrent leur approche d'Objets Sonores (*Sounding objects*) pour fournir l'information en applications interactives en utilisant l'audio basé sur des modèles sonores dynamiques. Cette technique comprend différents éléments tels que la modélisation physique du son, l'utilisation de mécanismes paramétrés de contrôle du son et l'application de mécanismes de validations basées sur la perception auditive. Le principe est d'associer les variations de l'information à transmettre aux paramètres correspondant à un modèle physique sonore en particulier. De cette manière, le son change en fonction des modifications de l'information. Il est nécessaire pour cette technique de créer et d'adapter des modèles de synthèse de son et de définir ses paramètres de contrôle ainsi que leurs intervalles. Pour l'approche développée dans le cadre de cette thèse, nous allons reprendre l'idée d'utiliser des paramètres de contrôle pour modifier les caractéristiques des stimuli sonores en fonction de l'interaction de l'utilisateur avec le système.

Brewster [2003] et Rocchesso [2003] ont présenté deux états de l'art des aspects théoriques et pratiques des différents modèles existants. Ces modèles ont eu un impact important sur la conception des interfaces auditives, par contre, ils n'ont pas proposé de solution pour la présentation d'information spatiale. Les techniques décrites dans ces travaux sont plutôt appropriées pour transmettre les types d'informations suivants : des signalisations sur un événement concret (par exemple, le début ou la fin d'une opération) et les changements d'une variable unidimensionnelle (par exemple, la progression d'une tâche). Un autre inconvénient important est la difficulté rencontrée pour mettre en place un mécanisme de validation basé sur des critères objectifs de performance. Il existe néanmoins des mécanismes subjectifs basés en général sur la perception auditive.

Dans cette thèse, nous proposons une approche qui permet de fournir une information spatiale et de mettre en place une évaluation basée sur des critères objectifs de performance. Cette approche est basée sur l'utilisation de techniques de spatialisation sonore avec des stimuli sonores enrichis par des effets audio paramétrés. L'intérêt de cette approche est de fournir l'information spatiale nécessaire à la réalisation d'une tâche interactive en environnement 3D.

Il existe différentes procédures d'évaluation du retour sonore : les procédures objectives et les procédures subjectives. Les procédures objectives (basée essentiellement sur la psychoacoustique) utilisent des mesures expérimentales obtenues généralement par des méthodes d'observation (automatiques ou non automatiques). Par ailleurs, les procédures subjectives (par exemple, l'évaluation de la perception) sont basées sur l'évaluation du retour d'informations de l'utilisateur obtenu à l'aide de questionnaires. Bech et Zakarov [2006] ont publié un livre sur l'évaluation de la perception auditive.

2.3.2. Interaction en environnement virtuel

L'interaction 3D se réfère à l'interaction entre l'utilisateur et les dispositifs spécialisés dont l'espace tridimensionnel est important. L'interaction 3D est liée à la RV et aux IHM. Dans cette section, nous allons présenter les concepts et les résultats de recherches les plus importants sur l'interaction 3D.

Il existe différentes techniques pour l'interaction 3D basées sur différents mécanismes de retour sensoriel. Ces techniques furent classées par Bowman (1998) comme les tâches canoniques suivantes : la sélection, la manipulation, la navigation et le système de commande. La plupart de ces techniques sont basées sur le retour visuel.

Foley et Van Dam (1982) ont présenté un livre de référence sur l'interaction avec les ordinateurs. Ce travail est restreint aux interfaces 2D et il discutait déjà une bonne partie des considérations qui sont aujourd'hui utilisées dans la conception des interfaces interactives.

Bowman et al. (2005) ont établi que la manipulation est une tâche fondamentale dans les environnements réels et virtuels. Selon les auteurs, la plupart des techniques de navigation et de système de commande sont basées sur les techniques de manipulation. La plupart des techniques d'interaction sont restreintes aux objets rigides qui ne changent pas de forme. Par ailleurs, la manipulation et la sélection sont souvent traitées ensemble. Pour quelques auteurs, la sélection représente une sous-tâche de la manipulation.

D'après les recherches de Bowman et al. (2005), les sous-tâches principales de la manipulation sont les suivantes : la sélection, le positionnement et la rotation. Le champ d'application de telles techniques est ample. Les auteurs ont décrit une série de techniques de manipulation existantes et valides. Selon eux, il n'est pas viable de concevoir de nouvelles techniques à moins que le bénéfice potentiel soit important.

2.3.3. Son 3D pour l'interaction

Le son 3D a été traditionnellement utilisé pour simuler des environnements acoustiques tridimensionnels. Néanmoins, les travaux portant sur l'utilisation du son 3D pour fournir des informations utiles à l'accomplissement d'une tâche interactive sont rares. Dans cette section, on présente une sélection de ces travaux et on discute du potentiel du son 3D.

Les techniques existantes de spatialisations sonore manquent de précision et elles ne permettent pas de fournir des informations suffisamment précises pour permettre la manipulation 3D. De ce fait, il est nécessaire de développer une nouvelle technique qui permet une localisation précise de source sonore à l'aide d'indices spatiaux (*spatial cues*) supplémentaires.

2.3.3.1. Applications existantes du retour sonore 3D

Les travaux de recherche publiés sur l'utilisation de son 3D dans les applications interactives sont rares et la plupart sont récents. Le son 3D peut être utilisé dans les systèmes interactifs pour améliorer la perception en environnement immersif, pour diffuser des alertes spatiales, pour permettre la détection des cibles ou pour assister l'utilisateur dans la manipulation et la navigation en environnement virtuel. La Figure 2.3 montre quelques applications potentielles et existantes du son 3D pour l'interaction en environnement virtuel.

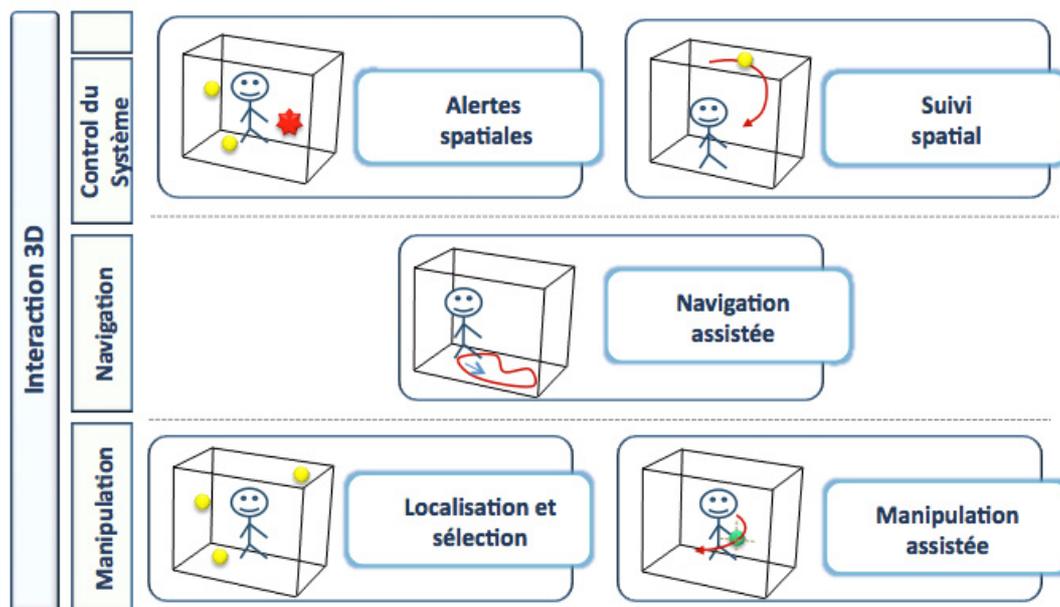


Figure 2.3. Différentes applications du son 3D pour l'interaction 3D.

E. M. Wenzel [1994] a présenté une discussion sur les avantages de l'utilisation de la représentation spatiale du son en termes de la performance de l'utilisateur. Selon l'auteur, la représentation spatiale permet d'améliorer et d'étendre la conscience situationnelle et la représentation multicanale. La conscience situationnelle se réfère à différents éléments : la représentation directe des données spatiales, l'amélioration des sensations de présence et de réalisme, la possibilité de suivre des événements dans toutes les directions et la possibilité de compléter ou de remplacer d'autres modalités. La représentation multicanale efficace permet de distinguer, d'interpréter et de sélectionner différentes sources sonores diffusées simultanément.

Cohen et Elizabeth M Wenzel [1995] ont publié une analyse sur l'utilisation du son spatial pour les interfaces interactives. Ils ont décrit le rôle potentiel du son 3D dans l'interaction multi-sensorielle en présentant un état de l'art intéressant sur ses applications existantes. Les auteurs ont proposé que la difficulté à localiser la direction et la profondeur des sources sonores peut être résolue en utilisant une combinaison d'information visuelle et auditive. Le problème de cette approche est que l'efficacité du retour sonore devient dépendante du retour visuel rendant complexe la détermination de la contribution du son.

Il existe des travaux de recherche sur le rôle du son 3D pour améliorer la perception. Hendrix et Barfield [1995] ont présenté une étude sur l'effet d'une sélection d'indices sonores et visuels dans la sensation de présence en environnement virtuel. Les auteurs ont montré que le son spatial contribue significativement à la sensation de présence mais qu'il n'a pas d'effet significatif sur le réalisme. Larsson et al. [2008] ont étudié l'effet de la directivité et de la réverbération pour les sources sonores qui se déplacent dans l'espace. Les auteurs ont conclu que la réverbération contribue à la sensation de présence. Ils ont également conclu que la contribution d'une source sonore spatiale dans la perception de l'utilisateur dépend plus du réalisme du stimulus sonore que de la qualité de la procédure de spatialisation utilisée.

L'effet du son 3D pour la détection spatiale de cibles fut étudié par McKinley et al. [1995]. Ce travail a permis de déterminer qu'il existait une contribution significative du son 3D dans la détection de cibles. Bronkhorst et al. [1996] ont étudié l'effet du son spatial pour le suivi de cibles volantes dans un simulateur de vol. Ils ont trouvé que le son 3D est approprié pour compléter et éventuellement remplacer des indicateurs visuels comme le radar. Ho et Spence [2005] ont effectué des travaux de recherche sur les avantages de l'utilisation du son spatial dans des situations potentiellement dangereuses de conduite automobile. D'après leurs résultats, le son 3D est un mécanisme efficace pour attirer l'attention de l'utilisateur et plus approprié que les alertes sonores non spatialisées.

McKinley et al. (1995) ont étudié le couplage de retours visuels et auditifs pour la tâche d'identification de cibles volantes. Les auteurs ont étudié une série de différentes combinaisons de retours immersifs visuels avec différents types de retours auditifs spatialisés. Ils ont découvert que le son spatial apporte une contribution dans toutes les conditions observées et en particulier lorsque le champ visuel est limité.

L'effet du son 3D dans la navigation en environnement virtuel a été étudié dans différents travaux de recherche. Sánchez et Sáenz [2005] ont proposé un système de son 3D pour naviguer dans un environnement virtuel et pour interagir avec ses éléments. Tapio Lokki et Matti Grohn [2005] ont évalué la contribution du son 3D pour la navigation en environnement immersif. Ils ont estimé que le son tridimensionnel, lorsqu'il se combine avec le retour visuel, apporte une amélioration significative de la performance de l'utilisateur dans la navigation. Walker et Lindsay [2006] ont montré que l'utilisation du son spatial contribue significativement à l'amélioration de la précision et de la rapidité dans la navigation en environnement virtuel.

Il existe quelques travaux de recherche sur l'application du son spatial avec la diffusion simultanée de multiples sources sonores. Kan et al. [2004] ont présenté un système qui permet de diffuser des sons directionnels pour des applications de communication. Vazquez-Alvarez et Brewster [2009] ont publié une évaluation sur le potentiel du son 3D pour diffuser des sources sonores multiples et simultanées dans la technologie mobile courante. Ils ont conclu que les sujets sont capables de distinguer jusqu'à cinq sources sonores simultanément.

Le son 3D pour la représentation de données complexes ou multidimensionnelles a été très peu étudié. Ménelas et al. [2009] ont présenté une approche pour enrichir l'information fournie par la HRTF dans le cadre de l'interaction multimodale. Il est à noter que la plupart des techniques de représentation sonore de données multidimensionnelles sont basées sur le son non spatial (par exemple, Shelley et al. [2009]).

En conclusion, les résultats des travaux de recherche décrits montrent qu'en général le son 3D contribue à la sensation de présence mais non pas au réalisme. Il existe des applications réussies utilisant le son 3D pour

l'interaction en particulier pour les fonctions suivantes : le suivi et les alertes spatiales, la navigation et la diffusion de multiples sources simultanées. Néanmoins, la contribution du son 3D n'a pas encore été démontrée pour la manipulation d'objets en environnement virtuel.

2.3.4. Approches existantes pour améliorer la localisation de sources sonores en environnement virtuel

Le système auditif humain a une faible résolution pour la localisation de sources sonores (N. I. Durlach et al. [1993]). La précision de la localisation de sources virtuelles est normalement affectée par les caractéristiques du modèle de spatialisation employé. Il existe différentes approches pour améliorer la précision de la localisation de sources sonores dans un environnement virtuel. Nous décrivons dans les paragraphes suivants certaines d'entre elles.

La localisation auditive supranormale est une approche proposée par N. I. Durlach et al. [1993]. Elle est basée sur la modification contrôlée de la valeur du paramètre d'azimut de la HRTF employée en utilisant une fonction d'association (*mapping*). La suite de ce travail a été publiée dans Barbara G. Shinn-Cunningham et al. [1998a] et Barbara G. Shinn-Cunningham et al. [1998b]. Ce travail est restreint aux sources sonores localisées dans le plan horizontal.

En M. Grohn et Takala [1994] et M. Grohn et Takala [1995] les auteurs ont présenté le projet nommé MagicMikes traitant de la sonification³ de données spatiales dans un contexte de navigation en environnement virtuel. Cette approche consiste à l'utilisation combinée du guidage auditif et de la spatialisation sonore. M. Grohn et al. [2002] ont présenté une étude pour évaluer la localisation sonore en environnement immersif avec un système multicanal. Les erreurs de localisation obtenues varient entre 6 et 8 degrés pour l'azimut et entre 12 et 25 degrés pour l'élévation.

Marentakis et Brewster [2005] ont mis en place une expérimentation pour comparer différentes indications sonores afin d'améliorer l'efficacité de la localisation des sources sonores. Les auteurs ont testé différentes sources sonores localisées dans le plan horizontal et ont obtenu des erreurs angulaires comprises entre 4 et 10 degrés.

Tapio Lokki et Matti Grohn [2005] ont présenté une étude sur l'effet du guidage audio dans la performance pour la navigation en environnement virtuel. L'évaluation de la performance est basée sur les mesures de temps d'exécution et la longueur des trajectoires entre autres mais la précision angulaire n'a pas été prise en compte. Les auteurs ont trouvé que le guidage audio a un effet significatif sur la performance et qu'il est possible de naviguer en utilisant seulement le retour auditif. Dans cette étude, un groupe d'indices sonores comme le gain et la hauteur musicale (*pitch*) a été testé. Ce travail a pris en compte la combinaison d'indices paramétrés et la spatialisation sonore pour améliorer la précision de localisation en élévation. Ce travail représente une contribution importante dans le domaine du son 3D. Cependant, les résultats obtenus ne permettent pas d'évaluer la précision dans la localisation des sources sonores.

³ La sonification se réfère à l'utilisation du retour audio non parlé pour transmettre une certaine information qui n'est pas auditive à l'origine.

La plupart des techniques existantes visent à améliorer la précision de localisation de sources sonores situées dans le plan horizontal. Par contre, l'objectif de notre travail est d'améliorer la précision de localisation de source sonore en azimut et en élévation.

2.3.5. Application potentielle du son pour la manipulation

Il n'existe pas de technique de manipulation 3D efficace basée sur le retour sonore. L'existence d'une telle technique aura les avantages suivants :

- Manipulation d'objets lorsque le retour visuel est perturbé ou indirect.
- Manipulation d'objets lorsque le retour visuel est partiellement ou complètement indisponible.
- Amélioration de la performance de certaines techniques de manipulation existantes basées sur d'autres retours sensoriels.
- Remplacement d'autres retours sensoriels comme la vision et le retour haptique qui sont techniquement complexes et nécessitent des équipements onéreux.
- Utilisation dans les nouvelles techniques de navigation et dans les systèmes de commande ainsi que pour la réalisation de tâches spécifiques : par exemple dans le suivi et la détection de cibles spatiales et l'exploration de données multidimensionnelles.

Les limites les plus importantes de l'écoute spatiale sont le manque de précision et l'ambiguïté inhérente (présence d'erreurs inverses et cônes de confusion). L'être humain n'est pas très précis dans la localisation de sources sonores. Il est significativement plus précis dans la localisation en azimut qu'en élévation. L'erreur absolue globale de localisation varie selon les différents auteurs entre 10 et 25 degrés environ et la dispersion des données est importante. Ces erreurs et leurs dispersions sont évidemment trop importantes pour transmettre des informations spatiales précises.

Le fait que l'écoute spatiale dépende des caractéristiques anatomiques de chaque individu peut représenter un inconvénient majeur pour les systèmes de son 3D. Dans ce contexte, nous proposons une technique basée sur des méthodes généralisées.

Les problèmes d'ambiguïté représentent aussi un obstacle pour utiliser le son pour l'interaction en environnement virtuel. Il n'est pas convenable que l'information sonore spatiale soit ambiguë car ceci peut affecter d'une manière importante la performance de l'utilisateur. Il est donc nécessaire de développer une technique sans ambiguïté.

L'identification de la provenance de source sonore par l'auditeur peut comprendre la réalisation d'un geste de mouvement (*localization by movement*) pour faciliter le processus. Marentakis et Brewster [2004] ont exploré différents gestes pour interagir avec des interfaces audio spatialisées dans la localisation de sources sonores. Les gestes considérés sont le pointage avec la main, le pointage avec la tête et l'indication avec une interface visuelle. Les auteurs ont défini l'erreur de localisation de chaque geste et ils ont conclu que les trois gestes peuvent être utilisés dans les interfaces auditives.

Les aspects discutés dans ce chapitre seront récapitulés dans les paragraphes suivants pour expliquer le contexte dans lequel notre approche a été proposée.

Nous avons envisagé de créer une technique qui permet la manipulation 3D basée sur le retour sonore. Les techniques existantes de guidage sonore sont particulièrement efficaces pour certaines tâches mais elles ne permettent pas de fournir des informations spatiales. Les technologies existantes de son 3D sont matures mais l'information spatiale qu'elles fournissent n'est pas suffisamment précise et elle présente des ambiguïtés. Nous envisageons de proposer une technique basée sur le guidage audio et le son 3D pour permettre la manipulation 3D en fournissant une information spatiale précise et sans ambiguïté.

Cette technique doit satisfaire certains critères. Il faut qu'elle soit intuitive, facile à utiliser et rapidement apprise par l'utilisateur. Il faut qu'elle permette également un bon niveau de performance en termes de précision et de temps d'accomplissement de la tâche. Enfin, il est nécessaire qu'elle réduise significativement la présence d'ambiguïtés dans la localisation de sources sonores.

Cette technique doit permettre une bonne précision pour fournir des informations spatiales utiles pour la manipulation d'objets en environnement 3D. Par contre, il n'est pas évident de déterminer ce niveau de précision puisqu'il dépend de l'application à étudier. Il a donc été décidé de fixer ce niveau en fonction des limites connues pour la localisation avec les techniques de spatialisation courantes. La précision moyenne en azimut avec les techniques existantes peut atteindre une valeur entre 1 et 3 degrés. Nous chercherons à réduire cette variabilité ou cette dispersion de données. Dans le cas de l'élévation, la meilleure précision moyenne obtenue se trouve autour de 10 degrés. Ainsi, il est nécessaire de réduire en particulier l'erreur en élévation. Par conséquent, notre objectif est d'obtenir une précision en élévation équivalente à celle obtenue en azimut avec les techniques existantes. L'erreur angulaire globale (azimut et élévation combinés) doit être de l'ordre de trois degrés obtenue pour l'azimut en utilisant ces mêmes techniques.

3. Techniques et outils employés

Ce chapitre a pour but de décrire les concepts, les techniques et les outils utilisés dans le cadre de cette thèse. Nous présentons tout d'abord les principales techniques de traitement de signal utilisées dans nos travaux. Ensuite, les fondements de la Fonction de Transfert de l'Effet de l'Anatomie (HRTF) sont décrits. Puis, on présente les concepts, les modèles et les méthodes statistiques utilisés dans l'analyse des résultats expérimentaux. Finalement, on décrit la plateforme expérimentale développée pour la réalisation des expérimentations.

3.1. Concepts et techniques de traitement de signal

Les modèles de spatialisation sonore impliquent l'utilisation de différentes techniques de traitement de signal. Pour faciliter la compréhension de l'implantation réalisée dans le cadre de cette thèse, nous allons décrire les techniques les plus importantes. Une description exhaustive de ces techniques a été publiée par Smith [1999], Oppenheim et al. [1999] et Proakis et Manolakis [2007].

Un signal discret $\{x[n]\}$ composé de N éléments est représenté comme suit :

$$\{x[n]\} = \{x[0], x[1], \dots, x[N-1]\} \quad (3.1)$$

où $x[n] \in \mathbb{R}$.

Le signal $\{x[n]\}$ est souvent représenté par $x[n]$ uniquement.

Transformée Discrète de Fourier

La Transformée Discrète de Fourier (TDF) est un outil mathématique permettant d'obtenir la représentation spectrale (dans le domaine de fréquences) d'un signal discret périodique. La théorie de cet outil est basée sur l'Analyse de Fourier.

Etant donné un signal discret $x[n]$, sa représentation spectrale $X[k]$ obtenue par la Transformée de Fourier est définie comme suit :

$$X[k] = \sum_{n=0}^{N-1} x[n] \cdot e^{-i2\pi \frac{kn}{N}} \quad (3.2)$$

où N est la longueur de $x[n]$ (ceci se représente par $|x[n]| = N$),

$$x[n] \in \mathbb{R},$$

les éléments de $X[k]$ sont des nombres complexes ($X[k] \in \mathbb{C}$), et

$$|X[k]| = N/2.$$

Transformée Inverse Discrète de Fourier

La Transformée Inverse Discrète de Fourier (TIDF) permet d'obtenir un signal discret dans le domaine temporel à partir de sa représentation spectrale. En considérant une série de nombres complexes $X[k]$, le signal discret correspondant peut être obtenu à partir de l'expression suivante :

$$x[n] = \frac{1}{N} \sum_{k=0}^{N/2} X[k] \cdot e^{i2\pi \frac{kn}{N}} \quad (3.3)$$

où $n = 0, 1, \dots, N-1$,

$$x[n] \in \mathbb{R}, \text{ et}$$

$$X[k] \in \mathbb{C}.$$

Transformée Rapide de Fourier

Le terme de Transformée Rapide de Fourier TRF se réfère à un algorithme de calcul de la TDF permettant de réduire le nombre d'opérations nécessaires. Il existe différentes implantations de la TRF. Nous avons utilisé une TRF basée sur l'algorithme de Cooley et Tukey (Cooley et Tukey [1965]). Cette méthode permet de réduire la complexité temporelle du calcul de la TDF de $O(n^2)$ à $O(n \log n)$.

Cet algorithme optimise le calcul en se basant sur le principe de la division et de la recombinaison. Ce principe se réfère à la séparation de l'entrée d'une manière convenable et à la composition de la sortie à partir d'un traitement appliqué lors de différentes étapes successives. Une description complète de cet algorithme a été publiée dans Smith [1999].

L'application de la TRF à un signal $x[n]$ pour obtenir sa représentation spectrale $X[k]$ peut être représentée par l'expression suivante :

$$X[k] = TRF(x[n]) \quad (3.4)$$

où $|x[n]| = N$, $|X[k]| = N/2$,

$$x[n] \in \mathbb{R} \text{ et}$$

$$X[k] \in \mathbb{C}.$$

De la même manière, la Transformée Inverse Rapide de Fourier (TIRF) d'un spectre pour obtenir le signal correspondant dans le domaine temporel peut être représentée comme suit :

$$x[n] = TIRF(X[k]) \quad (3.5)$$

où $|x[n]| = N$ et

$$|X[k]| = N/2.$$

Convolution Discrète

La convolution est une méthode mathématique qui permet de combiner deux fonctions. Dans sa version discrète elle permet de combiner deux signaux discrets. La convolution discrète compte différentes applications dans divers domaines comme l'électronique, la statistique et plus particulièrement le traitement de signal.

La convolution $y[n]$ de deux signaux discrets $h[n]$ et $x[n]$ est dénotée comme suit :

$$y[n] = h[n] * x[n] \quad (3.6)$$

où $|x[n]| = N$, $|h[n]| = L$ et $|y[n]| = N + L - 1$.

En mathématiques, la convolution discrète est définie comme suit :

$$h[n] * x[n] = (h * x)[n] = \sum_{m=0}^{\text{Min}(L-1, n)} h[m] \cdot x[n-m] \quad (3.7)$$

où $n = 0, 1, \dots, N-1$ et

$\text{Min}(L-1, n)$ représente la valeur minimale entre $L-1$ et n .

L'opérateur \cdot représente le produit arithmétique. La complexité de cette opération est dénotée par $O(nl)$ qui devient $O(n^2)$ si $n = l$.

Réponse impulsionnelle et convolution

Une réponse impulsionnelle d'un système est le signal sortant lorsque l'entrée est un signal court ou en forme d'impulsion (Figure 3.1a). La réponse impulsionnelle sert à représenter la réponse d'un système linéaire invariant dans le temps.

Parmi d'autres applications dans le traitement de signal, la convolution est utilisée pour appliquer une réponse impulsionnelle à un signal donné. De cette façon, la convolution entre un signal $x[n]$ et la réponse impulsionnelle $h[n]$ d'un système Z est équivalente à la sortie $y[n]$ du système Z lorsque l'entrée est le signal $x[n]$ (Figure 3.1b).

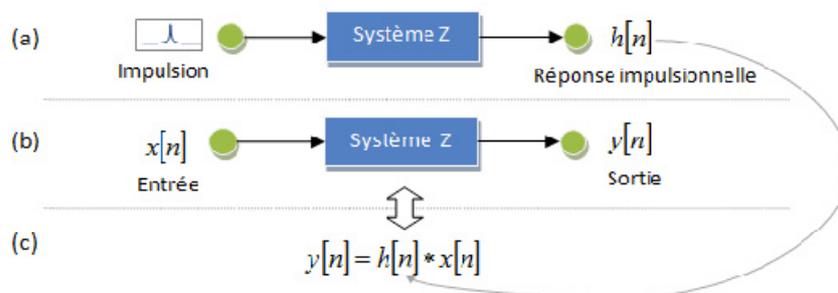


Figure 3.1. La convolution obtenue par l'application de la réponse impulsionnelle d'un système.

Signaux allongés

Un signal allongé (*zero pad*) se réfère à l'extension d'un signal discret à une longueur convenable en rajoutant des zéros à la fin. Les signaux allongés sont utiles pour effectuer différentes opérations de traitement de signal. Nous utilisons ces signaux allongés pour traiter les signaux audio et les réponses impulsionnelles associées à l'écoute binaurale.

Nous décrivons ci-après les types de signaux allongés que nous utiliserons dans notre implantation. Le signal allongé $x'[n]$ d'un signal discret $x[n]$ est défini alors comme suit :

$$\{x'[n]\} = \left\{ x[0], x[1], \dots, x[n-1], \underbrace{0, \dots, 0}_N \right\} \quad (3.8)$$

où $|x[n]| = N$ et

$$|x'[n]| = 2N.$$

Le signal allongé $h'[n]$ d'une réponse impulsionnelle $h[n]$ est défini alors comme suit :

$$\{h'[n]\} = \left\{ h[0], h[1], \dots, h[L], \underbrace{0, \dots, 0}_{L-1} \right\} \quad (3.9)$$

où $|h[n]| = L$ et

$$|h'[n]| = 2L - 1.$$

Obtention de la convolution discrète via la TRF et la TIRF

La multiplication des éléments des spectres de deux signaux discrets obtenus par la TDF est équivalente à la convolution de ces signaux (Oppenheim et al. [1999]). Ceci permet d'optimiser le calcul de la convolution avec une complexité gouvernée par l'expression $O(n \log n)$.

La convolution $y[n]$ entre un signal discret $x[n]$ et une réponse impulsionnelle $h[n]$ est représentée par :

$$y[n] = h[n] * x[n] \quad (3.10)$$

où $|x[n]| = N$, et

$$|h[n]| = N + 1.$$

Son spectre $Y[k]$ peut être obtenu par la multiplication des spectres des signaux allongés correspondants représentée comme suit :

$$Y[k] = H'[k] \times X'[k] \quad (3.11)$$

où

$$X'[k] = TRF(x'[n]) \quad (3.12)$$

$$H'[k] = TRF(h'[n]) \quad (3.13)$$

$$y[n] = TIRF(Y[k]) \quad (3.14)$$

\times représente la multiplication élément par élément des deux signaux et

$x'[n]$ et $h'[n]$ sont les signaux allongés de $x[n]$ et $h[n]$ respectivement.

La convolution entre le signal $x[n]$ et la réponse impulsionnelle $h[n]$ est alors définie comme suit :

$$y[n] = h[n] * x[n] = TIRF(H'[k] \times X'[k]) \quad (3.15)$$

où $|y[n]| = 2N$.

La Figure 3.2 montre le processus d'obtention de la convolution avec la Transformée Rapide de Fourier et la multiplication des spectres.

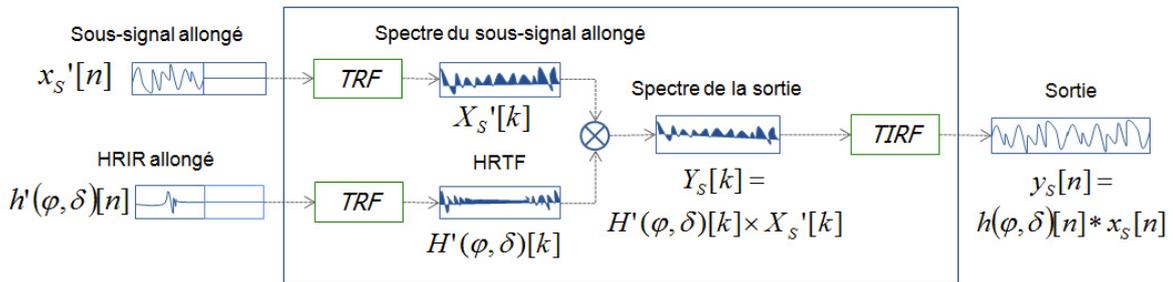


Figure 3.2. La convolution discrète obtenue par la Transformée Rapide de Fourier.

Méthode de Superposition et d'Addition

La méthode de superposition et d'addition (*overlap and add*) permet d'appliquer une réponse impulsionnelle (filtre) à un signal discret de longueur supérieure. Cette méthode est basée sur les étapes suivantes : la fragmentation du signal d'entrée en sous-signaux de la même taille, la génération des signaux allongés correspondants à ces sous-signaux, le traitement des sous-signaux allongés et la composition du signal de sortie en rassemblant les sous-signaux allongés traités. Ci-après, nous allons décrire les principaux concepts de cette technique.

Obtention de sous-signaux

Pour appliquer la méthode de superposition et d'addition entre un signal $x[n]$ et un filtre $h[n]$ de longueurs M et L respectivement où $M \gg L$, il est nécessaire de diviser le signal $x[n]$ en sous-signaux de longueur $N = L - 1$. Ces sous-signaux sont alors définis comme suit :

$$x_s[n] \stackrel{def}{=} x[S \times N + n] \quad (3.16)$$

où $|x_s[n]| = N$ et

et l'index S dénote le $S^{\text{ème}}$ sous-signal.

Les sous-signaux allongés de $h[n]$ et de chaque $x_s[n]$ sont dénotés par $h'[n]$ $x'_s[n]$ respectivement où $|h'[n]| = |x'_s[n]| = 2N$.

La convolution des sous-signaux

La convolution entre chaque sous-signal $x_s[n]$ et le filtre $h[n]$ est dénotée par :

$$y_s[n] = h[n] * x_s[n] \quad (3.17)$$

où $|y_s[n]| = 2N$.

Ceci peut être aussi obtenu par la formule suivante :

$$y_s[n] = TIRF(H'[k] \times X'_s[k]) \quad (3.18)$$

Cette formule permet d'optimiser le calcul via la TRF et la TIRF comme il a été expliqué précédemment.

Superposition et addition

Le signal de sortie $y[n]$ formé par la superposition et l'addition des sous-signaux traités est défini comme suit :

$$\begin{aligned} & \text{si } n < N \text{ ou } n > M & y[n] &= y'_q[r] \\ & \text{si } N \leq n \leq M & y[n] &= y'_q[r] + y'_{q-1}[N+r] \end{aligned} \quad (3.19)$$

où $n = 0, 1, \dots, (M + N - 1)$,

$q = n \text{ div } N$, et

$r = n \text{ mod } N$.

Les operateurs *div* et *mod* représentent la division entière et le résidu de la division respectivement. La méthode complète de superposition et d'addition est illustrée dans la Figure 3.3.

3.2. La fonction HRTF

3.2.1. Les fondements de la fonction HRTF

La fonction HRTF (ou la HRTF tout simplement) est une technique de restitution de l'écoute binaurale basée sur les étapes suivantes: la capture de l'effet de l'écoute spatiale, la création des filtres de transfert et finalement, l'application de ces filtres à un son. Cette technique est nommée d'après les Fonctions de Transfert de l'Effet de l'Anatomie (HRTF : *Head-Related Transfer Function*) qu'elle utilise.

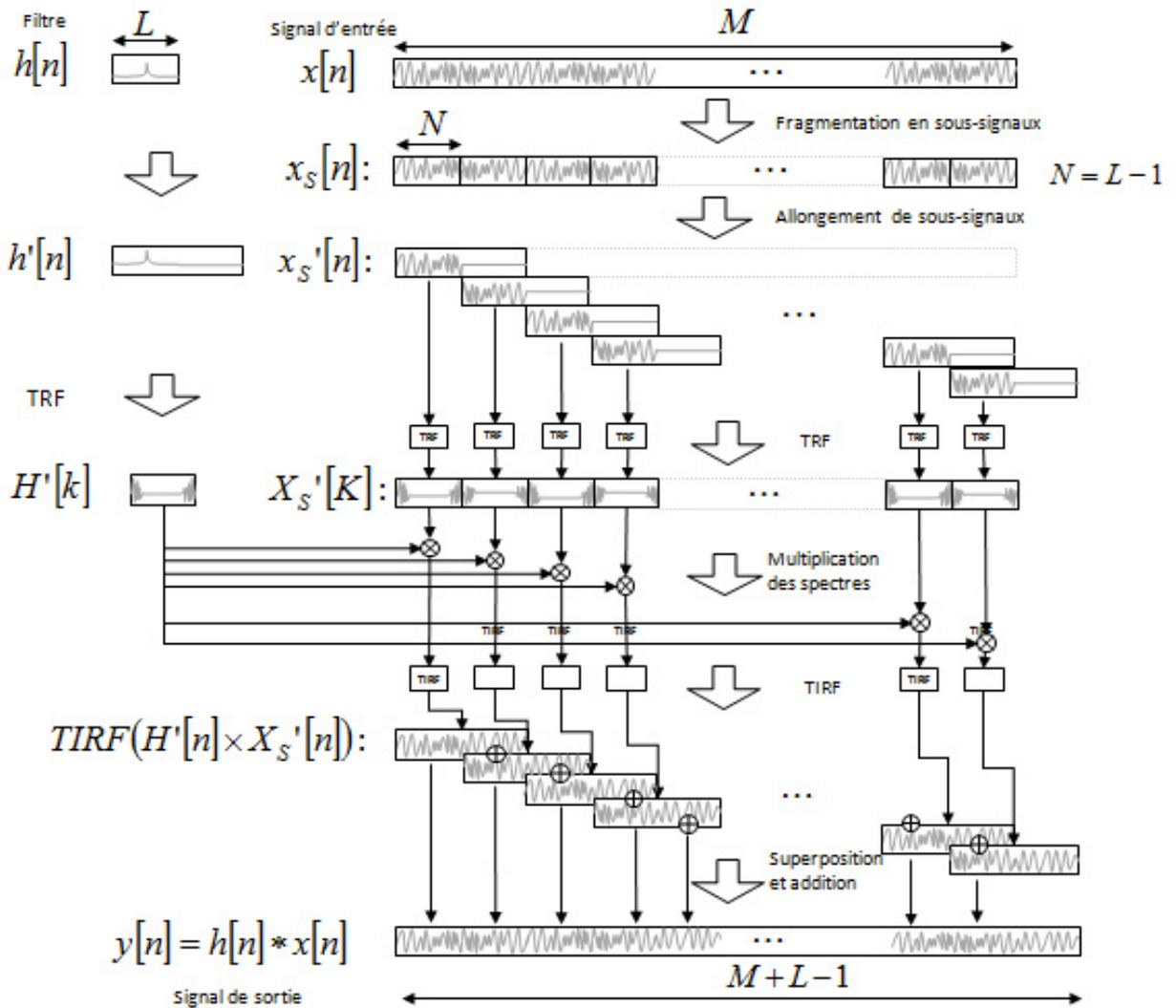


Figure 3.3. Diagramme de la méthode complète de superposition et d'addition pour l'application d'une réponse impulsionnelle à un signal discret long.

La capture se réfère à l'enregistrement des sons provenant d'une source spatiale donnée à l'entrée des canaux auditifs d'un individu. En général, le stimulus sonore utilisé a la forme d'une impulsion. Le signal obtenu à partir de ce processus est connu comme la Réponse Impulsionnelle de l'Anatomie de l'auditeur (HRIR : *Head-Related Impulse Response*). Ce signal représente principalement l'effet de l'anatomie de la tête et des oreilles dans l'écoute d'un stimulus sonore provenant d'une position spatiale donnée. Pour un individu en particulier, dans une position donnée et dans un environnement avec des propriétés acoustiques invariantes dans le temps, la forme de la HRIR dépend de la localisation de la source sonore.

Dans le processus de spatialisation sonore, le spectre d'une HRIR est calculé pour obtenir un filtre fréquentiel. Ce filtre est connu comme la Fonction de Transfert de l'Effet de l'Anatomie de l'auditeur (connu

également comme HRTF) qui permet de transférer à un signal audio les propriétés spatiales de la position d'une source sonore.

La technique de spatialisation d'un stimulus sonore pour une position donnée nécessite l'application de la HRIR correspondante à cette position. Lorsque la HRIR de la position désirée n'est pas disponible, les solutions courantes considèrent soit l'utilisation de la HRIR la plus proche soit le calcul d'une HRIR approximative en effectuant une interpolation des HRIRs les plus proches. Il est nécessaire de capturer les HRIRs d'un nombre suffisant de positions différentes pour avoir une bonne représentation des directions de provenance possibles de la source sonore virtuelle dans l'espace.

Les HRIRs d'une base de données sont normalement prises dans des intervalles réguliers d'élévation et d'azimut. Il est également courant que toutes les HRIRs soient prises à la même distance entre l'auditeur et la source sonore (en général 1.5 mètres environ).

L'utilisation de la HRTF dans des applications interactives nécessite un ensemble de processus en ligne. Il existe différentes approches pour optimiser le coût du calcul (Kapralos et Mekuz [2007]). R O Duda [1993] a décrit les besoins pour la modélisation efficace de la HRTF, dans la même publication, il a classé les différentes approches existantes pour la modélisation de la HRTF.

Brown et R. O Duda [1997] ont proposé des modèles mathématiques pour la simulation des effets de la tête et des pavillons de l'auditeur ainsi que de l'acoustique de la chambre dans laquelle le sujet se trouve. Les erreurs de localisation obtenues dans la validation expérimentale sont 12 degrés environ. Les auteurs ont conclu que, dans ces modèles, l'individualisation est importante pour l'élévation uniquement.

3.2.2. Individualisation et généralisation des fonctions HRTF

Il existe deux façons différentes pour implanter la technique de la HRTF : l'individualisation et la généralisation. L'individualisation se réfère aux processus de spatialisation basés sur des HRTFs mesurées ou adaptées à chaque utilisateur. La généralisation implique que la même HRTF est utilisée pour spatialiser les sources sonores pour tous les utilisateurs. Une HRTF généralisée peut être obtenue en traitant les HRTFs d'un ou plusieurs utilisateurs.

E. M. Wenzel et al. [1993] ont réalisé une expérience pour comparer la localisation de sources sonores dans l'écoute naturelle et en utilisant des HRTFs individualisées. Ils ont trouvé des résultats similaires dans les deux cas.

Il existe différentes techniques pour améliorer la perception obtenue avec des HRTFs généralisées. Par exemple, Martens [2003] a présenté une évaluation de la perception comparative des HRTFs généralisées et des HRTFs individualisées. L'auteur a proposé une solution qui prend en compte les différences individuelles dans la perception, elle consiste en la calibration basée sur des critères de perception des HRTFs individualisées. Selon l'auteur, cette solution permet d'éviter de mesurer un ensemble complet des HRTFs. L'inconvénient de cette approche est qu'elle nécessite d'adapter la fonction HRTF à chaque utilisateur.

L'utilisation des HRTFs généralisées n'est pas un choix technique très pratique pour les applications interactives. Il est nécessaire de mesurer la HRTF de chaque utilisateur, ce qui implique un temps de traitement important. De plus, la prise de HRTF individualisées nécessite des installations et équipements spécialisés. Par

conséquent, la majorité des systèmes de spatialisation sonore basés sur la HRTF utilisent des fonctions généralisées. Dans les travaux développés dans cette thèse, nous avons décidé d'utiliser des HRTFs généralisées pour éviter les problèmes liés à l'individualisation.

3.2.3. Base de données HRTF

Une base de données HRTF se compose d'un ensemble de HRIRs. Actuellement, il existe différentes bases de données de HRIRs disponibles. Elles se différencient principalement dans la résolution, le nombre de positions de source sonore mesurées, le nombre d'auditeurs disponibles et les mécanismes de capture utilisés.

Bill Gardner et K. Martin [1994] ont publié la base de données de KEMAR HRTF créée au MIT Media Lab. Cette base de données est composée d'un ensemble de HRIRs mesurées à partir de la tête artificielle KEMAR. Cette tête artificielle est produite par la société GRAS Sound & Vibration.

Dans le cadre de cette thèse, nous avons utilisé la base de données de KEMAR HRTF dans sa version compacte, considérée parmi les premières bases de données de HRIRs disponibles. Elle représente une référence importante dans le domaine, ses avantages et ses limites sont identifiés. Les HRIRs de la version compacte possèdent 128 points chacune. Les positions mesurées représentent un ensemble de 710 prises. Elles ont été mesurées successivement avec des pas de 6 degrés environ en azimut et 10 degrés en élévation. L'intervalle en azimut est de 360° ($-180^\circ, +180^\circ$), Il est de 130° ($-10^\circ, +90^\circ$) en élévation.

Il existe d'autres bases de données disponibles pour les applications scientifiques et commerciales. Pralong et Simon. Carlile [1994] ont créé une librairie HRTF au Laboratoire de Neurosciences Auditives comprenant environ 50 auditeurs. Algazi et al. [2001] ont présenté la base de données CIPIC qui comprend les HRTFs mesurées à partir de 45 participants y compris la tête artificielle KEMAR. Warusfel [2003] a publié la base de données de Listen HRTF créée au laboratoire IRCAM. Cette base de données contient également les mesures de 50 utilisateurs environ.

MacDonald et Tran [2008] ont réalisé une expérience pour évaluer quatre différentes bases de données de HRTFs y compris la KEMAR HRTF et le CIPIC. Les quatre bases de données ont été créées à partir des mesures de la tête artificielle KEMAR à l'aide de différents mécanismes de capture et de traitement. Les auteurs ont trouvé que, malgré la différence entre les techniques employées pour créer ces bases de données, ces dernières permettent des précisions de localisation sonore similaires. Ils ont conclu que les bases de données de HRTFs généralisées ne présentent pas de différences significatives et qu'elles peuvent être utilisées indifféremment. Ce résultat nous permet de considérer que la base de données de HRTFs utilisée dans les expérimentations présentées dans cette thèse n'aura pas d'effet significatif dans la localisation de sources sonores.

3.2.4. Modèle de la fonction HRTF

En général, le modèle de la HRTF est simplifié pour se restreindre aux sources sonores se trouvant toutes à la même distance de l'auditeur. Par conséquent, dans la base de données utilisée, les HRIRs varient seulement en fonction de l'azimut et l'élévation alors que la distance entre l'auditeur et la source sonore reste fixe.

La Figure 3.4 illustre la HRTF correspondante à une source sonore spatiale. Ce schéma est inspiré de la figure publiée par R.O. Duda [2000]. Le vecteur \vec{U} représente la localisation de la source sonore par rapport à l'auditeur. Le vecteur unitaire \hat{V} représente l'orientation de la tête de l'utilisateur. Les angles φ et δ représentent l'azimut et l'élévation de la source sonore. Le signal original discrétisé est représenté par $x[n]$. Les signaux $x_R[n]$ et $x_L[n]$ représentent les signaux sonores aux entrées des canaux auditifs gauche et droite respectivement.

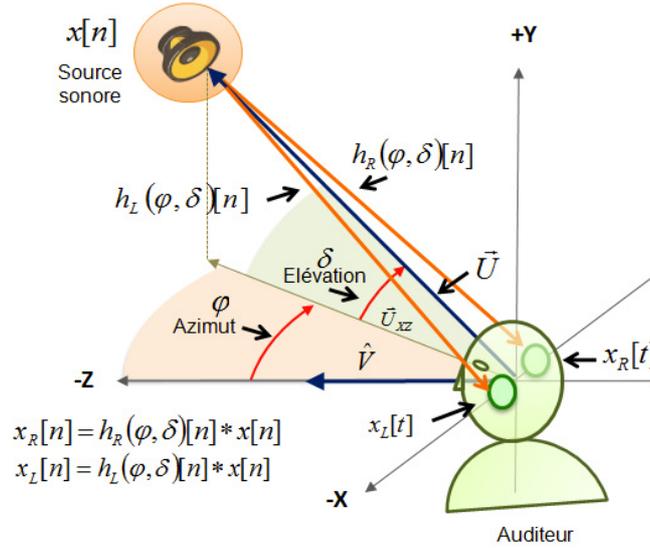


Figure 3.4. Fonction HRTF d'une source sonore spatiale.

Lorsqu'une source sonore est positionnée à φ en azimut et à δ en élévation par rapport à l'auditeur, l'effet de l'anatomie de la tête sur les sons captés par l'oreille droite et l'oreille gauche est déterminé par les HRIRs $h_R(\varphi, \delta)[n]$ et $h_L(\varphi, \delta)[n]$ respectivement. Ces HRIRs peuvent être appliquées à un stimulus sonore pour lui transférer des propriétés spatiales.

Les signaux captés $x_R[n]$ et $x_L[n]$ sont donc définis comme suit :

$$x_R[n] = h_R(\varphi, \delta)[n] * x[n] \quad (3.20)$$

$$x_L[n] = h_L(\varphi, \delta)[n] * x[n] \quad (3.21)$$

où $x[n]$ est le signal discret correspondant au stimulus original, et

* représente la convolution des deux signaux discrets.

Les HRIRs gauche et droite sont appliquées à l'aide de la procédure de superposition et d'addition. La Figure 3.5 présente un diagramme qui illustre le processus de l'application de la HRTF en temps réel. Les signaux $x_R[n]$ et $x_L[n]$ sont normalement diffusés directement aux entrées des oreilles de l'auditeur à l'aide d'un casque audio.

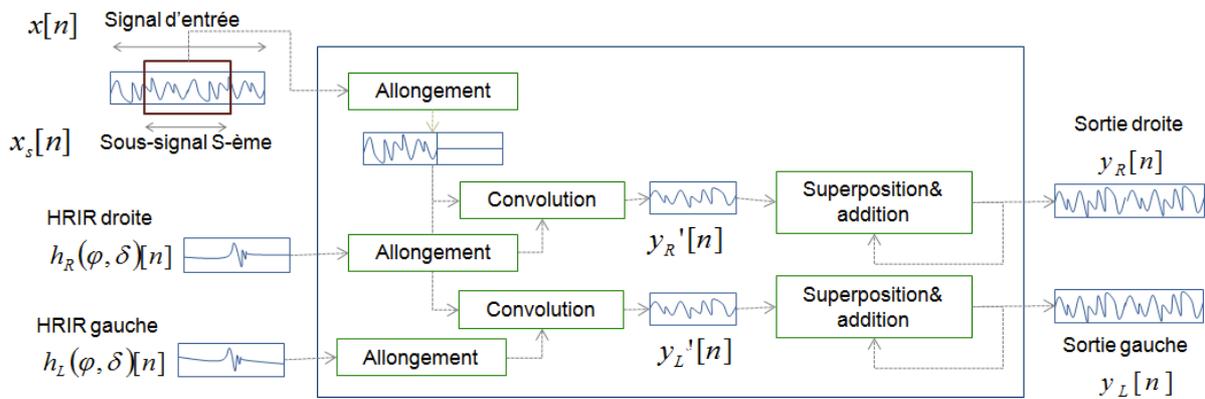


Figure 3.5. Processus d'application de la HRTF en temps réel.

3.3. Méthode d'analyse statistique

Dans cette section, nous allons présenter les concepts et les techniques statistiques utilisés pour l'analyse des données des expérimentations réalisées dans le cadre de cette thèse. Les aspects suivants sont traités : les mesures statistiques descriptives utilisées pour caractériser les données obtenues, les types de représentations graphiques employées, les méthodes de plan d'expérimentation et d'analyse statistique utilisées. Les analyses statistiques décrites ont été élaborées à l'aide des logiciels SPSS PASW 17 (IBM [2009]) et STATISTICA 9 (StatSoft [2009]).

3.3.1. Statistiques descriptives

Dans cette partie, nous allons décrire les concepts et les mesures statistiques utilisés pour caractériser les données obtenues à partir des expérimentations réalisées. Les échantillons de données sont caractérisés par les mesures suivantes : la moyenne, la médiane, l'écart-type, les quartiles et l'intervalle interquartile.

La moyenne arithmétique (ou simplement moyenne) d'un ensemble de données est le rapport de la somme des éléments de l'ensemble de données par le nombre d'éléments. Pour les distributions normales, elle est très utile pour caractériser l'ensemble des données.

La médiane d'un groupe de données arrangées en ordre croissant est l'élément qui sépare l'ensemble des données en deux parties égales. La médiane peut être utilisée pour localiser l'ensemble des données. C'est-à-dire qu'elle peut représenter le centre des données. La médiane et la moyenne sont utiles pour représenter et comparer graphiquement des groupes de données.

L'écart type d'un groupe de données est une mesure de la dispersion des éléments du groupe autour de la moyenne. L'écart type est la racine carrée de la variance. La variance est définie à son tour comme le rapport de la somme des carrés des différences entre chaque élément et la moyenne de l'ensemble par le nombre d'éléments. L'écart-type est utile pour représenter la dispersion de distributions normales et similaires.

Les quartiles d'un groupe de données arrangées en ordre croissant sont les trois éléments divisant l'ensemble des données en quatre parties de taille égale. Le premier quartile ($Q1$) sépare les premiers 25% de l'ensemble de données. Autrement dit, il limite le premier quart d'éléments. Le deuxième quartile ($Q2$) correspond à la médiane de l'ensemble des données. Le troisième quartile ($Q3$) sépare (à 75%) le quart supérieur des éléments du groupe de données. Les quartiles permettent de caractériser et représenter un ensemble de données et sa dispersion. L'intervalle interquartile (IQ) est une mesure de dispersion de données définie par la différence entre le troisième et le premier quartile ($IQ = Q3 - Q1$). Pour la caractérisation de groupes de données que ne suivent pas la distribution normale, les quartiles sont préférés à la moyenne et à l'écart-type.

3.3.2. Représentations graphiques

Nous utilisons deux types de représentations graphiques pour les données des expérimentations : les boîtes à moustaches et les graphiques de moyennes.

Une boîte à moustaches (*boxplots*) permet de représenter un échantillon de données sans faire aucune supposition sur la distribution statistique sous-jacente. Les boîtes à moustaches sont utiles pour identifier différentes propriétés de l'ensemble de données comme la dispersion et l'asymétrie (Mercier [2010], vol. 1).

Les informations explicitement décrites à travers les boîtes à moustaches sont les suivantes : la médiane et les autres quartiles, l'intervalle des valeurs non-atypiques, les valeurs atypiques et les valeurs extrêmes. Dans la Figure 3.6, nous montrons un exemple de graphique avec deux boîtes à moustaches. Dans cette figure, nous montrons aussi la moyenne de chaque échantillon.

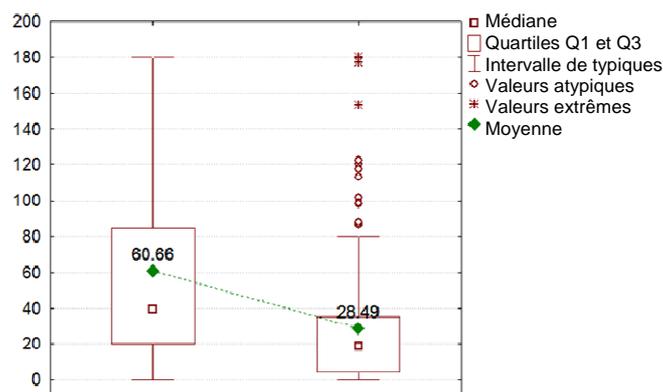


Figure 3.6. Exemple d'un graphique de type boîte à moustaches avec deux échantillons.

L'intervalle de valeurs typiques délimite l'espace des éléments considérés comme typiques. La limite inférieure de cet intervalle est donnée par l'élément le plus petit dans l'intervalle défini par $Q1 - 1.5 \times IQ$ et $Q1$. La limite supérieure est donnée par l'élément le plus grand entre $Q3$ et $Q3 + 1.5 \times IQ$.

Les valeurs atypiques sont les éléments inférieurs à $Q1 - 1.5 \times IQ$ ou supérieurs à $Q3 + 1.5 \times IQ$. Les valeurs extrêmes ou extrêmement atypiques sont les éléments inférieurs à $Q1 - 3 \times IQ$ ou supérieurs à $Q3 + 3 \times IQ$.

Les graphiques de moyennes se composent généralement des moyennes et des barres d'erreur indiquant la dispersion des données autour des moyennes. Ces graphiques permettent d'identifier la localisation de l'échantillon montré. Les barres d'erreur montrent normalement l'une des mesures statistiques suivantes : l'intervalle de confiance, l'écart-type ou l'erreur standard. Dans les graphiques de moyennes montrés dans cette thèse, nous utilisons l'intervalle de confiance à 95% pour montrer la dispersion des données. Un exemple du type de graphiques de moyennes utilisé dans cette thèse est présenté dans la Figure 3.7.

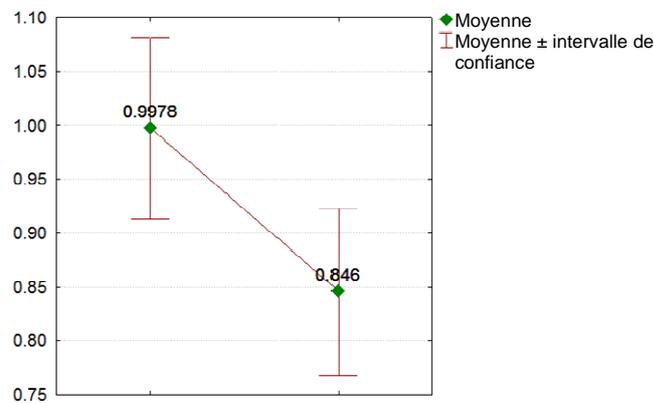


Figure 3.7. Exemple d'un graphique de moyennes avec deux échantillons.

L'intervalle de confiance est une mesure statistique indiquant la fiabilité de l'estimation de la moyenne de valeurs de plusieurs échantillons. Le calcul de l'intervalle de confiance nécessite la définition du niveau de confiance. Cette valeur est fixée à 95% pour les résultats et analyses effectués dans cette thèse. Le niveau de confiance exprime la probabilité que l'intervalle défini contient la valeur estimée.

3.3.3. Distribution fréquentielle

Les types de distribution fréquentielle (lois) utilisées dans nos expérimentations sont les suivantes : normale, log-normale, gamma, exponentielle et binomiale. Leurs définitions peuvent être consultées dans différents ouvrages comme Spiegel et al. [2000].

Pour déterminer le type de distribution d'un groupe de données, nous utilisons le test de Kolmogorov Smirnov (K-S). Ce test est un outil statistique permettant de comparer la distribution d'un groupe de données avec une distribution de référence. Il mesure la distance entre les deux groupes de valeurs. L'hypothèse nulle

associée à ce test est que les données suivent bien la distribution de référence. L'hypothèse nulle est rejetée si la probabilité p donnée par le test est inférieure à 0.05.

3.3.4. Expérimentation

Dans cette partie, nous décrivons les concepts utilisés pour définir une procédure expérimentale. Une expérience consiste à observer un événement pour un objectif précis comme celui de la validation d'une hypothèse. L'observation implique normalement la mesure des variables de l'événement après ou pendant l'expérience. Dans les expérimentations réalisées dans cette thèse, une expérience consiste à observer le déroulement d'une tâche en particulier par un groupe de sujets dans des conditions expérimentales différentes.

3.3.4.1. Conditions et facteurs expérimentaux

Une expérimentation est constituée de différentes conditions expérimentales dans lesquelles un test est effectué. Une condition expérimentale représente les circonstances dans lesquelles un essai expérimental (*experimental repetition*) ou un groupe d'essais expérimentaux se sont réalisés. Un facteur expérimental est une variable qui est délibérément variée entre différentes conditions expérimentales. Dans nos expérimentations, un facteur expérimental se réfère à un élément qui est présent ou absent en fonction de la condition expérimentale. Ainsi, nous pouvons par exemple, évaluer le facteur du son spatial lorsque qu'il est présent ou absent durant l'accomplissement d'une tâche. L'étude des effets statistiques des facteurs expérimentaux dans les variables mesurées représente souvent un des objectifs principaux d'une expérimentation.

L'effet d'apprentissage (*carry-over effect*) se réfère à l'expérience acquise par un individu lorsqu'il réalise successivement une même tâche dans des conditions expérimentales différentes. Cette expérience peut affecter les variables mesurées de façon non désirée. L'effet d'apprentissage est présent dans les expérimentations à échantillons dépendants. Changer l'ordre des conditions expérimentales pour chaque participant permet de réduire son influence dans les résultats. Bien que cette solution ne permette pas d'éliminer complètement cet effet, elle donne la possibilité de le distribuer dans l'ensemble des conditions de façon plus ou moins équilibrée. Il est donc conseillé d'utiliser un ordre aléatoire de conditions expérimentales pour chaque utilisateur.

3.3.4.2. Échantillon

Le terme échantillon de données se réfère à un groupe de données obtenu par des procédures expérimentales ayant certaines caractéristiques communes (ex. : les résultats obtenus pour une condition expérimentale en particulier). Il est important de différencier les échantillons indépendants des échantillons dépendants pour le type d'expérimentation réalisée. Le premier terme implique la participation d'un groupe formé par des individus différents pour chaque condition. Les échantillons dépendants se réfèrent à la participation d'un même groupe de participants dans toutes les conditions.

Les échantillons indépendants nécessitent d'être plus larges que les dépendants. Ceci est dû au fait que les indépendants peuvent présenter une grande variabilité inter sujets (*between-subjects*) de chaque échantillon. Une grande variabilité entre les individus peut représenter un problème dû au fait que les échantillons ne sont pas

suffisamment équilibrés. L'utilisation d'échantillons larges permet de réduire la probabilité d'obtenir de grandes variabilités entre les différents sujets. En revanche, les échantillons dépendants présentent généralement une variabilité plus petite entre les sujets puisqu'il s'agit des mêmes individus. De cette façon, l'utilisation d'échantillons dépendants permet de travailler avec des tailles d'échantillon plus réduites (Dallal [2001]). Pour cette raison, nous avons choisi d'utiliser des échantillons dépendants dans nos expérimentations.

3.3.4.3. Essais

On parle d'une expérimentation à multiples essais si chaque utilisateur réalise plus d'un test pour chaque condition expérimentale. Il est possible que les circonstances changent légèrement entre les différents essais de la tâche pour chaque condition expérimentale. Ceci permet, par exemple, d'étudier une sélection représentative des différentes conditions initiales possibles de la tâche observée. Les conditions initiales se réfèrent aux valeurs initiales (au moment du début de la tâche) des variables qui changent durant chaque essai expérimental. L'observation de divers essais avec des conditions initiales différentes permet une meilleure compréhension du phénomène puisqu'elle prend en compte une pluralité de situations. En revanche, l'utilisation des mêmes conditions initiales pour tous les essais conditionne fortement les résultats. Nous pouvons ainsi étudier, par exemple, la précision de la localisation d'un groupe de sources sonores positionnées dans différentes positions par rapport à l'auditeur (essais et leurs conditions initiales) pour différentes techniques de spatialisation sonore (conditions expérimentales). Les essais sont effectués par les sujets dans un ordre aléatoire pour réduire l'effet d'apprentissage.

Pendant la réalisation des expérimentations, nous avons constaté que les sujets n'ont pas montré une bonne performance lors des premiers essais de la tâche. Cet effet résulte du manque de maîtrise de l'outil d'interaction 3D. Les techniques d'interaction en réalité virtuelle nécessitent un temps d'apprentissage avant la réalisation de la tâche. Pour cela, nous avons décidé de permettre aux sujets d'effectuer quelques essais préliminaires « à blanc » avant la réalisation de la tâche expérimentale qui sera prise en considération dans l'analyse de données.

3.3.4.4. Considérations générales

L'utilisation de plusieurs conditions expérimentales, de plusieurs essais et d'échantillons dépendants permet d'augmenter le nombre de données à analyser pour chaque utilisateur. Un nombre important de données à analyser permet d'améliorer la fiabilité des résultats puisque la probabilité d'avoir proportionnellement trop de comportements marginaux est réduite.

Dans ce paragraphe, on résume les éléments les plus importants des expérimentations réalisées dans le cadre de cette thèse. Une expérimentation consiste à observer un sujet qui réalise une tâche dans différentes conditions expérimentales et avec un nombre d'essais prédéfini. L'observation se réfère à la prise de mesures pendant et à la fin de l'exécution de la tâche. Ces mesures seront analysées pour étudier la performance de l'utilisateur dans les conditions expérimentales. Ces dernières seront présentées aux sujets dans un ordre aléatoire. Chaque condition se compose du même groupe d'essais. Les essais ont chacun des conditions initiales différentes. Les essais sont réalisés dans un ordre aléatoire pour chaque condition expérimentale. Les essais préliminaires ne sont pas pris en compte dans l'analyse statistique. Nous utilisons le même groupe de sujets pour les différentes conditions expérimentales, ce qui constitue des échantillons dépendants (*within-subjects*

design). La Figure 3.8 montre un exemple d'expérimentation avec deux conditions expérimentales et un facteur expérimental.

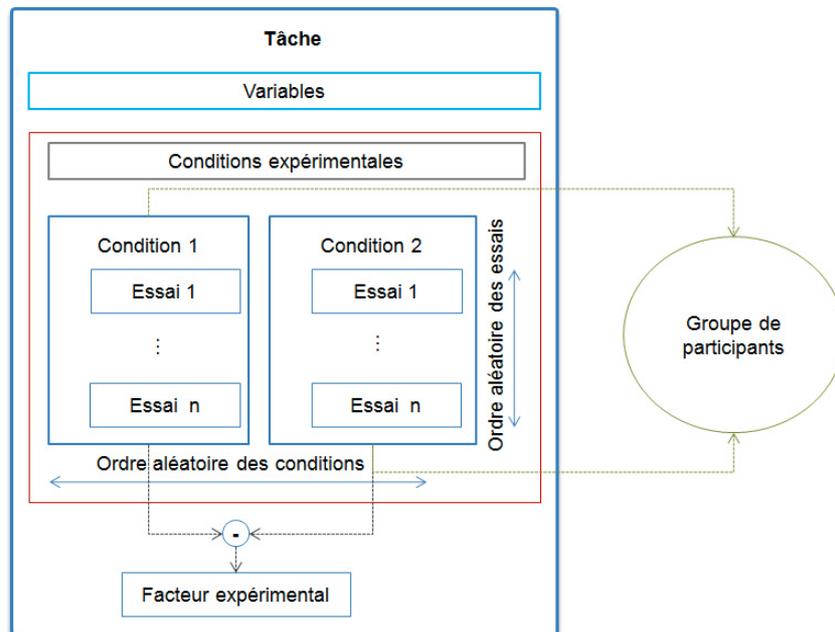


Figure 3.8. Exemple de protocole expérimental.

3.3.5. Méthode d'Equations Généralisées d'Estimation

La méthode d'Equations Généralisées d'Estimation (GEE : *General Estimating Equations*) est une procédure statistique permettant de déterminer la signifiante des effets des facteurs expérimentaux dans une expérience avec multiples essais et échantillons dépendants. La GEE offre ainsi la possibilité d'effectuer l'analyse statistique d'une expérimentation intra-facteurs (*within-factors*) et intra-participants (*within-subjects*).

La méthode GEE permet d'analyser des données qui ont une distribution de fréquences qui n'est pas forcément normale. Cette méthode nécessite de choisir le modèle de distribution de fréquences le plus proche possible de la distribution des données à analyser. De cette manière, la méthode GEE permet d'effectuer une analyse statistique mieux adaptée aux données que celle réalisée avec d'autres techniques telle que l'analyse de variance (ANOVA).

Le choix du modèle de distribution à utiliser est réalisé à travers l'application du test d'ajustement de Kolmogorov-Smirnov. Dans les analyses effectuées, les distributions de fréquences testées sont les suivantes : normal, log-normal, exponentiel, gamma, chi-square, poisson, géométrique, binomial et bernoulli. Pour chaque variable, le modèle de distribution de fréquences choisi est celui qui a le test de Kolmogorov-Smirnov le plus proche avec une probabilité d'ajustement positive.

La méthode GEE permet l'analyse d'un ensemble de données ayant une distribution qui n'est pas forcément normale. Elle nécessite de choisir un modèle de distribution de fréquences assimilable à la distribution des données à traiter. Cette méthode permet d'effectuer une analyse statistique mieux adaptée aux données par rapport à d'autres techniques existantes telle que l'ANOVA à mesures répétées.

Le principe des méthodes d'analyse statistique est la comparaison entre échantillons pour déterminer s'ils sont significativement différents. La méthode GEE, comme plusieurs autres méthodes d'analyse statistique, est basée sur l'ajustement des modèles aux échantillons évalués. Une description détaillée de cette méthode est publiée dans Norusis [2009]. Dans ce qui suit, nous décrivons les éléments qui composent sa sortie.

Le Degré de Liberté (DDL) est une mesure qui désigne le nombre de valeurs aléatoires qui ne peuvent pas être déterminées par le modèle linéaire calculé. Le DDL est utilisé dans l'interprétation de résultats statistiques.

La méthode GEE utilise le test de khi carré (χ^2) de Wald pour comparer les échantillons. Ce test fait partie du groupe de tests statistiques non paramétrés qui considèrent que la distribution des données est anormale. Il permet d'évaluer la distance entre des ensembles de données (échantillons) en fonction du type de distribution de fréquences spécifié et du nombre de degrés de liberté calculé. Dans les tableaux des résultats montrés dans cette thèse, la valeur de χ^2 se réfère à la distance calculée par le test.

L'hypothèse nulle signifie que la différence entre des échantillons à comparer n'est pas statistiquement significative. p représente la probabilité que l'hypothèse nulle est vérifiée. L'hypothèse nulle sert à déterminer si le facteur expérimental testé est statistiquement significatif sur la variable à analyser. Il est souvent convenu que si la valeur de p est inférieure à 0.05, le facteur évalué est significatif avec une fiabilité de 95% ($0.95 = 100 \times (1 - p)$). Si la valeur de p est supérieure à 0.05, on considère que la différence entre les échantillons n'est pas significative et qu'elle est due soit aux fluctuations d'échantillonnage soit au hasard.

3.3.6. Méthode d'Estimation Généralisée de Moyennes Marginales

La méthode d'Estimation Généralisée de Moyennes Marginales (*GEMMeans : Generalized Estimated Marginal Means*) est un outil d'analyse statistique basé sur le calcul de la réponse moyenne de chaque facteur expérimental ajustée à une ou à plusieurs variables. Les moyennes marginales de deux groupes de données dénotent l'ensemble de valeurs moyennes calculées à partir des éléments de chaque groupe.

La méthode GEMMeans sert à comparer des paires de groupes de données pour déterminer si leur différence est statistiquement significative. Par exemple, elle permet de comparer directement les résultats obtenus avec deux conditions expérimentales. En revanche, la GEE permet uniquement de comparer les données d'un ensemble de conditions expérimentales pour un facteur expérimental donné.

Pour plus d'informations sur cette méthode, le lecteur peut consulter Norusis [2009]. Dans ce qui suit, nous décrivons les concepts de base de l'emploi et de l'interprétation des valeurs de sortie de GEMMeans.

L'erreur de l'estimation (ou erreur standard) est une mesure de la précision du modèle linéaire obtenu par GEMMeans. Les petites valeurs de l'erreur standard indiquent que la différence entre les échantillons comparés n'est pas significative.

La différence moyenne est une mesure statistique égale à la moyenne des différences absolues entre les éléments de deux échantillons comparés. Cette mesure permet de caractériser la dispersion entre les deux échantillons.

L'intervalle de confiance de Wald est une façon de calculer les écarts majeur et mineur, par rapport à un échantillon normal, du modèle obtenu par la méthode d'analyse utilisée. La définition de cet intervalle nécessite

d'établir un niveau de confiance (associé au paramètre α). Le niveau de confiance peut varier mais il est normalement fixé à 95% ($0.95 = 100 \times (1 - \alpha)$). Les mesures p et DDL, déjà présentées précédemment, font aussi partie des valeurs de sortie de la méthode GEMMeans.

3.4. Plateforme expérimentale

Dans cette section, nous décrivons la plateforme expérimentale développée pour réaliser les expérimentations présentées dans cette thèse. Elle permet un rendu sonore spatial en temps réel et elle inclue le suivi de la tête, la restitution binaurale et multicanale de champs sonores ainsi que des effets de réverbération simplifiés.

Cette plateforme est similaire à d'autres solutions existantes telles que l'IKA-SIM, le DIVA et le VAP entres autres. La valeur ajoutée de la plateforme proposée dans le cadre de cette thèse, est sa capacité à combiner une spatialisation sonore et des indications auditives supplémentaires.

3.4.1. Architecture logicielle

La Figure 3.9 montre le schéma de l'architecture matérielle de la plateforme expérimentale. Les divers composants et leur connectivité sont montrés dans ce schéma.

Les caractéristiques les plus importantes de ces composants sont montrées dans le Tableau 3.1. La dernière colonne comprend des caractéristiques ayant un impact important sur les expériences. Ces informations sont importantes pour deux raisons. Premièrement, elles permettent la différenciation de la solution technologique utilisée des autres solutions disponibles sur le marché. Deuxièmement, elles permettent d'estimer la performance de chaque composant et de l'ensemble du système.

La Nintendo Wiimote (Nintendo [2008]) est un dispositif de commande sans fil avec retour de vibration et interface Bluetooth. Ce dispositif permet à l'utilisateur de communiquer facilement avec le système.

Le Xsens MTi (Xsens [2009]) est un traqueur d'orientation basé sur des accéléromètres à 3 degrés de liberté (DDL) : 360° en roulis et en lacet et 180° en tangage. L'interface du traqueur utilise la norme USB. Ce traqueur est fixé sur le casque pour permettre de capturer l'orientation de la tête de l'utilisateur. La résolution du traqueur permet un suivi de la tête approprié pour l'écoute spatiale puisque la résolution du capteur (1°) est nettement supérieure à la précision typique de l'utilisateur humain pour la localisation de sources sonores spatiales (10° - 25°).

La diffusion du son est effectuée à l'aide d'un casque audio de type fermé qui permet un bon niveau d'isolation avec l'environnement acoustique réel. Le casque permet la diffusion audio à haute fidélité, il est confortable et relativement léger.

Le Novint Falcon est une interface haptique à faible coût de 3 degrés de liberté produite par Novint Technologies Inc [2007]. Ce dispositif permet de capturer la position d'un curseur en trois dimensions et de restituer une force linéaire. Il est important de mentionner que ce dispositif est utilisé comme mécanisme de positionnement et non pas pour étudier le rendu haptique. Dans les expériences réalisées, cette interface a été utilisée pour indiquer la direction dans l'espace.

Les caractéristiques de la carte graphique et de la station de travail sont fournies dans le Tableau 3.1 pour donner une description complète de la plateforme.

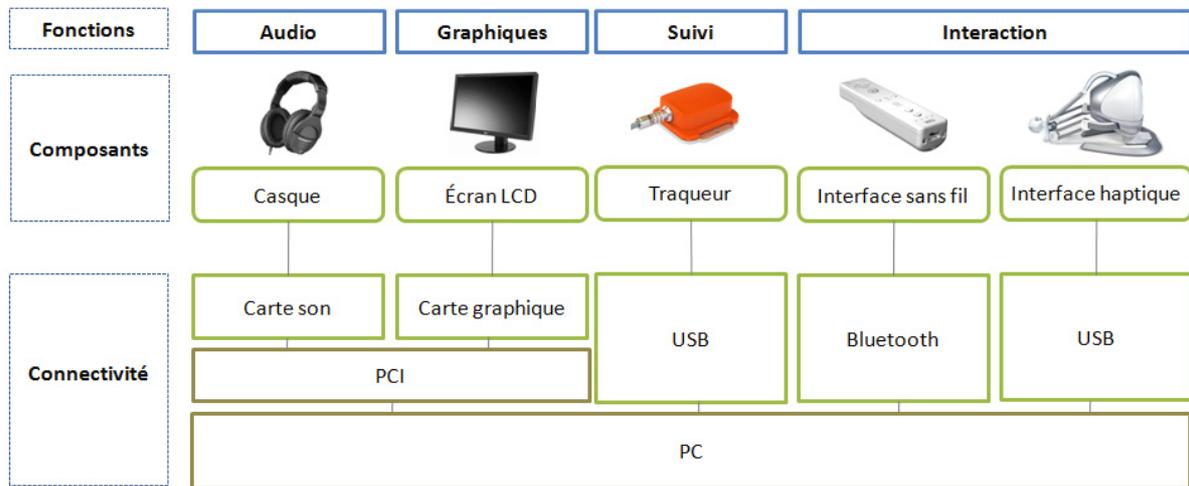


Figure 3.9. Schéma de l'architecture matérielle de la plateforme expérimentale utilisée.

Tableau 3.1. Architecture matérielle

Fonction	Dispositif	Description	Interface	Vitesse	Autres caractéristiques
Traitement audio	Creative Sound Blaster X-Fi Pro	Carte son	PCI	194 kHz	Processeur de 24 bits, effets EAX
Sortie audio	Sennheiser HD 201	Casque audio	Analogique	-	Dynamique fermé
Interaction dans l'espace 3D	Nintendo Wiimote	IHM sans fil	Bluetooth	≈100 Hz	10 boutons, retour de vibration
	Novint Falcon	Interface haptique	USB	480 Mbps	3 DDL ; espace de travail: 10.2cm×3 ; résolution>400 dpi ; force > 0.9 kg
Suivi de mouvement	XSens MTi	Traqueur basé sur accéléromètres	USB	120 Hz	2.5 degrés de liberté ; erreur<1.0°
Carte graphique	NVIDIA Quadro FX 560	Carte graphique	PCI	24 Hz	128 MB
Station de travail	HP xw8400	CPU	-	2×1.6 GHz	2 GB en RAM

3.4.2. Architecture logicielle

Le schéma de l'architecture logicielle est montré dans la Figure 3.10. Les principales caractéristiques des composants logiciels sont détaillées dans le Tableau 3.2.

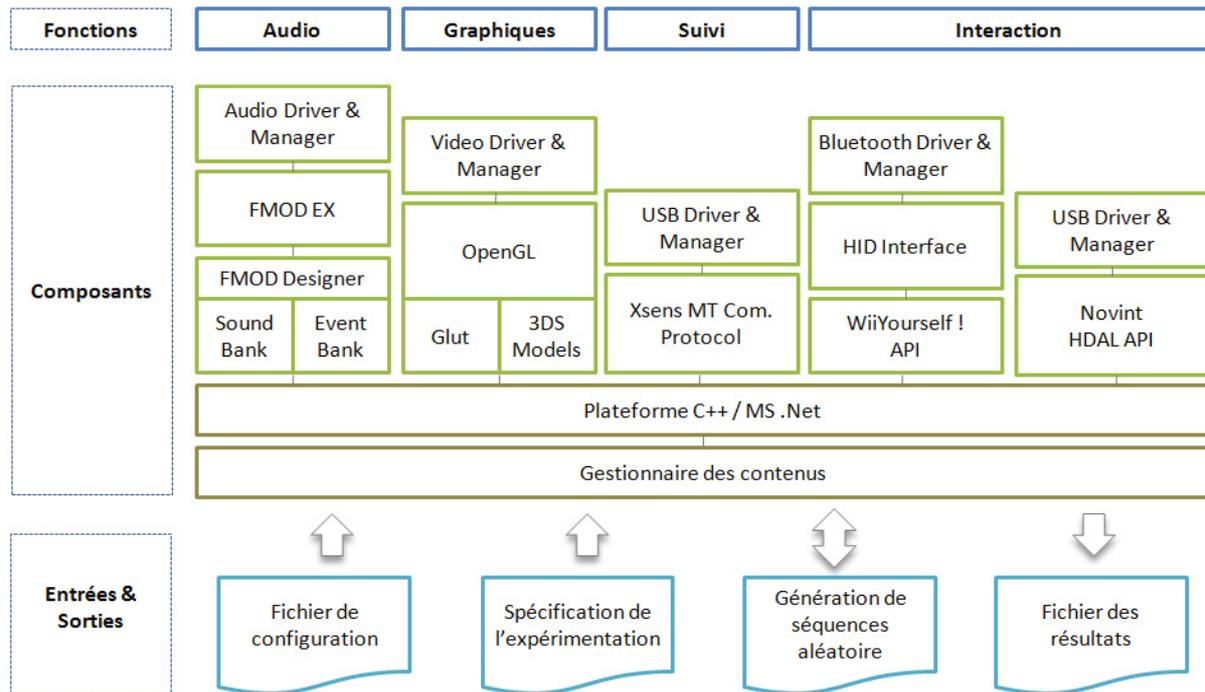


Figure 3.10. Schéma de l'architecture logicielle de la plateforme expérimentale développée.

Tableau 3.2. Architecture logicielle

Fonction	Composant	Version	Description	Auteur
Audio	FMOD EX API & Designer	4.33 & 4.23	Librairie de traitement de son	Firelight Technologies
Graphique	OpenGL & GLUT	2.1 & 3.7	Librairie graphique	Silicon Graphics & Mark Kilgard
Interaction	WiiYourself !	1.01	Interface pour la Wiimote	Brian Peek
	Haptic Device Abstraction Layer (HDAL)	2.1.3	Interface pour Novint Falcon	Novint Technologies Inc.
Suivi de mouvement	MTi Xsens Comm Protocol	1.2	Interface pour les traqueurs MTi/MTx	Xsens
Développement	Microsoft Visual C++ 2003	7.0	Environnement de développement	Microsoft

FMOD (FMOD Technologies [2005]) est une librairie de traitement de son en temps réel. Elle est couramment utilisée dans le développement de jeux vidéo. FMOD permet la communication avec les dispositifs de lecture, de traitement et de diffusion audio. La librairie offre la possibilité d'utiliser différentes techniques d'audio telles que le positionnement géométrique, le filtrage fréquentiel et la réverbération. Les effets de réverbération sont produits à l'aide de la technologie EAX (Environmental Audio eXtensions, Creative-Labs [1999]).

FMOD Designer est un module complémentaire de la bibliothèque qui permet la création des collections de sons et des collections d'effets sonores paramétrées. Nous avons créé une collection de son et une collection d'effets sonores qui sont utilisées en ligne dans notre système.

L'application logicielle a été développée sur le Microsoft Visual C++ (Microsoft [2003]). Le gestionnaire de contenus est le module logiciel principal de la plateforme qui permet la gestion des ressources et la spécification de l'expérimentation. Le gestionnaire reconnaît des fichiers de configuration et permet la génération des fichiers de résultats. Les fichiers de configuration permettent de définir les ressources à utiliser et de spécifier ses conditions de fonctionnement. Par exemple, il permet d'indiquer le type de système de diffusion sonore à utiliser et de configurer la connectivité des différents dispositifs. Le fichier de spécification de l'expérimentation est utilisé pour définir et reconfigurer les différents aspects de chaque préparation expérimentale sans avoir besoin de modifier le code de l'application. Ainsi, ce fichier permet de changer les conditions expérimentales, le nombre d'essais et les conditions initiales de chacun. Le générateur de séquences aléatoires permet de créer et de sauvegarder les séquences dans lesquelles les essais expérimentaux et les conditions expérimentales sont effectués. L'enregistrement de ces séquences permet d'éviter qu'un utilisateur exécute la même séquence de essais pour deux conditions expérimentales différentes.

D'autres composants sont utilisés pour développer les différentes fonctionnalités du système. Le module graphique a été développé à l'aide des bibliothèques OpenGL (Segal et Akeley [2006]) et GLUT (Kilgard [1996]). Celles-ci sont des bibliothèques graphiques génériques utilisées pour le rendu tridimensionnel et pour la gestion de la communication avec la carte graphique et l'interface graphique du système d'exploitation.

Le composant logiciel pour la communication avec le traqueur (MT Communication Protocol) via USB est fourni par XSens (Xsens [2009]). La communication avec la Wiimote est effectuée à l'aide de la librairie Wiiyourself (Peek [2009]). Ce composant logiciel est basé sur le protocole de communication sans fil Bluetooth en utilisant la librairie USB HID API (Delahoussaye [2010]).

Le taux de rafraîchissement de l'image et de l'audio est supérieur à 25 Hz. Ce facteur est important pour la mise à jour des positions des sources sonores. L'ensemble des composants logiciels et matériels constitue une solution à faible coût pour le rendu de son spatialisé en environnement virtuel.

4. Approche proposée : Sensations Auditives Spatiales Artificielles (SASA)

Dans ce chapitre, nous décrivons l'approche proposée pour fournir une information spatiale précise à l'aide du retour sonore. L'approche que nous proposons est appelée Sensations Auditives Spatiales Artificielles (SASA). Nous présentons les concepts les plus importants, les objectifs envisagés, les aspects à considérer, la structure des indices SASA et la méthode de conception. Une étude expérimentale est également présentée pour valider la contribution de notre approche dans la localisation précise de sources sonores.

4.1. L'approche SASA

4.1.1. Description générale de l'approche SASA

L'approche SASA a pour but d'offrir un mécanisme de transmission d'informations spatiales précises via le son pour assister l'utilisateur dans la réalisation d'une tâche en environnement virtuel. Cette technique a été développée à partir des besoins et des critères discutés précédemment dans la section 2.3.5.

L'utilisation du retour sonore pour transmettre des informations spatiales promet des applications intéressantes dans le domaine de l'interaction 3D. De nombreuses tâches interactives (comme la manipulation et la sélection d'objets) nécessitent une bonne précision de l'information spatiale transmise. Le son 3D est une technologie mature mais les méthodes de spatialisation sonore existantes ne permettent pas la localisation précise des sources sonores. De plus, ces techniques présentent des problèmes d'ambiguïté qui peuvent aussi affecter l'identification des informations spatiales transmises. Afin de pouvoir utiliser le retour sonore spatial pour transmettre une information spatiale précise et utile pour l'interaction en environnement 3D, il est nécessaire de développer une nouvelle approche qui bénéficie des avantages des techniques existantes du son 3D mais qui permet d'améliorer la précision et de réduire l'ambiguïté.

Le principe de la technique SASA est de combiner une source sonore spatialisée avec des indices sensoriels (*sensory cues*). Ces indices sensoriels sont générés à l'aide d'effets audio qui modifient le stimulus sonore original. Le but est de créer des sensations auditives permettant de mieux capter les propriétés spatiales de la source sonore ainsi que d'autres informations qui peuvent être éventuellement ajoutées. Comme l'indique le

nom de la technique SASA (Sensations Auditives Spatiales Artificielles), les sensations créées (ou générées) sont artificielles parce qu'elles ne correspondent pas forcément au phénomène naturel de l'écoute spatiale.

La technique SASA peut être utilisée dans différentes applications. Dans chaque cas, il est nécessaire d'identifier les indices sensoriels potentiellement utiles pour assister l'utilisateur dans l'accomplissement de sa tâche. De cette façon, un groupe d'indices SASA doit être défini ou adapté pour chaque application.

4.1.2. Structure d'un groupe SASA

Un groupe d'indices SASA, ou groupe SASA, se compose d'un ensemble d'indices ayant un objectif bien défini. Ces indices permettent d'enrichir un stimulus sonore qui sera spatialisé par la suite à l'aide d'une méthode appropriée. La Figure 4.1 montre la structure générale d'un groupe d'indices SASA, sa relation avec le stimulus sonore choisi et la méthode de spatialisation à utiliser.

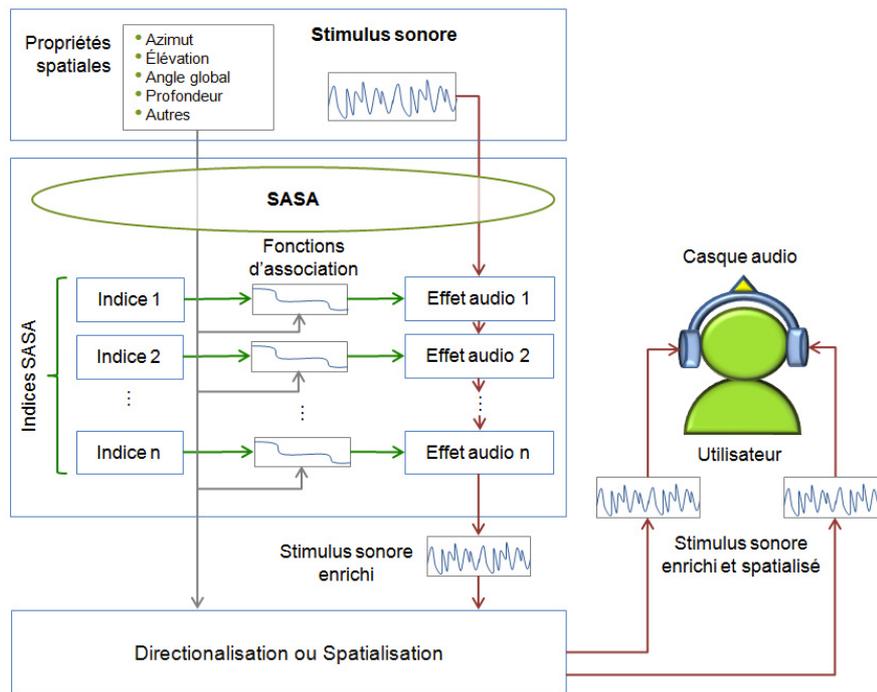


Figure 4.1. Structure générale d'un groupe d'indices SASA.

La source sonore associée au groupe SASA possède les propriétés spatiales suivantes : l'azimut, l'élévation, l'angle global et la profondeur ou distance (Figure 4.2). De plus, en fonction des besoins de la tâche à réaliser, il est possible d'ajouter d'autres propriétés spécifiques telles que des indicateurs sur le déroulement de la tâche (par exemple, son début, sa fin et les phases intermédiaires de sa réalisation) ou sur des aspects relatifs à l'interaction tels que la saisie d'un objet virtuel ou la collision entre objets.

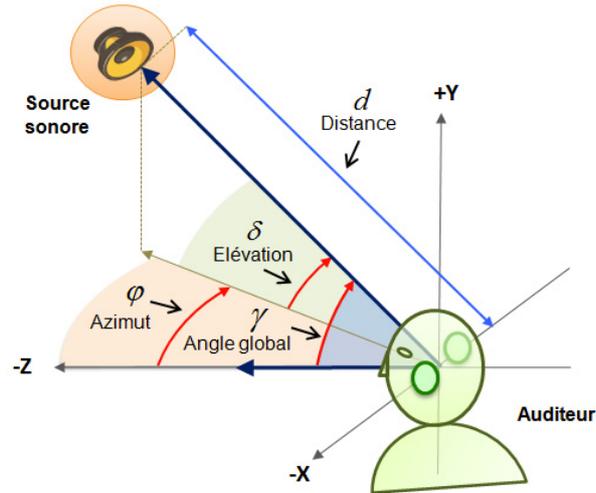


Figure 4.2. Propriétés spatiales d'une source sonore.

L'objectif de base d'un groupe d'indice SASA est de fournir une information spatiale précise. Néanmoins, le degré de précision nécessaire peut varier en fonction de l'application. De plus, il est possible d'utiliser la technique SASA pour fournir certaines informations supplémentaires qui ne sont pas intrinsèquement spatiales. De cette façon, il faut déterminer deux éléments : le niveau de précision de localisation recherché et les informations supplémentaires à transmettre.

En physiologie, un indice sensoriel est une information qui peut être extraite à partir d'une entrée sensorielle exprimant une propriété de l'environnement. Dans notre approche, un indice sensoriel SASA (ou indice SASA) est une information ajoutée au stimulus sonore pour que le perceuteur (l'auditeur) puisse l'identifier et s'en servir dans l'accomplissement d'une tâche.

Un groupe d'indices SASA est utilisé pour améliorer la performance d'un individu dans l'accomplissement d'une tâche interactive. Cette tâche sera réalisée dans un environnement virtuel tridimensionnel. De plus, il est nécessaire que l'utilisateur puisse tirer profit du caractère spatial de l'information transmise à l'aide de SASA. Les types de tâches interactives visées sont principalement la manipulation et la sélection d'objets en environnement tridimensionnel. Néanmoins, nous considérons qu'il est également possible d'adapter notre approche à des tâches de navigation assistée, au suivi spatial de cibles et à la diffusion d'alertes spatiales.

4.1.2.1. *Stimulus sonore et directionnalisation*

Au début du processus, il faut sélectionner un stimulus sonore monocanal. Ce stimulus sera traité pour lui ajouter des indices sensoriels et pour être ensuite spatialisé. Par ailleurs, le stimulus choisi doit satisfaire certains critères.

Il est nécessaire de faire le choix entre les sons verbalisés et les sons non-verbalisés. Selon Brewster [2003], le son non-verbalisé (*non-speech sound*) présente les avantages suivants : il permet d'obtenir un retour rapide suite aux actions de l'utilisateur, l'information à transmettre peut être comprise plus rapidement qu'une

phrase, il permet de fournir des informations structurées d'une manière continue. Pour ces raisons, nous préférons utiliser ce type de stimuli.

Il est également nécessaire que le stimulus soit cohérent avec les exigences de l'application. En particulier, il faut que le stimulus n'entre pas en conflit avec les autres flux d'information de l'application. En d'autres termes, le stimulus doit être élaboré sur des critères d'ergonomie, il faut qu'il soit facilement identifiable pour éviter qu'il perturbe l'utilisateur. Le stimulus doit être testé avec un groupe restreint d'utilisateurs pour s'assurer qu'il n'induit aucune perturbation..

Il est nécessaire de doter le stimulus sonore de la propriété de directivité. Pour cela, il faut sélectionner une méthode de spatialisation sonore. Les différentes méthodes existantes ont été déjà discutées dans la section 2.2. Quelque soit la méthode employée, si elle est utilisée toute seule, le niveau de précision pour la localisation sonore est toujours faible parce que la performance humaine en conditions naturelles est faible. C'est la raison principale pour laquelle les indices sensoriels sont ajoutés.

4.1.2.2. *Indices SASA*

L'information d'un indice SASA est transmise à travers un effet audio appliqué au stimulus sonore choisi. Les indices SASA sont définis avec la possibilité de les combiner avec d'autres indices. Néanmoins, il faut vérifier qu'il n'existe pas de conflits entre les indices après leur combinaison.

Un indice sonore peut correspondre à deux informations : une information manquante (qui n'existe ni dans le stimulus ni dans la source sonore) ou une information complémentaire (qui permet d'améliorer la compréhension d'une information existante dans le stimulus ou dans la source sonore). Les indices sensoriels doivent être complémentaires entre eux. De plus, chaque indice sensoriel doit avoir un objectif particulier bien précis.

Il faut définir une liste d'indices sensoriels qui permettront hypothétiquement d'atteindre le niveau de précision recherché et de transmettre les informations supplémentaires envisagées. L'ensemble d'indices SASA sera composé de deux sous-ensembles : les indices de base et les indices spécifiques. Les indices de base permettent d'améliorer la précision de la localisation sonore. Ils peuvent être adaptés en fonction du niveau de précision recherché dans la réalisation de la tâche. Les indices spécifiques se réfèrent aux informations additionnelles apportées qui ne sont pas directement liées à la précision. Les indices spécifiques dépendent de l'application à développer.

Les indices sonores dérivent du besoin de transmettre une information particulière. Cette information représente le message de l'indice. Ce message peut avoir pour but de mettre en valeur une certaine propriété spatiale ou d'exprimer l'état actuel d'une variable. Dans notre approche, ce message est transmis en faisant varier l'un des paramètres d'un effet audio pour le même stimulus sonore.

Après l'identification de l'indice, on détermine la propriété de la source sonore dont dépend l'information à transmettre. Cette propriété est prédominante car elle possède la plus grande influence sur l'information à transmettre. Lorsqu'il s'agit d'une information spatiale, la propriété prédominante est une des propriétés spatiales de la source sonore. Par contre, si l'information à transmettre n'est pas spatiale, la propriété prédominante est une des propriétés spécifiques de la source sonore.

Les valeurs critiques correspondent aux niveaux les plus significatifs dans la transmission de l'information de l'indice SASA. Les valeurs critiques de la propriété prédominante peuvent représenter, par

exemple, le déclenchement de l'émission d'un message, la transition entre deux états ou le rapprochement d'une valeur ayant une signification spécifique.

4.1.2.3. Effet audio

Un effet audio est un processus qui modifie un stimulus sonore. Les types d'effet audio utilisés dans l'approche SASA sont : le filtrage (filtres passe-bande, passe-bas et passe-haut), la réverbération, le gain (ou atténuation) et le pitch.

La propriété prédominante de l'indice SASA doit être associée au paramètre de contrôle de l'effet audio. Ainsi, l'intensité de l'effet audio appliqué dépend de la propriété prédominante de l'indice. Cette relation est spécifiée sous la forme d'une fonction d'association (*mapping function*). Cette association est effectuée en définissant deux éléments : les valeurs critiques de la propriété prédominante de l'indice et les niveaux distinguables du paramètre de contrôle de l'effet audio.

Les associations doivent être les plus intelligibles possibles. Si l'effet audio sélectionné possède plus qu'un paramètre, il est nécessaire de choisir celui qui permet d'obtenir les variations les plus facilement distinguables en fonction du stimulus choisi. Ce paramètre est considéré le paramètre de contrôle à varier de l'effet audio.

Après la détermination du nombre de valeurs critiques, il faut déterminer les niveaux distinguables du paramètre de contrôle de l'effet audio. Les niveaux distinguables sont les valeurs du paramètre de contrôle de l'effet audio produisant des sons facilement différenciables entre eux lorsque l'effet est appliqué au stimulus sonore. Le nombre de niveaux distinguables à identifier doit être égal au nombre de valeurs critiques de la propriété prédominante. Il est nécessaire de faire particulièrement attention aux valeurs extrêmes afin que le résultat obtenu ne produise pas de gênes pour les utilisateurs.

Le terme de niveau distinguable semble similaire à celui de seuil différentiel (JND : *Just Noticeable Difference*) utilisé dans le domaine de la psychophysique. Néanmoins, un seuil différentiel caractérise le changement minimal perceptible d'un stimulus (Booth et Freeman [1993]). En revanche, un niveau distinguable représente la valeur d'une variable produisant un changement facilement identifiable par rapport à une valeur de référence. De plus, on peut avoir plusieurs niveaux distinguables mais il existe un seul JND pour une variable physique. Il est nécessaire d'effectuer des tests systématiques pour déterminer un seuil différentiel. Par contre, pour définir un niveau distinguable, il faut que le concepteur s'assure que le son produit est effectivement facile à distinguer des autres niveaux distinguables déterminés précédemment.

4.1.2.4. La fonction d'association

La fonction d'association est définie par une courbe construite à partir de points de référence à l'aide d'une méthode d'interpolation. Les points de référence sont positionnés à l'intersection des valeurs critiques avec les niveaux distinguables. Un exemple de fonction d'association est montré dans la Figure 4.3. Les positions des points de référence peuvent varier pour ajuster et éventuellement améliorer l'effet obtenu.

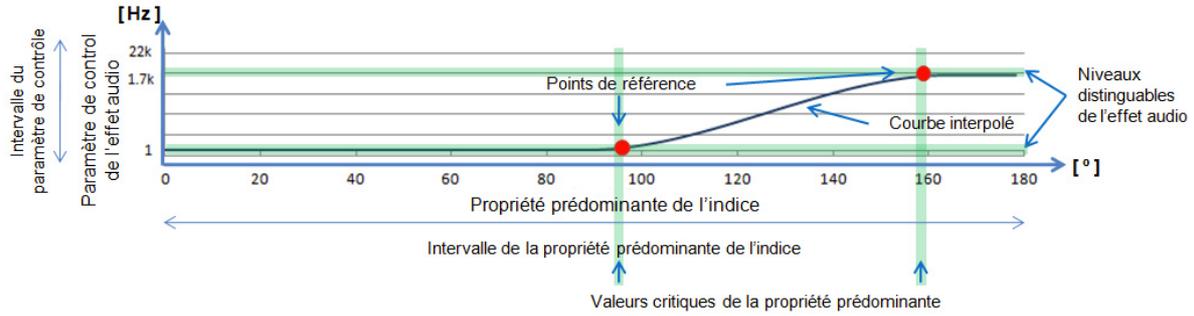


Figure 4.3. Exemple d'une fonction d'association d'un indice SASA et ses caractéristiques.

Dans ce paragraphe, nous décrivons un exemple d'indice sensoriel. Le problème de l'erreur inverse de devant/derrière est rencontré dans la localisation de sources sonores spatiales : l'utilisateur indique que la source sonore est devant lui alors que qu'elle se trouve derrière lui. Dans ce cas, il est possible d'ajouter un indice sensoriel qui permet à l'utilisateur de déterminer la position de la source sonore d'une façon correcte. Ainsi, l'effet audio choisi doit modifier le son d'une manière claire lorsque la provenance du son change passant de l'avant à l'arrière de l'utilisateur et inversement. Cet indice peut être nommé *Frontalité* parce qu'il permet de différencier les sources sonores en fonction de sa provenance frontale. Le paramètre prédominant de l'indice de Frontalité est l'azimut. Le paramètre spatial d'azimut de la source sonore sera associé au paramètre de contrôle de l'effet audio pour obtenir le résultat envisagé.

4.1.3. Méthode itérative de génération d'indices sensoriels SASA

Dans cette section, nous décrivons les méthodes utilisées pour générer chaque indice sensoriel SASA et leur intégration dans un groupe d'indices. Pour générer un indice il faut définir l'effet audio et les fonctions d'association à utiliser. Il faut également définir l'objectif de l'application, la tâche envisagée, la précision recherchée et les informations supplémentaires à transmettre à l'utilisateur. Des mesures empiriques seront effectuées au cours du processus de génération d'indices SASA pour confirmer que ces derniers remplissent les objectifs définis.

La méthode proposée pour la génération d'indices SASA permet de définir et ajuster les fonctions d'association. Cette méthode se contente de trouver une solution satisfaisante parmi les solutions possibles. La recherche de la meilleure combinaison d'indices et de fonctions d'association n'est pas un objectif vu le nombre important de solutions possibles et la difficulté de les évaluer empiriquement.

Cette méthode permet d'obtenir une solution acceptable basée sur un processus itératif, de critères définis à l'avance et sur de tests empiriques.

La méthode consiste à construire une première version de la fonction d'association et de tester de manière itérative son effet. Les tests empiriques consistent à évaluer l'effet obtenu par la fonction d'association avec un nombre limité d'utilisateurs (trois sujets dans notre étude) pour chaque itération. Les résultats obtenus sont analysés par rapport à la performance de l'utilisateur et son appréciation. La fonction d'association est

modifiée à chaque itération mais le processus est arrêté lorsque le concepteur considère que les objectifs recherchés sont atteints.

Les variables de performance mesurées dans les tests empiriques sont les erreurs angulaires de localisation de source sonore (azimut, élévation et angle global) et le temps d'accomplissement de la tâche. L'évaluation des informations supplémentaires transmises à l'utilisateur sont évaluées par une procédure d'évaluation subjective

Dans ce qui suit, nous décrivons les étapes de la méthode de génération d'un indice SASA et la détermination de sa fonction d'association. Les critères pris en compte sont décrits dans chaque étape. Dans la Figure 4.4, on présente le schéma de la méthode proposée.

- Etape 1. Spécification de l'information à transmettre ainsi que l'effet souhaité,
- Etape 2. Identification de la propriété prédominante de l'indice SASA. Cette variable peut être une des propriétés spatiales de la source sonore (azimut, élévation, angle global et profondeur) ou un paramètre relatif aux informations supplémentaires à transmettre à l'utilisateur,
- Etape 3. Identification des valeurs critiques de la propriété prédominante de l'indice,
- Etape 4. Choix de l'effet audio à appliquer au stimulus sonore. Cet effet audio doit produire un son facilement identifiable et il doit avoir au moins un paramètre de contrôle. L'effet audio choisi ne doit pas avoir été utilisé par un autre indice simultanément pour éviter la confusion,
- Etape 5. Sélection du paramètre de contrôle de l'effet audio à varier. Les autres paramètres de contrôle de l'effet audio doivent être maintenus aux valeurs par défaut ou à des valeurs convenables estimées par le concepteur. Le paramètre de contrôle choisi doit permettre d'obtenir les résultats les plus distinguables sans perturber le sujet,
- Etape 6. Identification des niveaux distinguables du paramètre de contrôle de l'effet audio appliqué au stimulus sonore,
- Etape 7. Définition et amélioration de la fonction d'association de l'indice SASA. Au cours des itérations, les positions des points de référence peuvent être ajustées selon l'estimation du concepteur,
- Etape 8. Réalisation du test empirique et détermination des améliorations possibles,
- Etape 9. Analyse de l'effet audio obtenu : nous avons décidé de rejeter l'effet audio généré si après six itérations on n'obtient pas la fonction d'association qui permet un résultat satisfaisant. Dans ce cas, le processus est repris à l'étape 4 pour choisir un autre effet audio. Si le concepteur constate des améliorations, le processus est repris à l'étape 7. Le processus est achevé lorsqu'aucune amélioration supplémentaire n'est nécessaire. A cette étape l'indice SASA, l'effet audio et la fonction d'association sont acceptés.

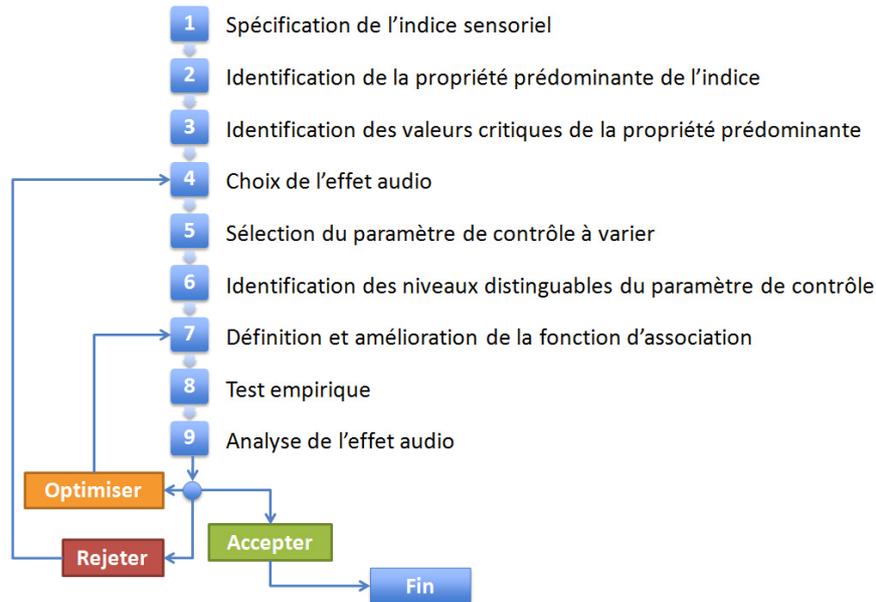


Figure 4.4. Méthode de génération d'un indice sensoriel SASA.

4.1.4. Intégration successive des indices SASA

Chaque indice doit être intégré dans un groupe d'indices SASA. Après l'intégration de chaque indice, il est nécessaire de vérifier qu'il n'entre pas en conflit avec les indices déjà intégrés dans le groupe SASA. Pour cela, nous avons défini un processus d'intégration décrit ci-dessous. La Figure 4.5 représente un schéma de la méthode proposée.

- Etape 1. Analyse de la tâche à réaliser,
- Etape 2. Identification des indices SASA à inclure dans la source sonore,
- Etape 3. Etablissement de la priorité de traitement des indices SASA : en général, les indices relatifs aux informations spatiales sont traités avant les informations spécifiques.
- Etape 4. Génération du premier indice SASA,
- Etape 5. Génération de l'indice SASA suivant,
 - Sous-étape 4a. Régénération de l'indice SASA qui implique la sélection d'un nouvel effet audio,
 - Sous-étape 4b. Amélioration de l'indice SASA qui consiste à ajuster les points de référence de la fonction d'association,
- Etape 6. Intégration de l'indice précédemment créé dans le groupe d'indices existant,
- Etape 7. Réalisation du test empirique et détermination des améliorations possibles,
- Etape 8. Analyse de l'effet obtenu : le nouvel indice est rejeté si après six itérations, le résultat obtenu n'est pas satisfaisant. Dans ce cas, le processus est repris à l'étape 4a. Le processus est repris à la

sous-étape 4b pour améliorer la fonction d'association. L'indice SASA est accepté si aucune amélioration supplémentaire n'est nécessaire,

Etape 9. Fin de processus. L'étape 4 est reprise s'il y a d'autres indices à intégrer, sinon le processus est terminé.

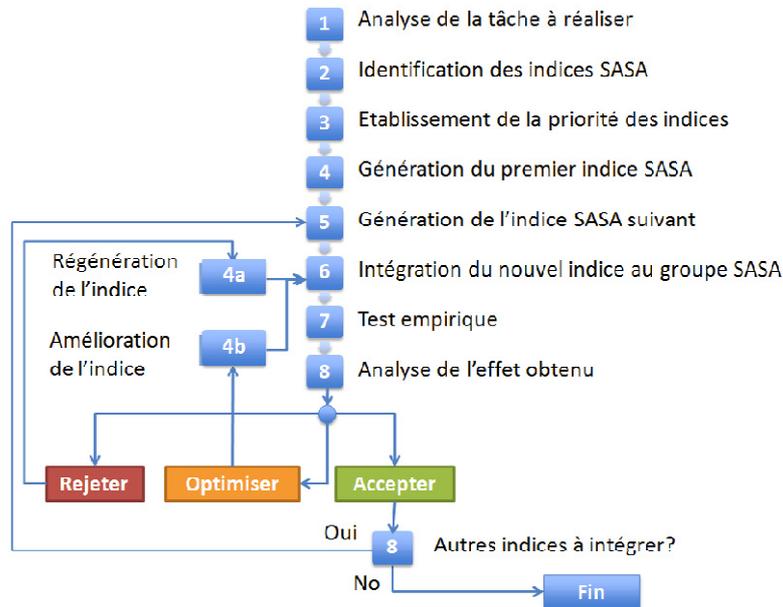


Figure 4.5. Méthode d'intégration d'un groupe d'indices SASA.

4.2. Proposition préliminaire : groupe d'indices SASA pour la localisation sonore précise

Dans cette section, on présente le groupe d'indices SASA de base que nous avons développé pour obtenir une localisation sonore précise. Les indices présentés et leur intégration ont été réalisés par les méthodes décrites précédemment. Un protocole expérimental a été défini pour valider la contribution de notre approche.

Le groupe d'indice SASA décrit dans cette section a pour objectif de permettre une localisation sonore précise. Ce groupe est appelé « SASA Localisation ».

4.2.1. Contexte

Plusieurs travaux ont été réalisés pour améliorer la localisation sonore basée sur la combinaison du son spatial et des indices supplémentaires (e.g. M. Grohn et Takala [1995], Tapio Lokki et Matti Grohn [2005], Marentakis et Brewster [2005] et Ménelas et al. [2009]). Cependant, les approches proposées se concentrent sur l'amélioration de la précision de la localisation en azimuth uniquement. Les travaux présentés dans cette section ont pour objectif d'améliorer la précision en azimuth et en élévation.

Notre approche se distingue de celles décrites dans la littérature par la structure des indices. Les approches existantes utilisent les effets audio typiques tels que le pitch, le gain et la variation de paramètres de synthèse de son. L'association des propriétés de l'indice et l'effet audio est définie par des équations fractionnelles. Les effets audio sont rarement combinés dans ces approches.

Dans notre approche, un indice SASA est structuré en fonction de l'objectif pour lequel il a été conçu. Il peut être combiné avec d'autres indices. Comparées aux équations fractionnelles, les fonctions d'association des indices SASA permettent de représenter des comportements plus complexes. L'intégration des indices SASA générés est effectuée par une procédure basée sur des tests empiriques.

4.2.2. Stimulus sonore

Les stimuli sonores utilisés dans le cadre de la recherche en stimulation auditive sont communément de quatre types : voix, musique, bruits et brefs (impulsions et tintements). Le stimulus sonore utilisé pour le groupe « SASA Localisation » est sélectionné en se basant sur des critères proposés par Silzle et al. [2004] et d'une série de tests que nous avons réalisés.

La forme d'onde et le spectre du stimulus sonore sélectionnés sont présentés dans la Figure 4.6. Le stimulus sonore est issu de la librairie sonore de Sound Rangers (Music [2009]). Il s'agit d'un stimulus sonore bref d'une durée de 0.5 secondes environ et il est reproduit continuellement. La forme d'onde du stimulus est un tintement. Les raisons et les critères principaux pris en compte pour sélectionner ce stimulus sont :

- un stimulus sonore bref est intelligible (facile à reconnaître),
- un stimulus sonore bref contient moins d'informations à décoder par l'auditeur par rapport à un stimulus vocal ou musical (Brewster [2003]),
- le tintement génère moins de perturbations que le bruit dont l'effet pourrait être confondu avec les situations problématiques comme l'interruption ou la perturbation dans le système de communication,
- les travaux de recherche publiés montrent que la localisation d'un tintement est plus précise que celle du bruit (Blauert [1997], page 39).
- Silzle et al. [2004] ont trouvé que l'utilisation du tintement permet de réduire la variabilité de l'erreur de localisation, en particulier pour l'élévation par rapport au bruit et au stimulus vocal.

La comparaison des différents types de stimuli sonores n'est pas l'objectif des travaux réalisés dans cette thèse. Par contre, l'utilisation d'un stimulus sonore, autre que celui employé dans cette thèse, risque de modifier la performance de l'utilisateur dans les applications que nous avons développées.

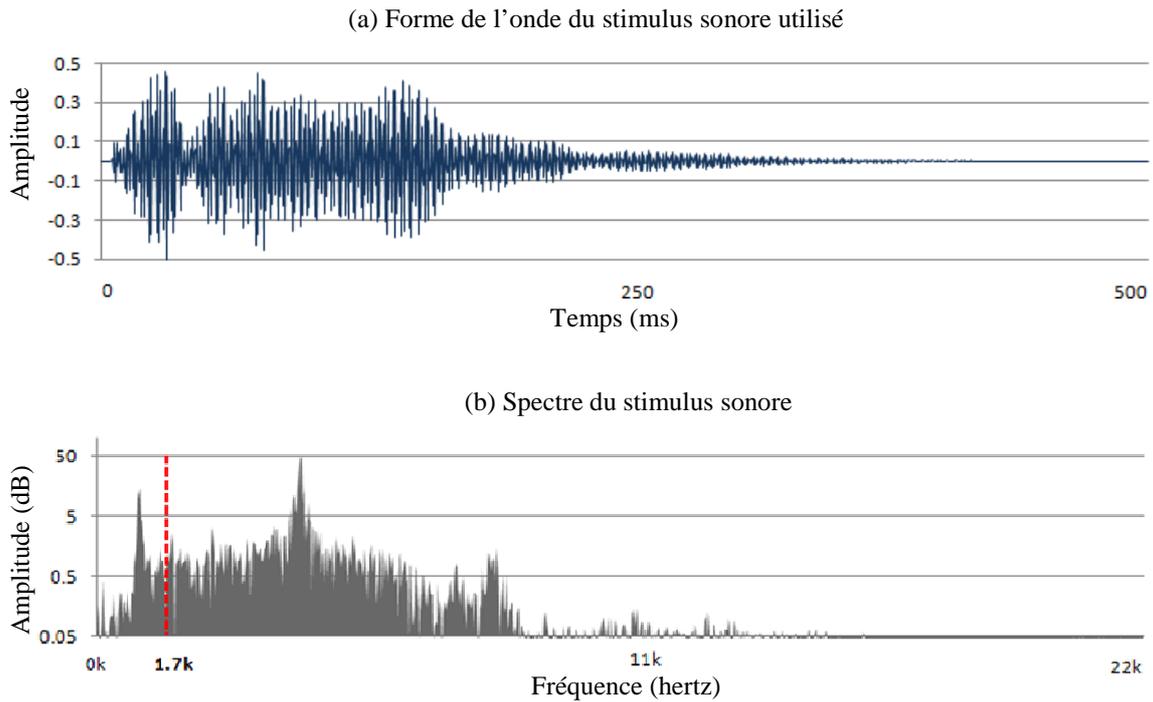


Figure 4.6. (a) La forme d'onde du stimulus sonore sélectionné. (b) Le spectre fréquentiel du stimulus sonore obtenu par la méthode TFR (Transformée de Fourier Rapide). La fréquence indiquée à 1.7 kHz représente la valeur de la fréquence de coupure utilisée comme valeur critique pour les indices SASA.

4.2.3. Indices de groupe SASA pour une localisation précise

Les caractéristiques des indices du groupe SASA sont décrites dans le Tableau 4.1. Pour chaque indice, l'effet audio choisi ainsi que le paramètre du groupe associé sont détaillés. Les paramètres du groupe d'indices SASA sont montrés dans le Tableau 4.2. Ces paramètres dépendent des différences angulaires entre l'orientation de la tête de l'utilisateur et la position de la source sonore.

L'horizontalité se réfère à la différence angulaire horizontale (erreur en azimut). Cet indice a été introduit pour fournir à l'utilisateur une meilleure sensation de la provenance de la source sonore. Le mécanisme de l'indice est décrit comme suit : l'auditeur positionne le centre de sa tête en face de la source sonore durant la phase de la localisation dans le plan horizontal. Pour faciliter cette étape, l'intensité du stimulus sonore augmente progressivement lorsque la source sonore se trouve dans l'intervalle $[-20,20]^\circ$ en azimut. La courbe qui caractérise l'association de l'effet avec le paramètre prédominant de l'indice est montrée dans la Figure 4.7a.

Tableau 4.1. Indices sensoriels du groupe SASA pour la localisation précise

Indices	Effet audio	Paramètre de l'effet audio	Intervalle du paramètre de l'effet audio	Propriété prédominante de l'indice
Verticalité	Réverbération	Niveau	$[-10k, 2k] \text{ mdB}$	Élévation
Horizontalité	Atténuation	Facteur	$[-40, 0] \text{ dB}$	Azimut
Frontalité	Filtre passe-haut	Fréquence de coupure	$[1, 22k] \text{ Hz}$	Azimut
Proximité angulaire	Filtre passe-bas	Fréquence de coupure	$[200, 22k] \text{ Hz}$	Angle global
Proximité angulaire renforcée	Cône de diffusion	Facteur d'atténuation	$[-10, 0] \text{ dB}$	Angle global

Tableau 4.2. Paramètres des indices SASA du groupe pour la localisation précise

Nom du paramètre	Effet audio	Intervalle du paramètre
Élévation	Différence angulaire verticale	$[-90, 90]^\circ$
Azimut	Différence angulaire horizontale	$[-180, 180]^\circ$
Angle global	Différence angulaire globale	$[0, 180]^\circ$

La frontalité se réfère à la distinction entre les sources sonores situées devant ou derrière l'auditeur. Cet indice a été introduit pour éliminer les erreurs inverses devant / derrière. Le mécanisme consiste en un filtre fréquentiel de type passe-haut qui modifie le son lorsque la source sonore est derrière l'auditeur. Ainsi, les composantes fréquentielles basses sont atténuées pour les sons situés derrière l'utilisateur. Cet effet audio permet de masquer le son pour faciliter la reconnaissance de cette information par l'utilisateur. La courbe de la fonction d'association correspondante est montrée dans la Figure 4.7b.

La verticalité se réfère à la distinction entre les sources sonores qui se trouvent au dessous ou au dessus du plan horizontal du repère de l'auditeur. L'effet de réverbération est utilisé pour indiquer si la source sonore se trouve au dessous de l'auditeur. Cet indice permet d'éliminer les erreurs inverses de type haut / bas. La fonction d'association correspondante est montrée dans la Figure 4.7c.

Si la source sonore se trouve dans l'intervalle dont la différence angulaire globale est inférieure à 10 degrés, la proximité angulaire permet de fournir à l'utilisateur une indication de rapprochement jusqu'à un degré. Cet indice est implanté à l'aide d'un filtre fréquentiel passe-bas. Le mécanisme de cet indice fait que les basses fréquences du son émis sont progressivement moins atténuées au fur et à mesure que l'auditeur s'approche de la source. La fonction d'association correspondante est montrée dans la Figure 4.7d. Si l'indice d'horizontalité permet de mieux identifier la localisation en azimuth, l'indice de proximité angulaire permet de le faire en azimuth et en élévation en même temps. Les filtres passe-haut et passe-bas utilisés dans cette implantation sont de deuxième ordre et ils atténuent les fréquences filtrées jusqu'à -40 dB.

La proximité angulaire renforcée est une information qui permet de consolider l'indice de proximité angulaire par une augmentation progressive de l'intensité du son émis. Pour ceci, nous avons utilisé un cône de

diffusion. Ce dernier possède deux paramètres : un angle extérieur et un angle intérieur (Doerr et al. [2007]). Un cône de diffusion atténue le son à un niveau défini lorsque l'angle global est supérieur à la valeur de l'angle extérieur. Le son est atténué proportionnellement lorsque l'angle global se trouve entre la valeur de l'angle intérieur et celle de l'angle extérieur. Les paramètres pour le cône de diffusion utilisé sont fixés à 1.0° pour l'angle intérieur, 5.0° pour l'angle extérieur et un facteur d'atténuation égal à -5dB. La fonction d'association correspondante est montrée dans la Figure 4.7e.

L'association des indices SASA à un stimulus sonore discret $x[n]$ est représentée symboliquement par \wp comme suit :

$$x^{SASA}[n] = \wp\{x[n]\} \quad (4.1)$$

où $x^{SASA}[n]$ est le stimulus sonore enrichi par les indices SASA.

4.2.4. Modèle de directionalisation utilisé

La méthode de directionalisation utilisée pour les indices sensoriels est celle de la Puissance Constante. Cette méthode se base sur des courbes sinusoïdales pour contrôler l'amplitude du signal dans les différents canaux de sortie pour maintenir une puissance totale constante (Roads [1996]). Les facteurs d'amplitude pour les sorties binaurales sont déterminés comme suit :

$$A_L = \frac{\sqrt{2}}{2} (\cos \varphi + \sin \varphi) \quad (4.2)$$

$$A_R = \frac{\sqrt{2}}{2} (\cos \varphi - \sin \varphi) \quad (4.3)$$

où φ est l'azimut, et

A_L et A_R sont les amplitudes du canal gauche et du canal droit.

Cette méthode fournit des indices interauraux uniquement. La sortie est invariante par rapport à l'élévation qui sera fournie par les indices SASA.

Il existe d'autres méthodes de positionnement géométrique pour la directionalisation tels que l'*Ambiosonics*, le Positionnement d'Amplitudes Basé sur Vecteurs (VBAP : Vector Based Amplitude Panning) et la Synthèse de Champs d'Onde (WFS : Wave Field Synthesis). *Ambiosonics* est une technique d'utilisation de microphones pour simuler la synthèse de son spatial (Gerzon [1973]). VBAP est une méthode basée sur des vecteurs (Pulkki [2001]). La technique de WFS (Berkhout [1988]) est basée sur le principe de l'holophonie qui implique l'utilisation d'un grand nombre de haut-parleurs pour reproduire des champs sonores à haute fidélité.

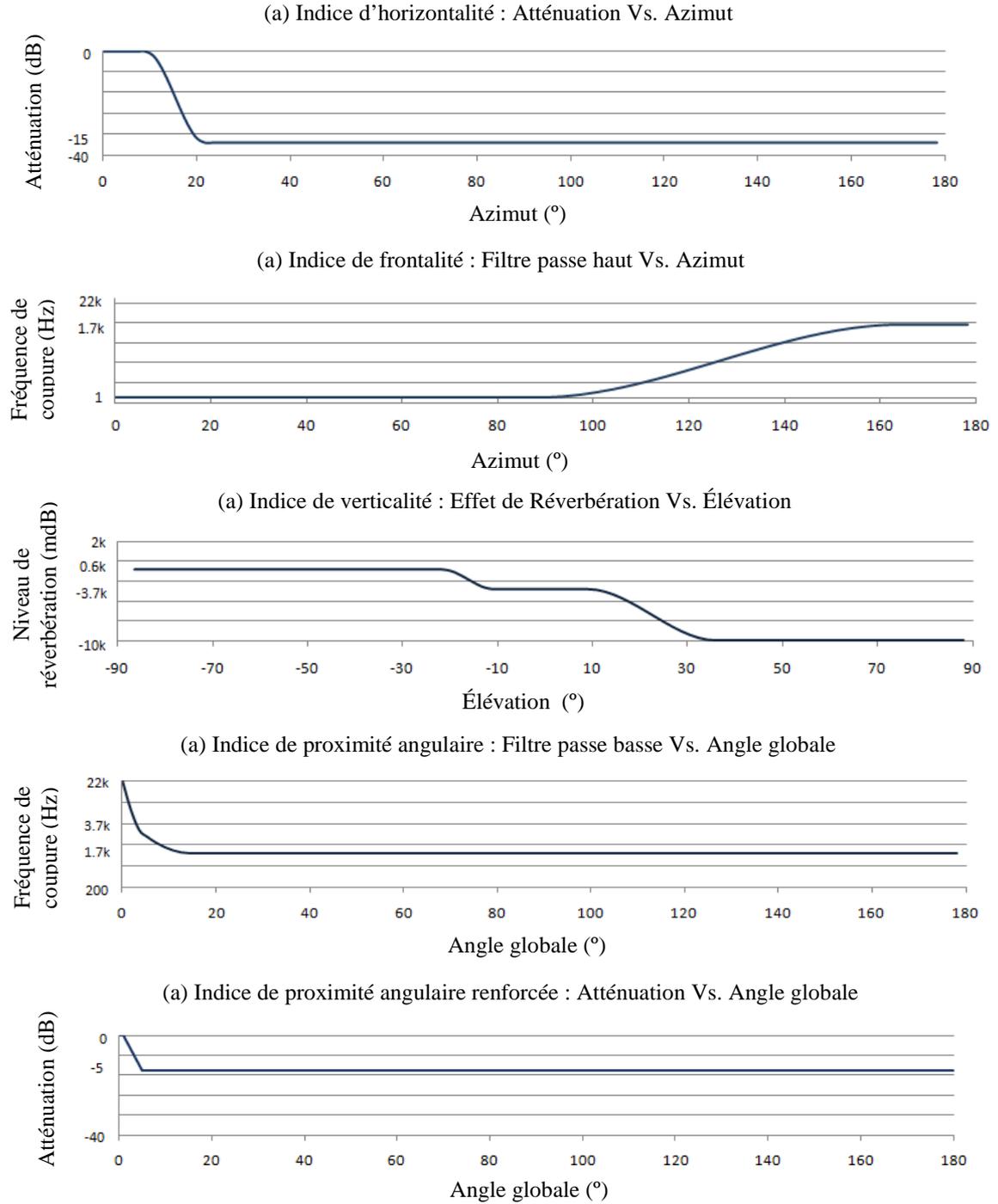


Figure 4.7. Fonctions d'association des indices sensoriels SASA. Les axes verticaux des figures (b) et (d) sont tracés en échelle logarithmique.

Les sorties binaurales $x_L^{SASA}[n]$ et $x_R^{SASA}[n]$ de l'approche SASA pour les canaux gauche et droit d'un casque sont respectivement calculées comme suit :

$$x_L^{SASA}[n] = A_L \cdot x^{SASA}[n] \quad (4.4)$$

$$x_R^{SASA} [n] = A_R \cdot x^{SASA} [n] \quad (4.5)$$

où $x^{SASA} [n]$ est le stimulus sonore enrichi avec les indices SASA,

A_L et A_R sont les coefficients d'amplitude de la méthode de positionnement géométrique de Puissance Constante,

\cdot représente l'opérateur de multiplication d'une valeur scalaire par un signal discret.

4.2.5. Protocole de validation expérimentale

L'objectif de cette procédure expérimentale est d'étudier la performance de l'utilisateur dans la tâche de localisation de source sonore à travers une évaluation objective.

4.2.5.1. Contexte

Dans cette expérimentation, les participants doivent indiquer avec la tête la provenance de la source sonore dans l'espace 3D. Tourner la tête vers la source sonore est un geste hautement écologique (Simon Carlile et al. [1997]). Ce geste est aussi utilisé très couramment dans le domaine de la recherche de son spatial (Makous et Middlebrooks [1990]). Néanmoins, il existe d'autres approches utilisant différents mécanismes tels que les interfaces graphiques, l'estimation verbale, le pointé par la main et le pointé par les outils manuels (Begault et al. [2001], E. M. Wenzel et al. [1993], Jiann-Rong Wu et al. [1997] et Oldfield et Parker [1984]).

Le mouvement de la tête a une importance particulière dans l'écoute humaine comme il a été démontré par Begault et al. [2001] et par Jiann-Rong Wu et al. [1997]. Dans ce sens, nous avons décidé d'incorporer le suivi de mouvement de la tête dans notre plan d'expérimentation.

Pour la restitution sonore, nous avons choisi d'utiliser un casque fermé. Ce choix est motivé principalement par les recommandations faites par Begault [2005] : indépendance de l'acoustique du contexte réel et meilleure fidélité pour le rendu sonore. La Figure 4.8a présente l'interface graphique de l'application, affichée pour l'expérimentateur uniquement. L'utilisateur porte un casque pour l'écoute binaurale, un masque sur les yeux et l'interface sans fils Wiimote dans la main (Figure 4.8b).

4.2.5.2. Description des essais expérimentaux

Avant de participer aux expériences, le fonctionnement des indices SASA est expliqué aux sujets par l'expérimentateur. Les sujets sont informés que l'objectif des indices SASA est de permettre à l'utilisateur de reconnaître les informations spatiales et les informations spécifiques d'une façon intuitive.

Le Tableau 4.3 présente les valeurs des positions utilisées pour les essais expérimentaux. Les provenances hémisphériques (gauche ou droite, devant ou derrière et haut ou bas) de chaque position sont indiquées.

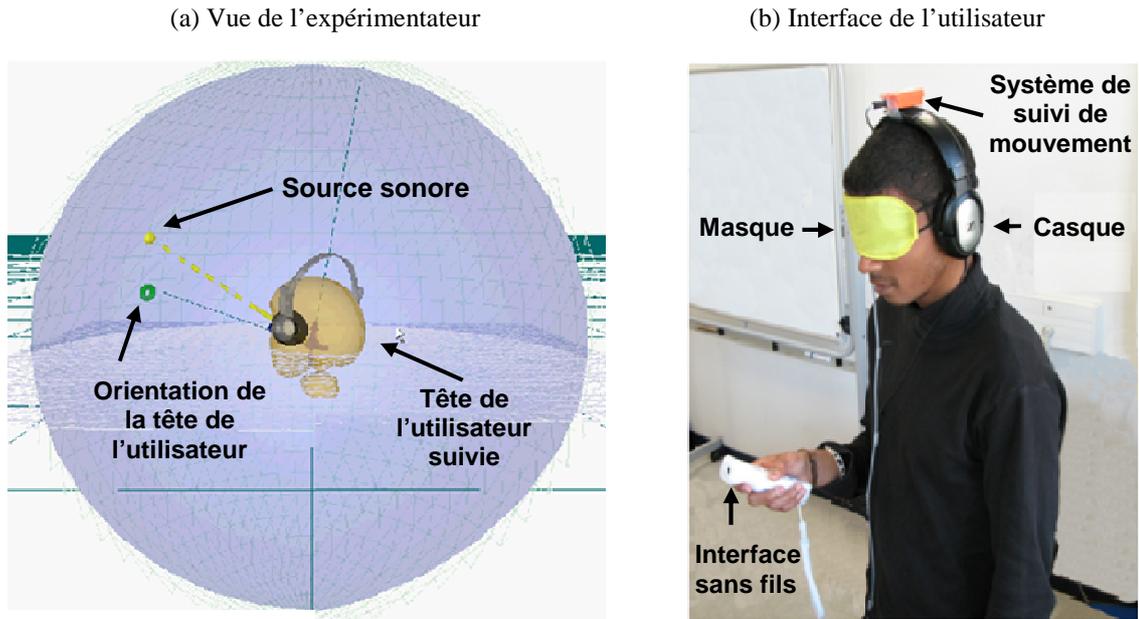


Figure 4.8. (a) Interface graphique affichée pour l'expérimentateur unique ; (b) Photo d'un participant en train de réaliser l'expérience.

Pour que l'utilisateur puisse se familiariser avec la tâche et l'interface, il doit exécuter deux essais préliminaires (qui ne seront pas pris en compte dans l'analyse des résultats). Le sujet réalise le premier essai préliminaire sans mettre le masque sur les yeux. Le deuxième essai préliminaire ainsi que les essais suivants sont réalisés avec le masque sur les yeux du sujet.

Tableau 4.3. Essais expérimentaux : position de la source sonore à chercher

Essai	Azimet (φ_0)	Élévation (δ_0)	Avant / Derrière	Gauche / Droite	Haut / Bas
1	5.0	15.0	Devant	Droite	Haut
2	45.0	5.0	Devant	Droite	Haut
3	45.0	-5.0	Devant	Droite	Bas
4	5.0	-30.0	Devant	Gauche	Bas
5	-5.0	30.0	Devant	Gauche	Haut
6	-60.0	-30.0	Devant	Gauche	Bas
7	60.0	30.0	Devant	Droite	Haut
8	90.0	30.0	Derrière	Droite	Haut
9	-90.0	-30.0	Derrière	Gauche	Bas
10	-135.0	-5.0	Derrière	Gauche	Bas
11	135.0	5.0	Derrière	Droite	Haut

Les utilisateurs doivent appuyer sur un bouton de l'interface Wiimote lorsqu'ils considèrent que la source sonore est localisée.

4.2.5.3. *Stratégie de localisation*

L'expérimentateur donne des recommandations spécifiques aux utilisateurs avant de commencer l'expérience. Ces recommandations ont pour objectif d'aider l'utilisateur dans l'accomplissement de sa tâche :

1. effectuer un mouvement lent de la tête et du corps pour localiser la source sonore,
2. commencer par effectuer une recherche dans le plan horizontal,
3. effectuer une recherche dans le plan vertical après avoir déterminé la provenance de la source sonore dans le plan horizontal,
4. maintenir l'orientation de la tête dans de la plan horizontal pendant la recherche dans le plan vertical,
5. utiliser la Wiimote pour indiquer au système que la source sonore est localisée et rester en position statique jusqu'à la vibration du Wiimote,
6. revenir à la recherche dans le plan horizontal en cas de sensation de confusion.

Il a été demandé aux sujets d'appliquer la stratégie recommandée pour éviter que la stratégie de localisation de la source sonore ne constitue un facteur expérimental elle-même.

4.2.5.4. *Groupe de participants et mesures expérimentales*

Le groupe de participants était composé de vingt sujets (dix hommes et dix femmes). Ils sont tous droitiers et âgés entre 20 et 25 ans. Ils sont des étudiants et ils ont déclaré qu'ils ont une vision et une audition normales.

La performance de l'utilisateur est mesurée par les variables objectives suivantes : le temps d'accomplissement de la tâche (en secondes) et les erreurs angulaires (en degrés). Les erreurs angulaires prises en compte sont : l'erreur en azimuth (Figure 4.9a), l'erreur en élévation (Figure 4.9b) et l'erreur angulaire globale (Figure 4.9c). L'analyse de l'erreur angulaire a été effectuée en tenant compte de sa valeur absolue.

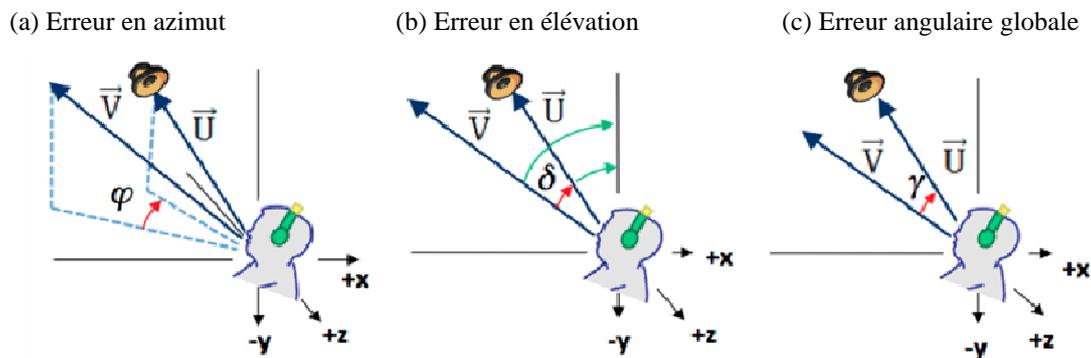


Figure 4.9. Erreurs angulaires mesurées (en rouge).

L'erreur angulaire globale γ est calculée comme suit :

$$\gamma = \arccos\left(\frac{\vec{U} \cdot \vec{V}}{|\vec{U}||\vec{V}|}\right) \quad (4.6)$$

où \vec{U} et \vec{V} sont les vecteurs qui représentent la position de la source sonore et celle de la tête de l'utilisateur respectivement.

4.2.6. Résultats et analyse des données

Dans cette section, on présente les mesures statistiques descriptives des différents groupes de données (définis en fonction des facteurs expérimentaux). Ensuite, nous présentons les tests statistiques pour déterminer la nature des distributions fréquentielles (connus également comme tests de qualité d'ajustement) pour chaque mesure expérimentale. Finalement, on présente une analyse de la signification statistique des différences entre l'effet des différents facteurs expérimentaux.

4.2.6.1. Mesures statistiques descriptives

Le Tableau 4.4 présente les mesures statistiques descriptives des facteurs expérimentaux. Les mesures statistiques montrées sont la moyenne, l'écart type, la médiane et l'information de quartiles.

La répartition autour de zéro pour les valeurs non absolues des erreurs angulaires sont insignifiants pour l'azimut (médiane = 0.50°) et pour l'élévation (médiane = -0.16°). Compte tenu de ce résultat et du fait que la valeur absolue est plus significative pour l'analyse, nous n'avons pas pris en compte le signe de cette valeur.

Chaque ligne du Tableau 4.4 comprend un facteur expérimental étudié. Ces facteurs sont le sexe des sujets et la provenance hémisphérique latérale (gauche et droite), horizontale (devant et derrière) et verticale (haut et bas) de la source sonore.

Les mesures angulaires sont affectées particulièrement par la précision du système de suivi. L'erreur globale de mesure du système de suivi est de $\pm 0.5^\circ$. Aucune erreur inverse de localisation n'a été observée pendant les expériences. L'approche SASA a permis d'éliminer les ambiguïtés de la localisation sonore (erreurs inverses).

La moyenne et la médiane du temps d'accomplissement de la tâche est de 26.7 et 22.5 secondes respectivement avec un écart type de 15.2 secondes. Le temps d'accomplissement de la tâche obtenu est supérieur à celui obtenu avec les approches citées dans le Chapitre 2. Pour la variable d'erreur angulaire globale, la moyenne, la médiane et l'écart type sont de 3.2, 2.6 et 2.0 degrés respectivement. La précision angulaire obtenue est supérieure à celle obtenue avec les approches existantes. La moyenne, la médiane et l'écart type pour l'erreur absolue en azimut sont 2.1, 1.5 et 1.8 respectivement ; celles d'élévation sont 2.1, 1.6 et 1.7.

Dans les études rencontrées dans la littérature, la précision en azimut est significativement supérieure à celle en élévation. Il est important de noter que les résultats obtenus montrent que la technique SASA permet d'obtenir une précision en élévation similaire à celle obtenue en azimut.

Tableau 4.4. Mesures statistiques des données obtenues

Groupe	Variable	N ^{bre}	Moyenne	Médiane	Quartile			Écart type
					Inf.	Sup.	Inter.	
Tous les groupes	Temps d'accomplissement (s)	220	26.78	22.51	15.32	34.34	19.02	15.47
	Erreur angulaire globale (°)	220	3.24	2.65	1.75	4.34	2.59	2.04
	Erreur en azimut (°)	220	2.10	1.52	0.76	2.84	2.09	1.84
	Erreur en élévation (°)	220	2.13	1.65	0.83	3.06	2.23	1.79
Masculin	Temps d'accomplissement (s)	110	2.87	2.39	1.59	3.77	2.18	1.78
	Erreur angulaire globale (°)	110	2.87	2.39	1.59	3.77	2.18	1.78
	Erreur en azimut (°)	110	2.03	1.50	0.83	2.81	1.98	1.72
	Erreur en élévation (°)	110	1.75	1.38	0.65	2.28	1.63	1.44
Féminin	Temps d'accomplissement (s)	110	28.77	24.91	16.81	37.24	20.43	16.05
	Erreur angulaire globale (°)	110	3.61	3.15	1.85	5.12	3.27	2.22
	Erreur en azimut (°)	110	2.16	1.65	0.69	2.89	2.20	1.95
	Erreur en élévation (°)	110	2.51	2.04	0.94	3.69	2.75	2.02
Devant	Temps d'accomplissement (s)	140	24.49	20.45	14.36	29.94	15.58	14.89
	Erreur angulaire globale (°)	140	3.12	2.65	1.60	4.29	2.69	1.91
Derrière	Temps d'accomplissement (s)	80	30.78	27.12	19.12	39.74	20.62	15.73
	Erreur angulaire globale (°)	80	3.44	2.64	1.82	4.71	2.88	2.24
Gauche	Temps d'accomplissement (s)	100	29.09	23.85	17.29	38.33	21.04	16.80
	Erreur angulaire globale (°)	100	3.26	2.71	1.73	4.55	2.82	2.03
Droite	Temps d'accomplissement (s)	120	24.85	21.59	14.54	29.38	14.85	14.04
	Erreur angulaire globale (°)	120	3.21	2.61	1.75	4.32	2.57	2.05
Haut	Temps d'accomplissement (s)	120	25.83	21.45	14.62	33.53	18.91	15.16
	Erreur angulaire globale (°)	120	3.41	2.74	1.98	4.68	2.70	2.07
Bas	Temps d'accomplissement (s)	100	27.91	23.37	17.14	36.41	19.27	15.82
	Erreur angulaire globale (°)	100	3.02	2.45	1.42	4.12	2.70	1.99

La précision du système de suivi de la tête est de $\pm 0.5^\circ$. Les mesures statistiques de l'erreur angulaire (azimut, élévation et angle global) ont été calculées sur la base des valeurs absolues de l'erreur obtenue.

La Figure 4.10a montre la moyenne et les boîtes à moustaches pour la variable de temps d'accomplissement de la tâche. Il existe une différence importante des temps : 24.6 s pour les sujets de sexe masculin et 28.7 pour les sujets de sexe féminin.

La Figure 4.10a montre la moyenne et les boîtes à moustaches pour la variable d'erreur angulaire globale. Il existe une différence potentiellement importante entre les hommes et les femmes (2.8° et 3.6° respectivement). La signification statistique des différences observées sera vérifiée plus tard dans le texte.

La Figure 4.11 montre la moyenne et les boîtes à moustaches de l'erreur angulaire absolue en azimut et en élévation pour les sujets masculins et les sujets féminins. Il est à noter que les sujets de sexe masculin ont enregistré une erreur en élévation nettement inférieure à celle enregistrée par les sujets de sexe féminin (masculin : 1.7° vs. féminin : 2.5°) ; l'erreur en azimut est similaire pour les deux groupes de sujets (masculin : 2.0° vs. féminin : 2.1°).

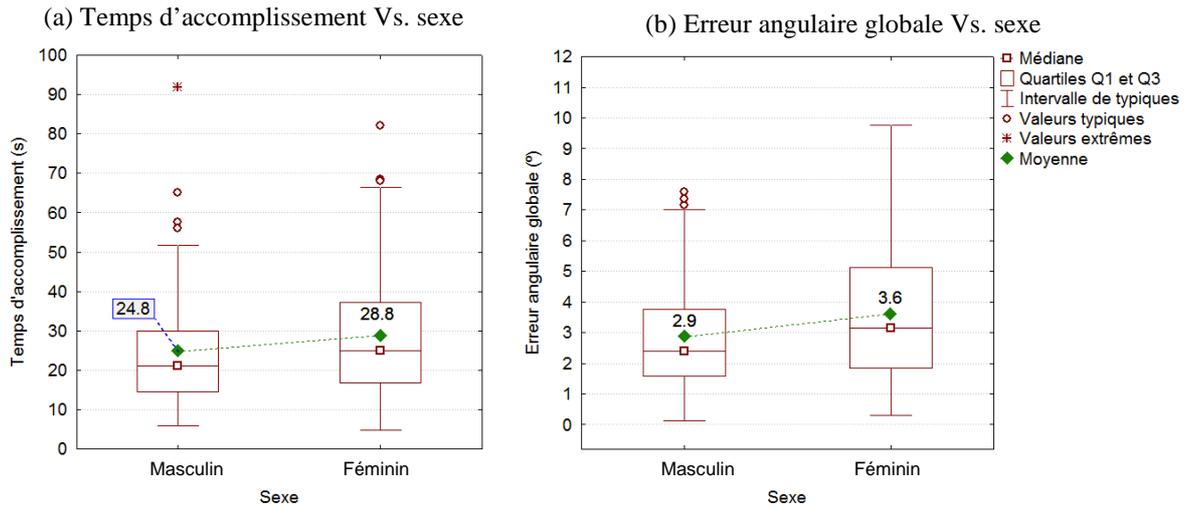


Figure 4.10. (a) Temps d'accomplissement de la tâche et (b) erreur angulaire globale selon le sexe des utilisateurs.

La performance individuelle en termes de temps d'accomplissement de la tâche et de l'erreur angulaire globale sont montrées dans la Figure 4.12a et la Figure 4.12b respectivement.

Nous avons réalisé une étude expérimentale de type intra-sujets avec essais multiples. Cette expérimentation est appropriée pour réduire la variabilité intra-sujets. La réalisation d'essais multiples réduit la variabilité intra-sujets.

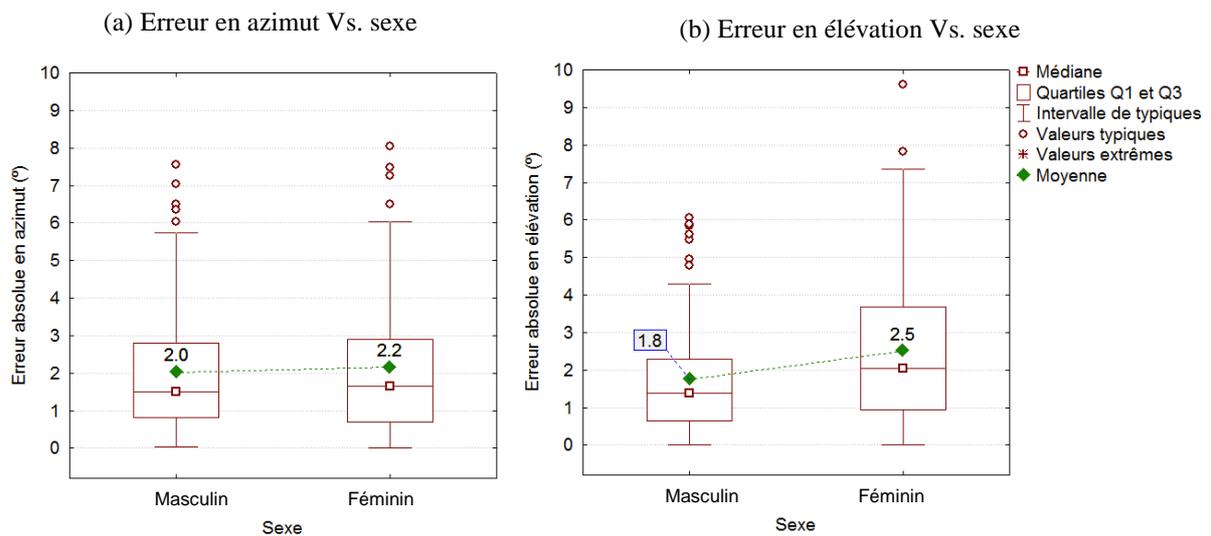


Figure 4.11. (a) Erreur absolue en azimut et (b) erreur absolue en élévation selon le sexe des utilisateurs.

La performance individuelle en termes de temps d'accomplissement de la tâche et de l'erreur angulaire globale sont montrées dans la Figure 4.12a et la Figure 4.12b respectivement.

Nous avons réalisé une étude expérimentale de type intra-sujets avec essais multiples. Cette expérimentation est appropriée pour réduire la variabilité intra-sujets. La réalisation d'essais multiples réduit la variabilité intra-sujets.

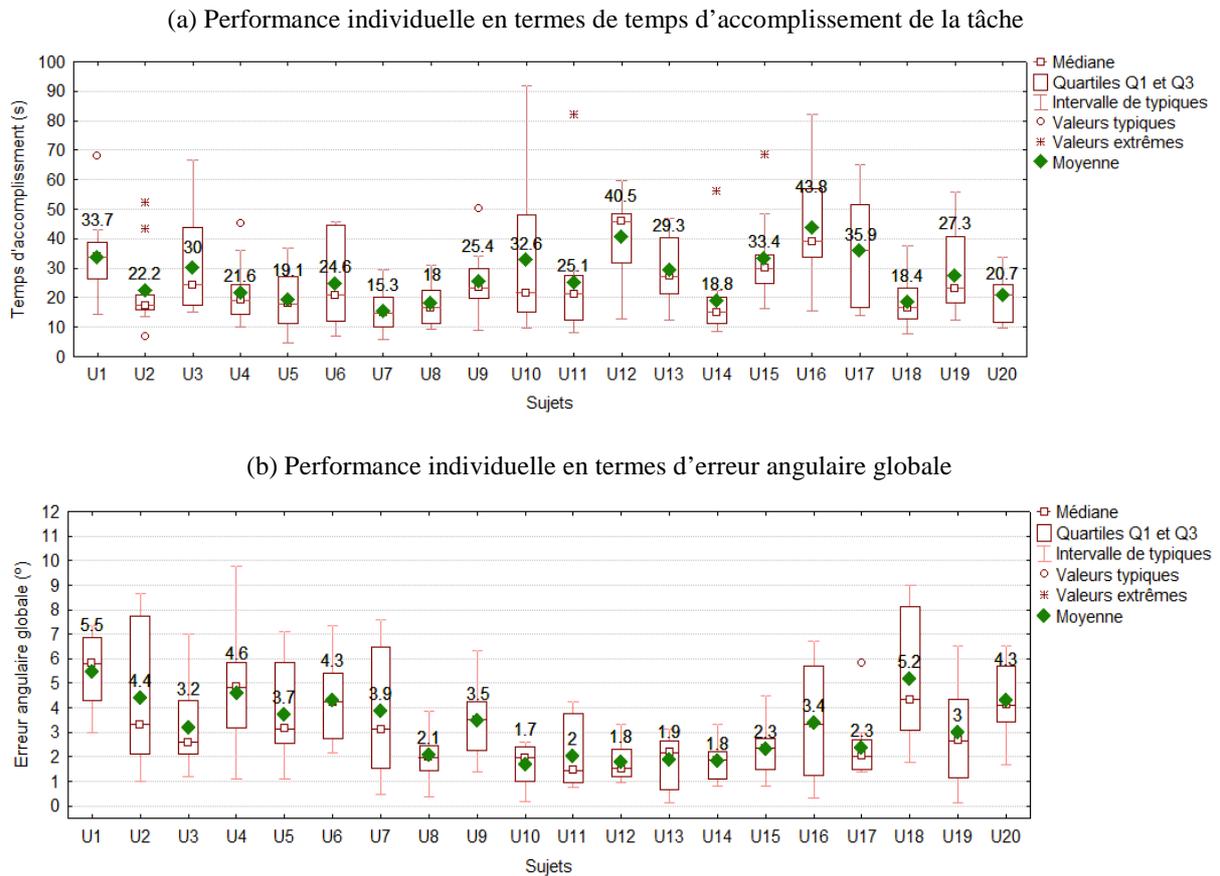


Figure 4.12. Performance individuelle du (a) temps d'accomplissement de la tâche et (b) l'erreur angulaire globale.

D'après la Figure 4.12b, il ressort que les sujets qui ont enregistré une faible précision angulaire (erreur > 3,0° : sujets U1, U2, U3, U4, U5, U6, U7, U9, U16, U18, U19 et U20) ont enregistré une variabilité intra-sujets plus importante que les sujets qui ont montré une meilleure précision. Ce résultat montre l'existence d'une relation directe entre la précision angulaire de la localisation et la variabilité intra-sujets.

D'après la Figure 4.12a, on note que les sujets qui ont enregistré un temps d'accomplissement de la tâche supérieur à 27 secondes (U1, U3, U10, U12, U13, U15, U16, U17 et U19) ont une variabilité intra-sujets plus importante comparée aux sujets qui ont enregistré un temps d'accomplissement de la tâche inférieur. Ceci démontre qu'il existe une relation entre le temps d'accomplissement de la tâche et la variabilité intra-sujets

4.2.7. Distribution fréquentielle des données

Le Tableau 4.5 montre les résultats des tests de qualité d'ajustement effectués pour chaque variable expérimentale. Ces tests ont été effectués pour déterminer la nature de la distribution de chaque variable. Pour effectuer la comparaison, nous avons utilisé le test de Kolmogorov-Smirnov (test K-S).

La détermination de la nature de la distribution des variables expérimentales est importante pour le choix de la méthode d'analyse statistique. Cette méthode permet de déterminer la signification des différences éventuelles entre les facteurs expérimentaux étudiés.

Les résultats du test K-S montrent qu'aucune des variables expérimentales ne suit la loi de la distribution normale. D'après le Tableau 4.5, le temps d'accomplissement de la tâche et l'erreur angulaire globale suivent la distribution log-normale (voir Figure 4.13a et Figure 4.13b). Les résultats montrent également que l'erreur absolue en azimut et en élévation suit la distribution gamma (voir Figure 4.13c et Figure 4.13d). Puisqu'aucune de ces quatre variables ne suit la loi normale et que leurs distributions ne sont pas symétriques, on ne peut pas utiliser les méthodes statistiques paramétrées telle que l'analyse par ANOVA.

Tableau 4.5. Tests de qualité d'ajustement.

Fonction	Variable	Moyenne	Variancé	Moyenne ajustée	Variancé ajusté	K-S distribution	p	Ajustée (o/n) ?
Normale	Temps d'accomplissement (s)	26.77	239.19	26.77	239.19	0.132	<0.01	N
	Erreur angulaire globale (°)	3.23	4.16	3.23	4.16	0.123	<0.01	N
	Erreur en azimut (°)	2.09	3.37	2.09	3.37	0.129	<0.01	N
	Erreur en élévation (°)	2.13	3.20	2.13	3.20	0.136	<0.01	N
Log-normale	Temps d'accomplissement (s)	26.77	239.19	3.13	0.31	0.040	>0.05	O
	Erreur angulaire globale (°)	3.23	4.16	0.94	0.55	0.060	>0.05	O
Fonction	Variable	Moyenne	Variancé	Moyenne ajustée	Variancé ajusté	K-S distribution	p	Ajustée (o/n) ?
Gamma	Temps d'accomplissement (s)	2.09	3.37	1.83	1.14	0.053	>0.05	O
	Erreur angulaire globale (°)	2.13	3.20	1.75	1.21	0.042	>0.05	O

La méthode de comparaison utilisée est le test de Kolmogorov-Smirnov (K-S) à un niveau de confiance de 95%.

4.2.8. Analyse statistique

Pour le choix de la méthode d'analyse statistique appropriée, il est nécessaire de prendre en compte les caractéristiques de l'expérimentation ainsi que la nature de la distribution des variables expérimentales. Nous

avons utilisé la méthode des Equations Généralisées d'Estimation (GEE) qui permet de d'analyser des données avec des distributions de type log-normale, gamma et exponentielle.

La Tableau 4.6 montre les résultats de l'analyse statistiques pour les différentes variables et les différents facteurs expérimentaux. L'effet du sexe n'est pas statistiquement significatif pour le temps d'accomplissement de la tâche ($p=0.258$) ni pour l'erreur angulaire globale ($p=0.142$).

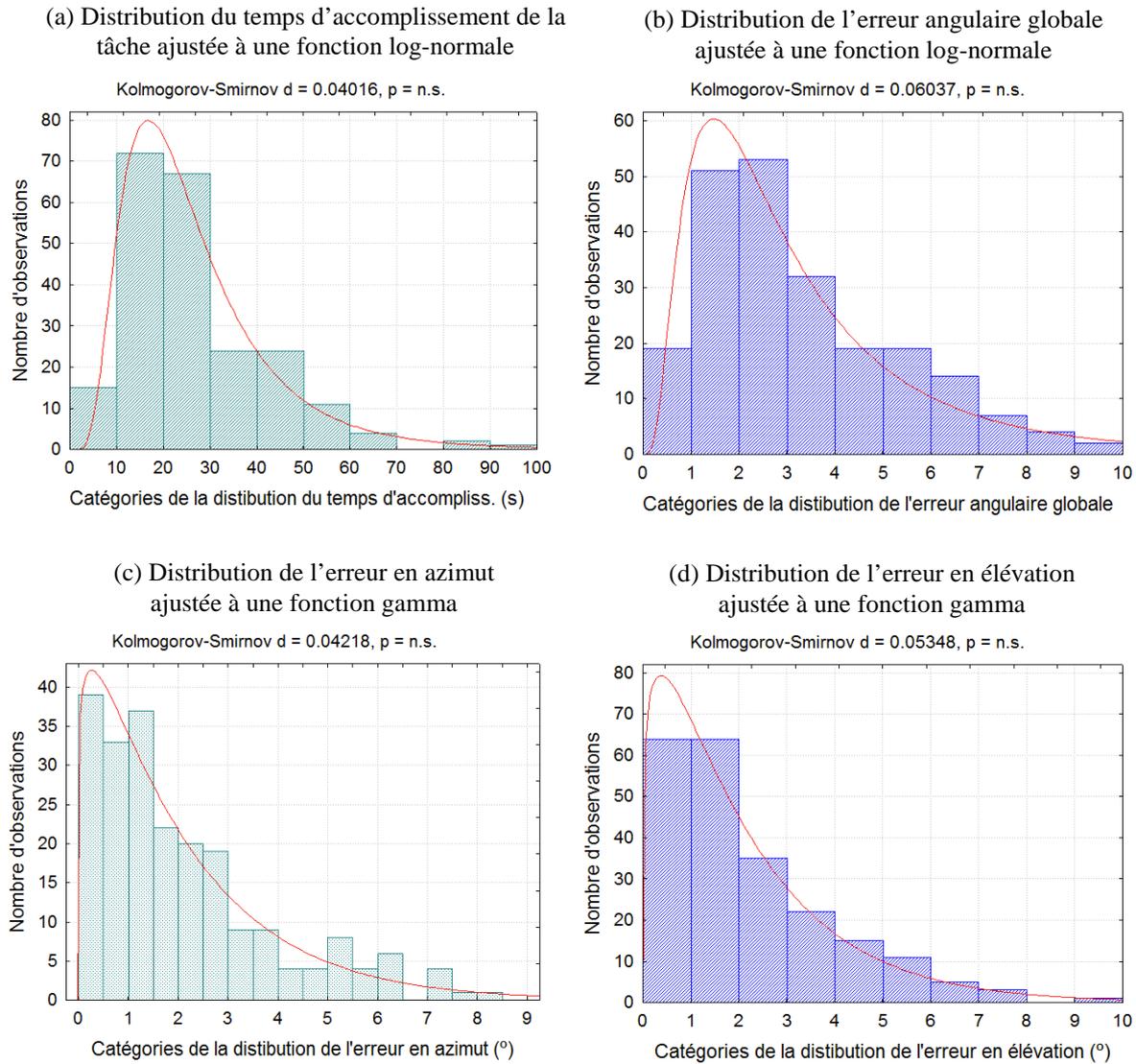


Figure 4.13. Distribution de la fréquence pour : (a) temps d'accomplissement de la tâche, (b) erreur angulaire globale, (c) erreur absolue en azimut et (d) erreur absolue en élévation.

Les sujets de sexe masculin ont enregistré une erreur en élévation inférieure à celle enregistrée pour les sujets de sexe féminin. Cette différence est statistiquement significative ($p=0.040$). Par contre, cette différence par rapport à l'erreur en azimut n'est pas statistiquement significative ($p=0.599$).

Ce résultat nous permet d'établir qu'il n'existe pas de différence significative entre la performance des sujets de sexe masculin et celle des sujets de sexe féminin. Par conséquent, nous considérons qu'il n'est pas nécessaire de prendre en compte le sexe du sujet dans les prochaines analyses de cette thèse.

La différence de temps d'accomplissement de la tâche pour les provenances hémisphérique latérale (gauche et droite) et verticale (haut et bas) n'est pas statistiquement significative ($p=0.070$ et $p=0.896$ respectivement). Par contre, pour la provenance hémisphérique frontale (devant et derrière), cette différence est statistiquement significative ($p=0.007$). Ceci implique que si le temps d'accomplissement de la tâche est un critère important dans une application de son spatialisé, il est recommandé d'utiliser des sources sonores localisées devant l'écouteur.

La différence de l'erreur angulaire globale pour les provenances hémisphériques frontale et latérale n'est pas statistiquement significative ($p=0.242$ et $p=0.453$ respectivement). Néanmoins, la différence entre les provenances verticales (haut et bas) est significative ($p=0.021$). Ce résultat montre que les sources sonores positionnées en haut de l'auditeur sont localisées avec une précision moindre.

Tableau 4.6. Analyse de l'effet des facteurs expérimentaux par l'Analyse de Mesures Répétées avec le modèle d'Equations Généralisées d'Estimation

Facteur	Variable	Modèle	Wald Chi-square (χ^2)	DDL	p	Significatif (o/n) ?
Sexe	Temps d'accomplissement (s)	Log-normale	1.282	1	.258	N
	Erreur angulaire globale (°)	Log-normale	2.153	1	.142	N
	Erreur en azimuth (°)	Gamma	0.276	1	.599	N
	Erreur en élévation (°)	Gamma	4.199	1	.040	O
Avant / Derrière	Temps d'accomplissement (s)	Log-normale	7.301	1	.007	O
	Erreur angulaire globale (°)	Log-normale	1.367	1	.242	N
Gauche / Droite	Temps d'accomplissement (s)	Log-normale	3.293	1	.070	N
	Erreur angulaire globale (°)	Log-normale	0.564	1	.453	N
Haut / Bas	Temps d'accomplissement (s)	Log-normale	0.017	1	.896	N
	Erreur angulaire globale (°)	Log-normale	5.296	1	.021	O

L'estimation des effets est basée sur la mesure statistique Wald chi-square. Le test est de type III pour lequel l'ordre d'agrégation des facteurs n'est pas important. Le modèle de distribution de fréquence utilisé est indiqué pour chaque variable dans la colonne Modèle. Le niveau de confiance est de 95%.

Ce résultat montre que la technique basée sur les indices SASA permet d'atteindre une localisation de sources sonores avec une précision supérieure à celle rapportée dans la littérature. Cependant, cette technique produit un temps d'accomplissement de la tâche supérieur à celui des approches développées par les autres auteurs.

D'après ces résultats, nous considérons que la technique d'indices SASA peut remplacer les modèles d'écoute spatiale humaine pour la tâche de localisation de source sonore. Par rapport au modèle de l'écoute humaine, cette approche présente l'avantage d'être généralisée et elle utilise une technologie à faible coût.

4.2.9. Conclusions

L'approche que nous avons proposée, permet d'obtenir une localisation de source sonore précise en azimut et en élévation. Les approches développées dans les travaux publiés permettent une précision équivalente en azimut uniquement.

Notre approche permet également de réduire l'erreur en azimut. L'erreur angulaire globale obtenue est inférieure à celle obtenue par les autres approches. Nous constatons également que notre approche a permis d'éliminer les erreurs inverses dans la tâche de localisation de source sonore.

Le résultat de l'étude expérimentale, que nous avons menée, montre qu'il n'existe pas de différence entre la performance des sujets de sexe masculin et les sujets de sexe féminin dans la tâche de localisation de source sonore.

Les mécanismes de stimulation auditive existants présentent des difficultés dans l'interaction 3D. Ces mécanismes actuels de stimulation auditive présentent des limites pour transmettre efficacement une information spatiale. L'approche que nous proposons permet de fournir à l'utilisateur une information spatiale utile pour l'interaction en espace tridimensionnel.

L'étude expérimentale que nous avons menée montre qu'il est nécessaire d'optimiser l'approche que nous proposons afin de réduire le temps d'accomplissement de la tâche de localisation de source sonore. Dans le chapitre suivant, nous proposons une technique basée sur la combinaison de cette approche et le modèle de l'écoute spatiale HRTF pour améliorer le temps d'accomplissement de la tâche.

5. Étude comparative et intégration des techniques SASA et HRTF

Dans ce chapitre, nous présentons une étude comparative entre la technique basée sur la fonction HRTF et notre approche basée sur les indices SASA. Cette étude comparative est effectuée pour la tâche de localisation de source sonore en environnement tridimensionnel. Afin d'exploiter les avantages de chaque technique nous proposons une approche qui intègre la fonction HRTF et les indices SASA.

5.1. Motivation de la recherche d'un modèle performant

D'après les expériences décrites dans les chapitres précédents, il ressort que les utilisateurs effectuent la recherche de la source sonore en deux phases : (1) une phase de localisation grossière au cours de laquelle l'utilisateur effectue un mouvement rapide de la tête et du corps pour chercher la zone de provenance de la source sonore ; et (2) une phase de localisation fine pendant laquelle l'utilisateur effectue un mouvement fin de la tête pour localiser la source sonore de façon précise.

L'observation des expériences précédentes nous permet de constater que les indices auditifs SASA sont efficaces dans la phase de localisation précise mais non appropriés à la localisation grossière. Dans ce chapitre, nous proposons de vérifier si la fonction HRTF est efficace dans la phase de la localisation grossière de source sonore. L'idée est de chercher à développer un modèle combiné qui améliore le temps d'accomplissement de la tâche dans les deux phases.

Ce chapitre décrit la procédure d'intégration de deux modèles : la fonction HRTF et l'approche basée sur les indices sonores SASA.

5.2. Modèle combiné HRTF - SASA

L'intégration consiste à inclure les indices SASA dans le signal sonore avant de le spatialiser avec le modèle HRTF. Dans cette intégration, la méthode de directionnalisation de la technique SASA (Puissance Constante) sera substituée au modèle de la HRTF. La Figure 5.1 présente les trois différentes procédures de spatialisation utilisées dans cette étude. La Figure 5.1a présente la procédure de traitement sonore de l'approche

SASA. La Figure 5.1b montre la méthode de spatialisation de la HRTF. La Figure 5.1c présente le diagramme de la procédure de stimulation spatiale qui utilise le modèle combiné de la HRTF et les indices SASA.

Étant donné un stimulus sonore enrichi avec les indices SASA $x^{SASA}[n]$, les sorties spatiales du modèle combiné pour les canaux droit $x_R^C[n]$ et gauche $x_L^C[n]$ sont définies par :

$$x_R^C[n] = h_R(\varphi, \delta)[n] * x^{SASA}[n] \quad (5.1)$$

$$x_L^C[n] = h_L(\varphi, \delta)[n] * x^{SASA}[n] \quad (5.2)$$

où $h_R(\varphi, \delta)[n]$ est la réponse impulsionnelle de l'oreille droite pour la source sonore située en (φ, δ) ,

$h_L(\varphi, \delta)[n]$ est la réponse impulsionnelle de l'oreille gauche, et

* représente la convolution de deux signaux discrets.

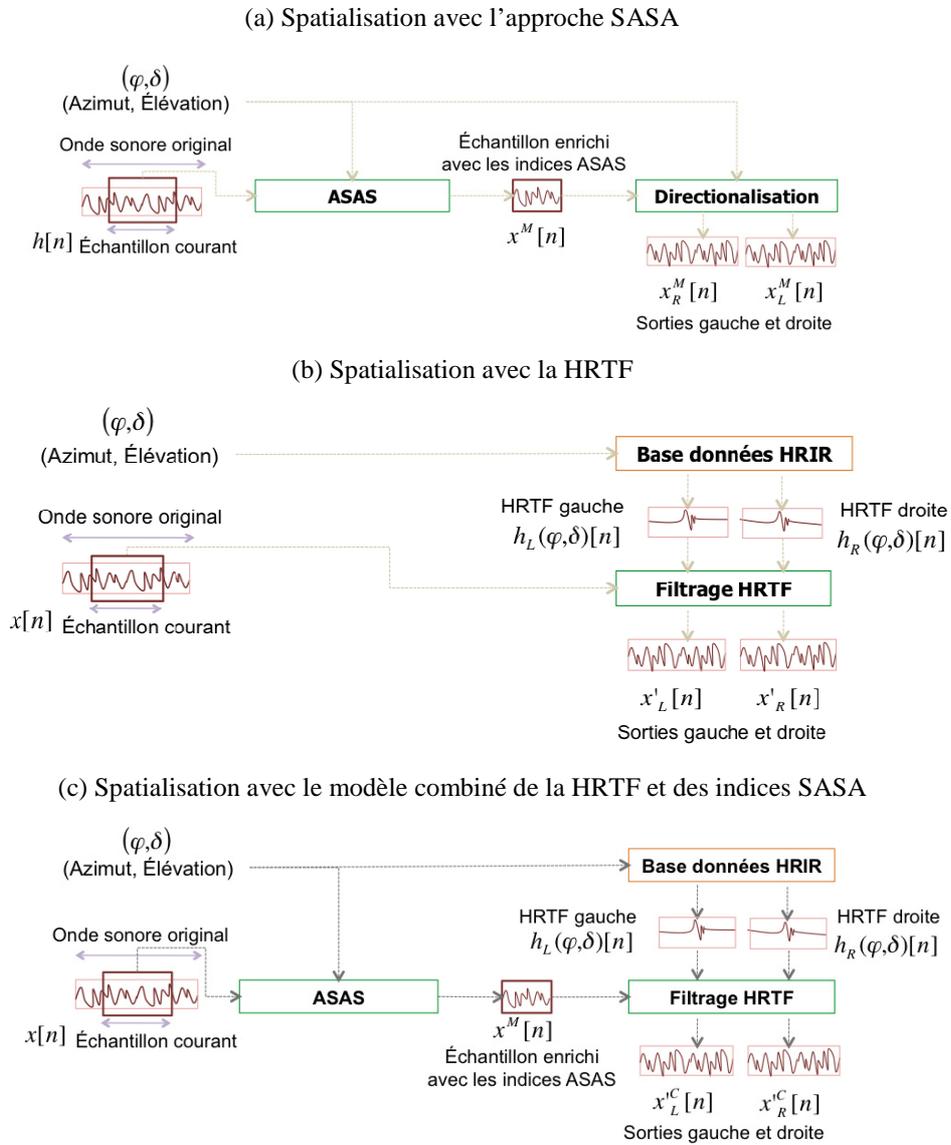


Figure 5.1. Les différentes méthodes implantées pour la stimulation auditive spatiale : (a) SASA, (b) HRTF et (c) HRTF+SASA.

L'inconvénient principal du modèle de la HRTF est sa faible précision dans la localisation de source sonore. D'après les résultats des expériences décrites précédemment, nous avons constaté que les indices SASA permettent une bonne précision de localisation mais un temps d'accomplissement de la tâche plus important que les approches existantes. Si le modèle combiné permet de pallier aux inconvénients de chaque modèle utilisé séparément, nous considérons que l'intégration de deux approches permettra d'améliorer la précision de localisation de source sonore et de réduire le temps d'accomplissement de la tâche.

Il est important de noter que le modèle combiné est moins compatible avec le phénomène de l'écoute binaurale que la fonction HRTF. Ce manque de compatibilité provient des informations supplémentaires dans les indices SASA, non existantes dans l'écoute réelle. Par conséquent, nous considérons que l'approche du modèle combiné n'est pas approprié aux applications dont le réalisme sonore est un critère important.

Pour réaliser les expériences décrites dans ce chapitre, nous avons utilisé les indices sonores SASA décrits dans le chapitre précédent. L'indice du cône de diffusion (indice de proximité angulaire renforcée) n'est pas pris en compte dans ces expériences car il induit un retard du temps de localisation de source sonore observable dans les expériences décrites dans la section 4.2.6.

5.3. Étude expérimentale

L'objectif de cette étude expérimentale est d'évaluer trois techniques de spatialisation auditive pour la localisation sonore : le modèle classique de la HRTF, l'approche des indices auditifs SASA et le modèle combiné de la HRTF et les indices SASA. L'idée est de comparer la performance de la fonction HRTF avec celles des indices auditifs SASA et par la suite comparer la performance du modèle combiné à celles des autres techniques lorsqu'elles sont utilisées séparément.

Pour l'étude de la performance, nous avons effectué une évaluation objective et une évaluation subjective. Concernant l'évaluation objective, nous avons mesuré le temps d'accomplissement de la tâche et l'erreur angulaire (azimut, élévation et angle globale). L'évaluation subjective est basée sur le retour des utilisateurs tels que le réalisme de l'interaction et la facilité d'utilisation. La tâche étudiée consiste en la localisation de source sonore spatiale.

Chaque sujet exécute la tâche avec trois modèles de spatialisation sonore : SASA, HRTF et modèle combiné (SASA+HRTF). La tâche est exécutée dans les trois conditions expérimentales dans un ordre aléatoire.

La recherche de la source sonore par pointage est effectuée avec le mouvement de la tête. L'utilisateur porte un casque dans lequel est fixé le système de suivi de mouvement. Le sujet se sert de la Wiimote pour indiquer au système la fin de la tâche, ceci implique que la source sonore est localisée. L'architecture logicielle et matérielle de la plateforme expérimentale est la même que celle décrite dans le chapitre précédent. Dans cette étude, la tâche est réalisée avec une occlusion visuelle totale.

5.3.1. Essais et stratégie de localisation de source sonore

Avant de réaliser la tâche dans chaque condition expérimentale, le sujet doit effectuer deux essais pour se familiariser avec le système. Après ces essais préliminaires, le sujet porte un masque jusqu'à la fin des essais. Les essais pour cette expérimentation sont ceux décrits dans le chapitre précédent.

Les conditions expérimentales sont identifiées comme suit (A) : HRTF ; (B) : SASA ; et (C) : modèle combiné HRTF + SASA. L'expérimentation demande aux sujets de suivre la même stratégie que celle décrite dans le chapitre 4.

5.3.2. Évaluation expérimentale

Pour étudier l'effet de chaque modèle de stimulation auditive sur la performance de l'utilisateur et sur les indicateurs de perception, nous avons défini une série de mesures objectives et une série de mesures subjectives.

5.3.2.1. Évaluation quantitative

Pour l'évaluation objective, on mesure le temps d'accomplissement de la tâche (en secondes) et l'erreur angulaire en azimut, en évaluation et l'angle global (en degrés).

5.3.2.2. Retour d'informations de l'utilisateur

Pour recueillir le retour d'informations des utilisateurs, on demande à chaque sujet de répondre à un questionnaire à la fin de l'expérience. Les questions posées portent sur le réalisme de l'interaction et la facilité d'utilisation de la technique d'interaction.

Les questions ont été élaborées pour permettre au sujet de comparer les trois paires de conditions expérimentales : (A et B), (B et C) et (A et C). Les modèles de questions et les réponses possibles sont présentés ci-dessous. Les modèles de question 1 et 2 correspondent aux indicateurs de réalisme et de facilité d'utilisation.

Modèle de question 1 : Parmi les conditions X et Y, quelle est la condition la plus réaliste?

- 3 Y est beaucoup plus réaliste
- 2 Y est plus réaliste
- 1 Y est un peu plus réaliste
- 0 X est aussi réaliste que Y
- 1 X est un peu plus réaliste
- 2 X est plus réaliste
- 3 X est beaucoup plus réaliste

où (X,Y) représente les combinaisons possibles des conditions A, B et C : (A,B), (B,C) et (A,C).

Modèle de question 2 : Parmi les conditions X et Y, quelle est la condition la plus facile à utiliser?

- 3 Y est beaucoup plus facile à utiliser
- 2 Y est plus facile à utiliser
- 1 Y est un peu plus facile à utiliser
- 0 X est aussi facile à utiliser que Y

- 1 X est un peu plus facile à utiliser
- 2 X est plus facile à utiliser
- 3 X est beaucoup plus facile à utiliser

où (X,Y) représente les combinaisons possibles des conditions A, B et C.

Pour définir les réponses possibles, nous avons utilisé le modèle de la Catégorie de Notation en Dégradation (DCR : *Degradation Category Rating*) recommandée par l'ITU-T [1996]. Au total, chaque sujet répond à six questions (c.f. Annexe : 10.1).

5.3.3. Groupe de participants et hypothèses

Quinze sujets âgés entre 20 et 30 ans (douze hommes et trois femmes) ont participé à ces expérimentations. Ils sont tous droitiers et ils ont déclaré ne pas avoir de maladies ni visuelles, ni auditives, ni psychomotrices. Ils sont tous étudiants et ils ont une vision et une audition normales et ne souffrent d'aucune maladie psychomotrice. Ils sont étudiants et n'ont pas d'expérience en réalité virtuelle. Chaque sujet passe 40 minutes environ pour réaliser la tâche dans les trois conditions expérimentales avec une pause de trois minutes à la fin de chaque série d'essais.

Nous avons posé les hypothèses suivantes pour cette étude expérimentale :

1. Comparée au modèle HRTF, l'approche SASA permet d'améliorer la précision angulaire par rapport au modèle HRTF.
2. L'utilisation de la HRTF permet de réaliser un le temps d'accomplissement de la tâche inférieur à celle obtenue avec l'approche SASA.
3. Le modèle combiné permet de bénéficier des avantages de chaque technique utilisée séparément. Ceci conduit à une localisation de source sonore rapide et précise.

5.4. Résultats expérimentaux et analyse de données

5.4.1. Évaluation de la performance

5.4.1.1. Caractérisation des techniques évaluées

Le Tableau 5.1 présente le résultat statistique des variables de l'évaluation de la performance.

La Figure 5.2a présente la moyenne du temps d'accomplissement de la tâche pour chaque condition expérimentale. Cette figure montre que la HRTF permet d'obtenir un temps d'accomplissement de la tâche inférieur à celui obtenu avec l'approche SASA. Le modèle combiné permet d'obtenir un temps d'accomplissement de la tâche similaire à celui obtenu avec l'approche SASA.

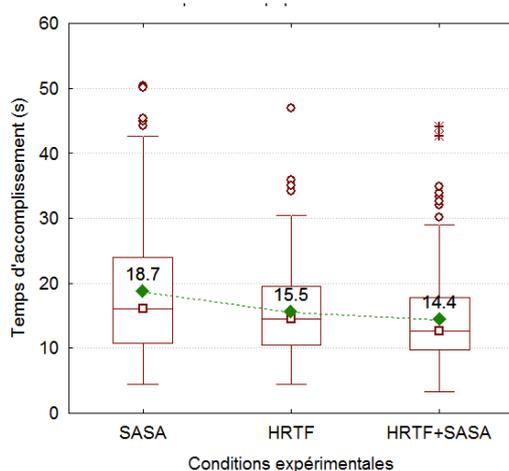
Tableau 5.1. Mesures statistiques des données obtenues

Groupe	Variable	N ^{bre}	Moyenne	Médiane	Quartile			Écart type
					Inf.	Sup.	Inter.	
Toutes les conditions	Temps d'accomplissement (s)	491	16.21	14.14	10.16	20.04	9.88	8.73
	Erreur angulaire globale (°)	491	11.33	4.98	2.52	13.61	11.09	14.49
	Erreur en azimut (°)	491	3.59	2.12	1.04	4.07	3.03	4.91
	Erreur en élévation (°)	491	10.04	3.20	1.35	11.29	9.94	14.65
A. SASA	Temps d'accomplissement (s)	165	18.72	16.12	10.81	23.92	13.11	10.61
	Erreur angulaire globale (°)	165	4.97	4.34	2.69	6.21	3.52	4.51
	Erreur en azimut (°)	165	4.06	2.49	1.11	5.04	3.93	5.68
	Erreur en élévation (°)	165	3.17	2.20	1.09	4.18	3.09	4.39
B. HRTF	Temps d'accomplissement (s)	161	15.46	14.52	10.44	19.59	9.15	7.01
	Erreur angulaire globale (°)	161	25.76	23.41	12.76	35.39	22.63	17.02
	Erreur en azimut (°)	161	4.96	2.97	1.40	5.20	3.80	5.74
	Erreur en élévation (°)	161	24.51	21.20	10.06	34.86	24.80	17.40
C. HRTF+SASA	Temps d'accomplissement (s)	165	14.42	12.66	9.75	17.79	8.04	7.54
	Erreur angulaire globale (°)	165	3.62	2.55	1.86	3.99	2.13	4.46
	Erreur en azimut (°)	165	1.78	1.43	0.85	2.32	1.47	1.50
	Erreur en élévation (°)	165	2.77	1.67	0.86	3.20	2.34	4.47

La précision du système de suivi de la tête est de $\pm 0.5^\circ$. Le calcul statistique de l'erreur angulaire (azimut, élévation et angle global) est effectuée sur la base de la valeur absolue de l'angle mesuré.

La Figure 5.2b présente la moyenne de l'erreur angulaire pour chaque condition expérimentale. D'après cette figure, la précision obtenue avec la fonction HRTF est inférieure à celle obtenue avec l'approche SASA. Le modèle combiné permet une précision angulaire supérieure à celle obtenue avec l'approche SASA. La signification statistique des différences observées sera déterminée plus tard dans le texte.

(a) Temps d'accomplissement Vs. conditions exp.



(b) Erreur angulaire globale Vs. conditions exp.

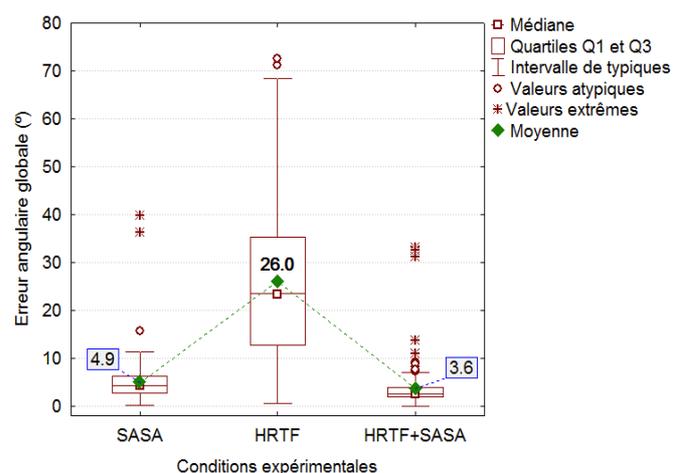


Figure 5.2. (a) Temps d'accomplissement et (b) Erreur angulaire globale pour les conditions expérimentales évaluées.

Les Figure 5.3a et Figure 5.3b montrent respectivement les moyennes des erreurs en azimut et en élévation pour les trois conditions expérimentales. L'utilisation de l'approche SASA et le modèle combiné permettent de réduire l'erreur moyenne en azimut de 1.1° et de 3.4° respectivement. Il est également à noter que, comparée au modèle HRTF, l'approche SASA et le modèle combiné permettent de réduire l'erreur moyenne en élévation de 21.5° et de 21.9° respectivement. Il est important de noter la présence d'une différence relative entre l'approche SASA et le modèle combiné en termes d'erreurs en azimut et en élévation.

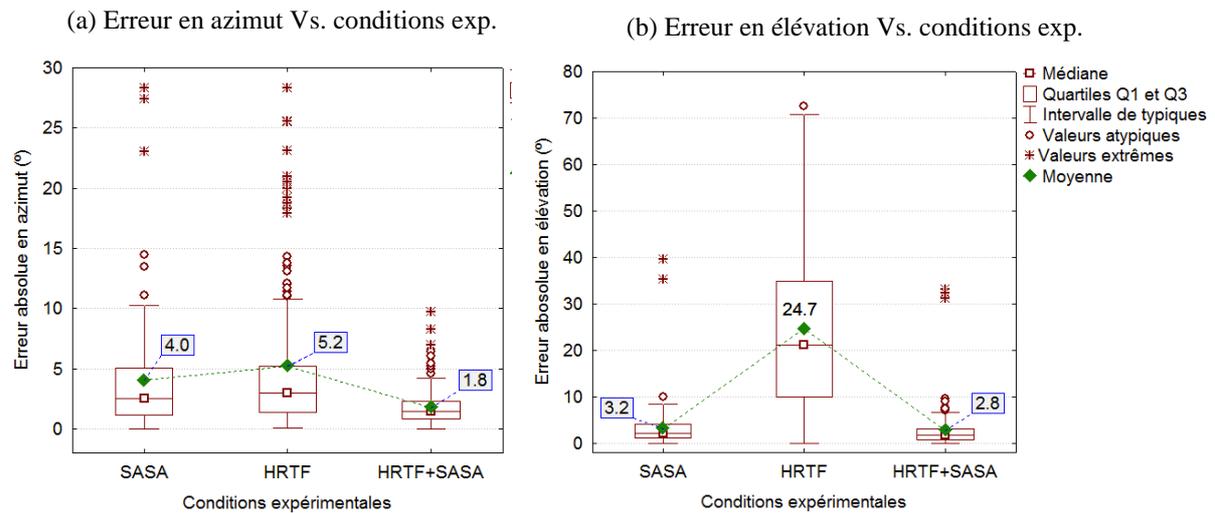


Figure 5.3. (a) Erreur en azimut et (b) Erreur en élévation pour les conditions expérimentales évaluées.

La Figure 5.4a ,montre le temps d'accomplissement de la tâche pour chaque sujet. On remarque que pour la plupart des sujets (11 sur 15), la variabilité intra-sujet est plus importante pour l'approche de SASA par rapport à celle enregistrée pour les autres conditions.

La Figure 5.4b montre l'erreur angulaire globale pour chaque sujet. D'après cette figure, les médianes obtenues pour la HRTF sont supérieures à celles par les autres conditions. On remarque également que tous les utilisateurs ont une variabilité intra-sujet similaire à celle obtenue avec l'approche SASA et le modèle combiné.

5.4.1.2. Distribution de fréquence des données

Le Tableau 5.2 montre les résultats des tests de qualité d'ajustement pour le temps d'accomplissement de la tâche et pour l'erreur angulaire globale. Ces tests ont été effectués pour déterminer la distribution de fréquences de ces variables. Toutes les mesures valides ont été prises en compte dans l'analyse.

Les résultats montrent que les données de ces variables ne sont pas normales. Ces données suivent plutôt une distribution de type log-normale (Figure 5.5a et Figure 5.5b). Il n'est pas possible de déterminer le type de distribution pour l'erreur en azimut et l'erreur en élévation. Ceci est dû à l'effet de chaque condition expérimentale sur la distribution de ces variables. Pour cette raison, il est nécessaire d'utiliser des méthodes d'analyse statistique non-paramétrées pour déterminer les effets statistiques de ces variables.

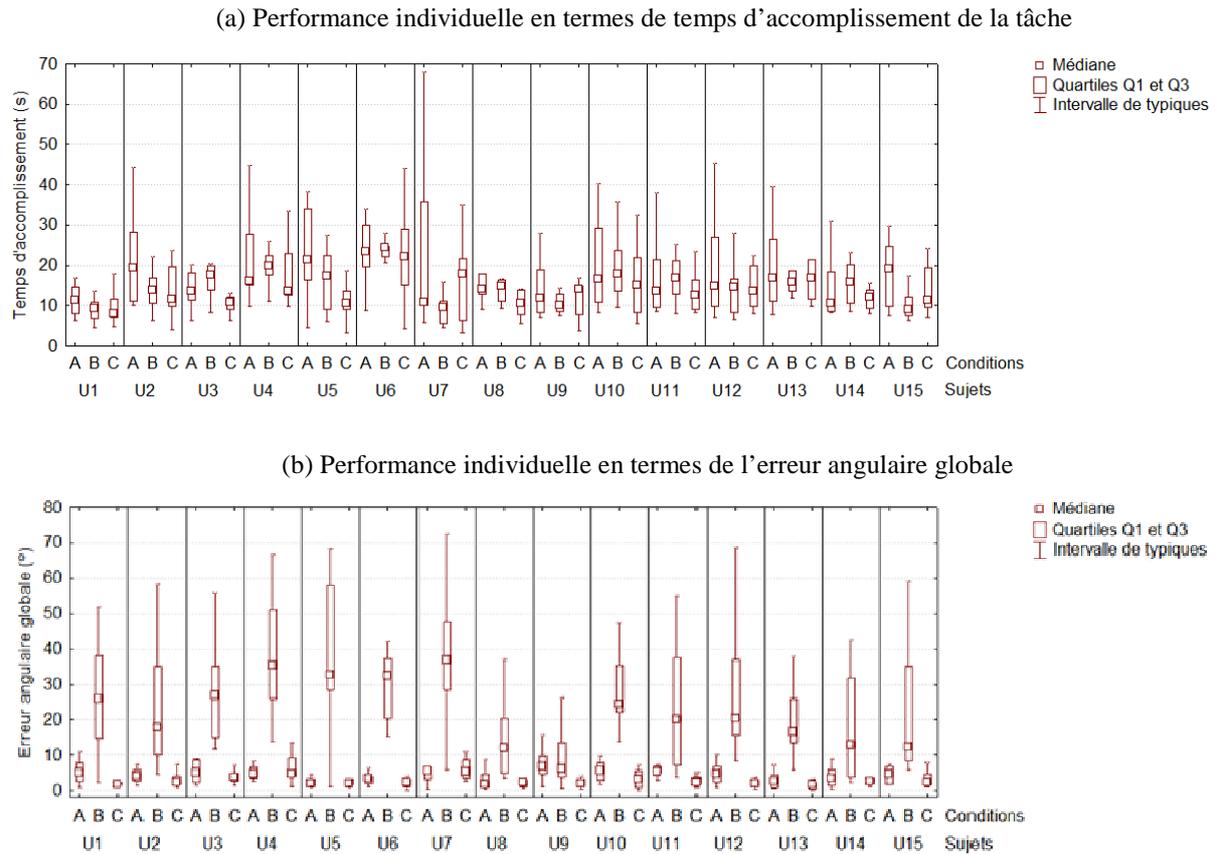


Figure 5.4. Performance individuelle pour le (a) temps d'accomplissement de la tâche et pour (b) l'erreur angulaire globale. A, B et C représentent les conditions expérimentales de SASA, HRTF et HRTF+SASA respectivement.

Tableau 5.2. Tests de qualité d'ajustement

Fonction	Variable	Moyen ne	Varian ce	Moyen ne ajustée	Varian ce ajustée	Distan ce K-S	p	Ajustée (o/n) ?
Normale	Temps d'accomplissement (s)	16.20	76.23	16.29	76.23	0.116	<0.01	N
	Erreur angulaire globale (°)	11.33	210.06	11.33	210.06	0.265	<0.01	N
Log-normale	Temps d'accomplissement (s)	16.20	76.23	2.65	0.25	0.032	>0.05	O
	Erreur angulaire globale (°)	11.33	210.06	1.72	1.52	0.061	0.05 < p < 0.1	O

La méthode de comparaison utilisée est le test Kolmogorov-Smirnov (K-S) avec un niveau de confiance de 95%. Pour l'erreur angulaire, nous avons tenu compte de la valeur absolue des données traitées.

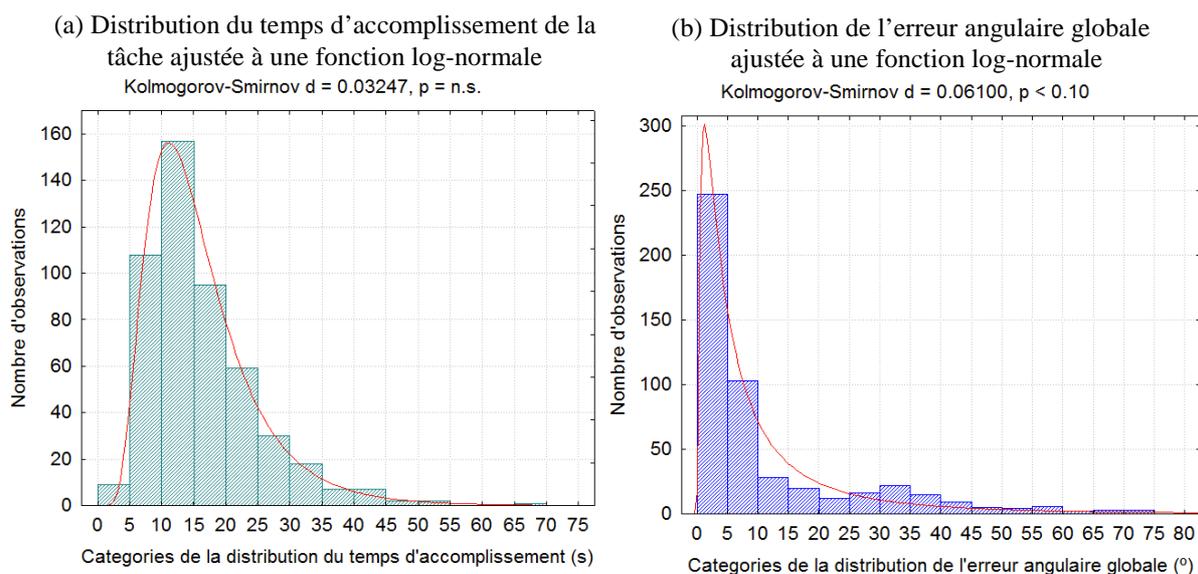


Figure 5.5. Distribution de fréquence pour : (a) le temps d'accomplissement de la tâche et (b) l'erreur angulaire globale.

5.4.1.3. Analyse statistique

Pour l'analyse statistique des données, nous avons utilisé la méthode des Equations d'Estimation Généralisées.

Le Tableau 5.3 montre le résultat de l'analyse statistique de la comparaison des différentes conditions expérimentales basée sur la méthode d'Estimation des Moyennes Marginales.

D'après cette analyse, la fonction HRTF, comparée à l'approche basée sur les indices sonores SASA, a un effet statistiquement significatif sur le temps d'accomplissement de la tâche ($p=0.002$). Le modèle combiné (HRTF+SASA) a également un effet statistiquement significatif sur le temps d'accomplissement de la tâche ($p=0.000$) comparée à l'approche des indices SASA. Cette analyse ne montre pas une différence statistiquement significative entre la HRTF et le modèle combiné (HRTF+SASA) par rapport le temps d'accomplissement de la tâche.

Le résultat de l'analyse montre que, comparée à la HRTF, l'approche SASA réduit significativement l'erreur angulaire ($p=0.000$). Par rapport à la HRTF, le modèle combiné a un effet significatif sur l'erreur angulaire ($p=0.000$). La différence entre l'approche SASA et le modèle combiné est également statistiquement significative ($p=0.000$) pour cette erreur.

La comparaison entre l'approche SASA et la HRTF montre que la différence de l'erreur angulaire n'est pas significative en azimut ($p=0.409$) et elle significative en élévation ($p=0.000$). Par rapport à l'approche des indices SASA, le modèle combiné a un effet significatif sur l'erreur angulaire en azimut ($p=0.028$) mais il n'a

pas en effet significatif sur l'erreur angulaire en élévation ($p=0.387$). Le modèle combiné, comparé à la HRTF, a un effet significatif sur l'erreur en azimut ($p=0.000$) et sur l'erreur en élévation ($p=0.000$).

Tableau 5.3. Estimation Généralisée de Moyennes Marginales par l'Analyse de Mesures Répétées avec le modèle d'Equations Généralisées d'Estimation

Facteurs		Variable	Différence moyenne	Erreur standard	DDL	p	Wald IC 95%		Significatif (o/n) ?
(1)	(2)						Inf.	Sup.	
SASA	HRTF	Temps d'accomplissement (s)	2.85	.92	1	.002	1.05	4.65	O
		Erreur angulaire globale (°)	-18.98	2.14	1	.000	-23.17	14.79	O
		Erreur en azimut (°)	-1.17	1.42	1	.409	-3.97	1.62	N
		Erreur en élévation (°)	-21.51	2.01	1	.000	-25.46	17.57	O
SASA	HRTF +SASA	Temps d'accomplissement (s)	4.14	.64		.000	2.89	5.39	O
		Erreur angulaire globale (°)	1.33	.46	1	.004	0.42	2.23	O
		Erreur en azimut (°)	2.28	1.04	1	.028	0.24	4.32	O
		Erreur en élévation (°)	0.39	.45	1	.387	-0.50	1.28	N
HRTF	HRTF +SASA	Temps d'accomplissement (s)	1.29	.81	1	.113	-0.31	2.89	N
		Erreur angulaire globale (°)	20.31	1.96	1	.000	16.46	24.15	O
		Erreur en azimut (°)	3.45	.83	1	.000	1.84	5.07	O
		Erreur en élévation (°)	21.90	2.00	1	.000	17.98	25.83	O

Comparaison des moyennes marginales pour les différentes conditions expérimentales basée sur les échelles originales des variables analysées. Dans cette analyse, nous avons pris en compte la valeur absolue de l'erreur angulaire (azimut, élévation et angle global). Le niveau de confiance est de 95%.

L'analyse montre que l'approche SASA est plus précise que la HRTF en élévation et elle permet de réduire l'erreur angulaire globale. La HRTF permet de réaliser la tâche plus rapidement que l'approche SASA et elle a une précision en azimut similaire à celle obtenue avec l'approche SASA. Le modèle combiné permet une rapidité et une précision angulaire en élévation similaires à celles obtenues avec l'approche SASA. On remarque également que le modèle combiné permet une précision angulaire en azimut plus importante que celle obtenue avec la fonction HRTF et l'approche SASA utilisées séparément. On constate que le modèle combiné HRTF+SASA prend les avantages de la HRTF et de l'approche SASA lorsqu'ils sont utilisés séparément. Ce résultat confirme notre hypothèse pour l'intégration de deux modèles.

L'intégration de la fonction HRTF et le modèle SASA permet de transmettre à l'auditeur des informations spatiales précises, ce qui réduit le temps d'accomplissement de la tâche interactive en environnement virtuel.

5.4.2. Retour d'informations de l'utilisateur

L'évaluation subjective effectuée est basée sur le retour d'informations de l'utilisateur. L'expérimentateur demande aux sujets ayant participé aux essais de répondre à un questionnaire. Ce dernier est constitué d'une série de questions focalisées sur le réalisme de l'interaction et sur la facilité d'utilisation de la technique de stimulation sonore dans la réalisation de la tâche.

La Figure 5.6 montre les préférences des utilisateurs parmi les paires possibles de conditions expérimentales. Pour faciliter la lecture du diagramme, les réponses de type : « ...est aussi réaliste » et « ...est aussi facile à utiliser » n'ont pas été présentées. Pour la même raison, les réponses ont été combinées en fonction du signe de leurs valeurs (positives ou négatives) dans la notation DCR.

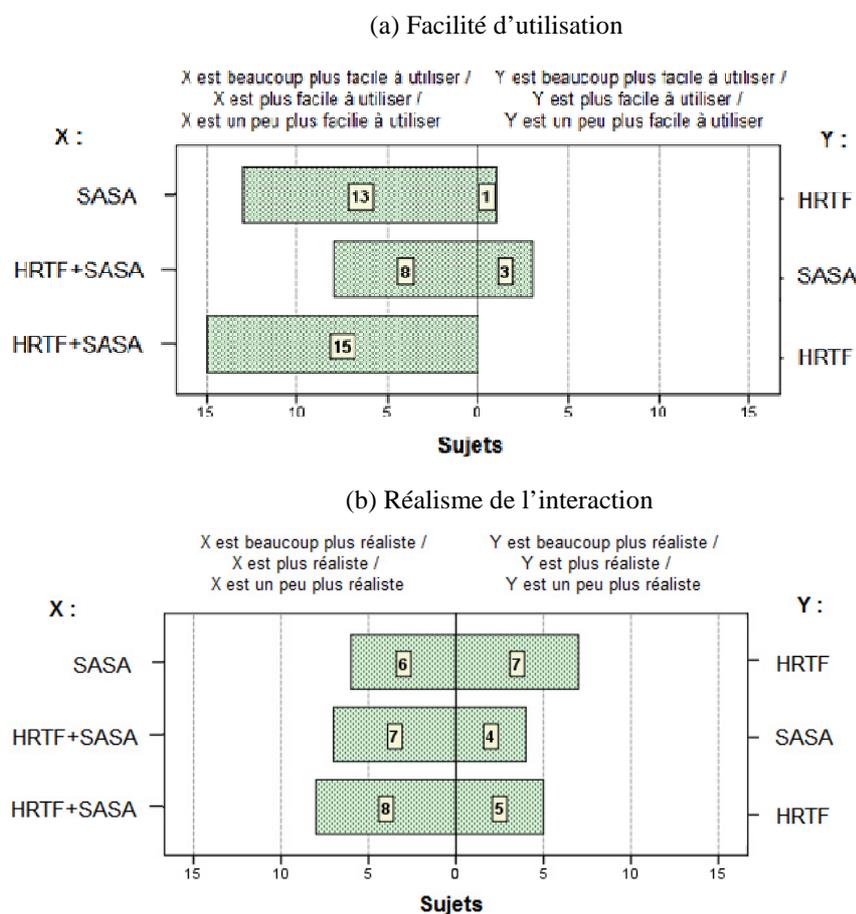


Figure 5.6. L'appréciation des utilisateurs par rapport à (a) la facilité d'utilisation et (b) au réalisme de l'interaction.

D'après la Figure 5.6, la plupart des sujets considèrent que l'approche SASA est plus facile à utiliser que la HRTF et que le modèle combiné est plus facile que la HRTF et l'approche SASA lorsqu'elles sont utilisées séparément.

La fonction HRTF est un modèle destiné à la restitution du phénomène de l'écoute binaurale. En revanche, l'approche SASA a pour but de recréer des indices auditifs artificiels. L'évaluation subjective montre que le réalisme de l'approche SASA est proche de celui de la HRTF (6 sujets ont préféré le réalisme de l'approche SASA et 7 sujets ont préféré celui de la HRTF). Sept sujets (contre trois) considèrent que le modèle combiné est plus réaliste que l'approche SASA. D'un autre côté, huit sujets (contre quatre) considèrent que le modèle combiné est plus réaliste que la fonction HRTF.

Le résultat de l'évaluation subjective fournit des informations utiles pour la compréhension des mécanismes de stimulation auditive proposés. Nous constatons que les sujets apprécient la facilité d'utilisation et le réalisme de l'approche SASA. D'après cette évaluation, il ressort que le modèle combiné rend plus facile la localisation de source sonore.

Dans l'appréciation des utilisateurs, le réalisme de l'approche SASA est pratiquement équivalent à celui de la HRTF et le modèle combiné permet d'augmenter le réalisme perçu. Ce résultat nous permet de considérer que la combinaison de la HRTF et de l'approche SASA permet d'améliorer les indicateurs de perception évalués.

5.4.3. Comparaison de notre approche avec les techniques de spatialisation sonore existantes

Le Tableau 5.4 présente les résultats de l'évaluation de la performance pour l'approche SASA et pour d'autres travaux publiés. Le tableau présente les techniques les plus connues par la communauté scientifique.

Pour chaque technique, nous avons recueilli les mesures de temps d'accomplissement de la tâche et de l'erreur angulaire absolue lorsqu'elles sont disponibles dans les publications originales. Il est à noter que dans certaines publications, les variables objectives définies dans notre étude ne sont pas identiquement prises en compte.

Les techniques de spatialisation prises en compte dans cette comparaison ont été déjà décrites dans les chapitres précédents. Les différences majeures entre ces techniques sont les suivantes : l'utilisation des sources sonores réelles ou virtuelles, l'interaction avec retour visuel ou avec occlusion visuelle et l'utilisation de modèles d'écoute humaine généralisés ou individualisés dans les cas où les sources virtuelles sont utilisées.

Parmi les techniques de spatialisation sonore utilisant des indices supplémentaires, le modèle combiné est celui qui permet une meilleure précision de localisation sonore.

Par rapport aux autres travaux publiés, le modèle combiné permet de réduire l'erreur angulaire en azimut et en élévation. Par contre, la précision est plus importante pour l'élévation que pour l'azimut. Ceci est également valide lorsque le modèle combiné est comparé avec la HRTF individualisée. Néanmoins, l'approche du modèle combiné est plus précise que la HRTF individualisée pour la localisation de source sonore en environnement virtuel. Par conséquent, nous pouvons considérer que le modèle combiné permet d'éviter

l'utilisation des HRTFs individualisées, considérées jusqu'à aujourd'hui comme la méthode de spatialisation la plus précise.

Le temps d'accomplissement de la tâche obtenu en utilisant le modèle combiné est très proche des temps obtenus dans les travaux rencontrés dans la littérature. Il est important de mentionner que les travaux qui ont abouti à un temps d'accomplissement de la tâche plus court portent sur la localisation de source sonore dans le plan horizontal uniquement.

Tableau 5.4. Comparaison du modèle combiné avec d'autres approches publiées

Référence	R/ V	G/ I	Occlusion visuelle	Temps d'accomplissement (s)	Erreur en azimut (°)	Erreur en élévation (°)	Erreur angulaire globale (°)
E. M. Wenzel et al. [1993]	V	G	-	-	-	-	20-30
Jiann-Rong Wu et al. [1997]	V	G	-	3-6	9-18	-	-
Brown et R. O Duda [1997]	V	C&G	-	-	-	-	12-23
Simon Carlile et al. [1997]	R	-	-	-	3	4-12	-
Begault et al. [2001]	V	C&G	N	-	15-25	15-30	-
Langendijk [2002]	V	C	-	-	-	15-30	-
M. Grohn [2002]	V+	-	-	6-8	12-25	-	-
Marentakis et Brewster [2005]	V+	G	-	3-8	4-10	-	-
Doerr et al. [2007]	V+	G	N	9-13	3-4	11-17	-
Modèle combiné (HRTF+SAA)	V+	G	Y	14.41	1.77	2.77	3.61

Légende :
 R : sources sonores réelles
 V : sources sonores virtuelles
 V+ : sources sonores virtuelles + indications supplémentaires
 G : modèle généralisé d'écoute humaine
 I : modèle individualisé d'écoute humaine
 - : information non disponible ou non spécifiée

5.5. Conclusions

Dans ce chapitre, nous avons montré les avantages de l'intégration de la fonction HRTF et l'approche de stimulation auditive SASA que nous avons développée. L'intégration des deux modèles découle de la comparaison de la HRTF et le modèle SASA sur la base du temps d'accomplissement de la tâche et la précision angulaire.

Cette étude comparative montre que la HRTF est un peu plus réaliste que l'approche SASA et qu'elle permet un temps d'accomplissement de la tâche plus court mais avec une précision angulaire inférieure à celle obtenue avec l'approche SASA. L'approche SASA est plus facile à utiliser et plus précise que la HRTF mais le temps nécessaire à la localisation de la source sonore est plus élevé.

La combinaison de la HRTF et de l'approche SASA permet de réduire le temps d'accomplissement de la tâche et elle améliore la précision angulaire par rapport à la HRTF et à l'approche SASA utilisées séparément. Cependant le modèle combiné est moins cohérent avec l'acoustique réelle lorsqu'il est comparé à la HRTF. Le temps nécessaire pour le traitement audio est plus important avec le modèle combiné qu'avec l'approche SASA ou la HRTF utilisées séparément.

Les approches décrites dans la littérature et qui permettent une localisation de source sonore rapide et précise sont restreintes aux sources sonores situées dans le plan horizontal uniquement.

Comparée aux approches décrites dans la littérature, le modèle combiné (HRTF+SASA) permet de localiser une source sonore en environnement tridimensionnel avec une meilleure précision et un temps d'accomplissement de la tâche plus court. L'approche SASA et le modèle combiné permettent une localisation précise en azimut et en élévation. Le résultat de l'étude comparative montre également que le modèle combiné permet d'éviter l'utilisation de modèles individualisés d'écoute binaurale.

Malgré la précision de localisation angulaire de source sonore obtenue par l'approche SASA, le temps nécessaire pour accomplir la tâche est relativement important. Ce dernier est significativement réduit en adoptant le modèle combiné (HRTF+SASA). Compte tenu de ce résultat, nous envisageons d'utiliser le modèle combiné pour restituer à l'utilisateur des informations spatiales dans la simulation de chirurgie arthroscopique qui sera décrite dans le chapitre 7.

6. Approche CITD pour améliorer la simulation de la profondeur

Dans ce chapitre, nous présentons une approche pour améliorer la simulation de la profondeur de sources sonores spatiales virtuelles. L'approche est basée sur la compensation des Différences Interaurales de Temps (CITD) d'une HRTF de rayon fixe⁴ en fonction des variations de la profondeur. Cette approche s'applique à des sources sonores localisées en dehors du champ intermédiaire ; c'est-à-dire situées à des distances entre 1 et 5 mètres. Nous proposons de mettre en œuvre l'approche CITD afin d'améliorer la localisation en profondeur de la source sonore ainsi que sa perception.

Dans les chapitres précédents, nous avons décrit et validé notre approche pour la localisation précise d'une source sonore en fonction de sa direction de provenance. Néanmoins, les techniques existantes pour localiser une source sonore en profondeur (détermination de la distance entre l'auditeur et la source sonore) ne sont pas encore très performantes.

6.1. Profondeur de sources sonores spatiales

Dans des conditions réelles, l'être humain tend à surestimer la profondeur des sources sonores situées à moins d'un mètre (avec une erreur typique de 0.15 à 0.3 mètres) (Cooke et Yan-Chen [2007]). L'auditeur tend également à sous-estimer les sources sonores situées à une distance supérieure à 1 mètre et l'erreur de localisation augmente selon la distance (Loomis et al. [1999]).

Zahorik [2002a] a proposé un modèle qui exprime la profondeur perçue en fonction de la distance physique dans des conditions d'écoute réelles. Le modèle est défini par $p' = k \cdot p^a$ où $k = 1.3$, $a = 0.5$ et p et p' représentent la profondeur réelle et la profondeur perçue par l'auditeur respectivement. Les erreurs absolues typiques définies par ce modèle pour les distances de 1, 2, 3, 4, 5, 6 et 7 mètres sont respectivement de 0.3, 0.1, 0.6, 1.3, 2.6 et 3.3 mètres avec une moyenne de 1.4 mètres.

Il existe quatre groupes principaux d'indices sensoriels pour la détermination de la localisation en profondeur de sources sonores (Zahorik [2002b]). Le premier groupe est l'intensité et il représente la perte de puissance en fonction de la distance parcourue par le son. Le deuxième groupe est formé par les indices binauraux : différence interaurale de temps et différence interaurale d'intensité. Le troisième groupe est le

⁴ Une basé de données de fonctions HRTF à rayon fixe se compose de fonctions de transfert des sources sonores qui ont été toutes mesurées à la même distance de l'auditeur.

rapport d'intensité entre la source sonore directe et les réverbérations correspondantes. Cet indice est important pour les sources sonores localisées dans le champ proche (distances inférieures à 1 mètre) et il est significativement moins important pour les sources sonores plus éloignées. Le dernier groupe d'indices est celui des indices spectraux. Les hautes fréquences sont atténuées par l'absorption de l'air à grandes distances. En moyenne, ces fréquences sont atténuées à un taux de 6 dB tous les 100 mètres. Cet effet peut donc être considéré négligeable pour les sources sonores proches (localisées à moins de 20 mètres).

Nous allons centrer notre étude sur des sources sonores utilisables pour le guidage auditif en environnement virtuel. Par conséquent, nous allons nous limiter à l'étude des sources sonores situées à une distance de 1 à 5 mètres de l'auditeur où les indices spectraux sont négligeables. Nous appelons cet intervalle « champ intermédiaire ».

Afin de simplifier l'étude, l'environnement acoustique virtuel est considéré comme anti échoïque. C'est-à-dire que le contact entre les ondes sonores et les surfaces ne produit aucun effet. Il est nécessaire de mentionner que dans l'approche SASA, l'effet d'écho est utilisé pour transmettre des informations supplémentaires. Autrement dit, les réverbérations utilisées sont artificielles et non réalistes. Étant donné que nous envisageons d'intégrer l'approche SASA, la fonction HRTF et l'approche proposée dans ce chapitre, nous avons décidé de ne pas inclure à ce stade des réverbérations réalistes puisqu'elles n'étaient pas prises en compte dans les modèles étudiés dans les chapitres précédents. L'absence des réverbérations réalistes élimine à son tour le troisième groupe d'indices pour la détermination de la profondeur : le rapport entre le son direct et les réverbérations correspondantes. De cette manière, l'approche présentée dans ce chapitre est basée uniquement sur deux groupes d'indices : les indices interaurales et l'indice d'intensité.

L'intensité du son représente l'indice le plus important dans la détermination de la profondeur par un auditeur. La perte de puissance associée est due à l'effet d'absorption du son par l'air. Dans des conditions idéales, de température ambiante et en champ libre, cette perte Γ est déterminée en dB (décibels) comme suit :

$$\Gamma = -10 \log_{10}(4\pi r^2) \text{ [dB]} \quad (6.1)$$

où r [mètres] représente la distance entre l'auditeur et la source sonore.

La plupart des systèmes de simulation de profondeur sont basés sur le modèle de perte de puissance. L'intensité est un indice ambigu puisque les variations peuvent être également dues aux changements de la puissance originale de la source sonore.

Les changements de profondeur d'une source sonore peuvent contribuer à son identification. Ceci implique que la profondeur des sources sonores en mouvement peut être plus facilement déterminée que celle des sources sonores fixes.

L'indice de différences binaurales se réfère aux différences interaurales (entre les deux oreilles). La Figure 6.1 montre un exemple des différences interaurales de temps (ITD) et d'intensité (ILD). Ces différences se manifestent lorsque les distances entre la source sonore et les oreilles de l'auditeur ne sont pas égales. L'ITD est le retard d'arrivée d'un son à une oreille par rapport à l'autre. L'ILD est la différence d'intensité entre les sons arrivant à chaque oreille et elles sont dues principalement à l'effet d'occlusion (*shadowing effect*) de la tête

(Goldstein [1999]). Les différences interaurales dépendent de la direction de provenance de la source sonore mais aussi de sa profondeur.

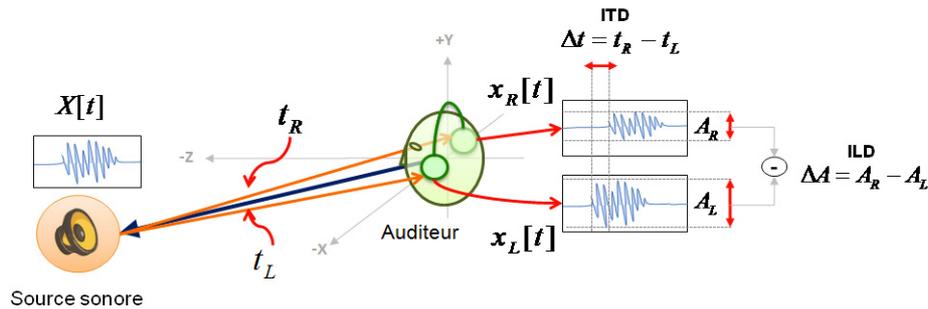


Figure 6.1. Les différences interaurales de temps (ITD) et d'intensité (ILD).

Les implantations de la HRTF les plus courantes utilisent des HRIRs obtenues à une distance fixe. Par conséquent, elles ne peuvent pas restituer les variations des différences interaurales dues aux changements de profondeur. Quelques implantations permettent de restituer les indices d'intensité de profondeur uniquement. Dans l'approche présentée dans ce chapitre, nous envisageons d'intégrer l'indice interaural de temps et l'indice d'intensité dans le modèle de la fonction HRTF.

6.2. Simulation de la profondeur des sources sonores spatiales

Dans cette section, nous décrivons quelques travaux publiés sur l'amélioration de la simulation de la profondeur de sources sonores virtuelles.

Zahorik et al. [2001] a publié un travail sur la simulation de la distance en acoustique virtuelle. Les auteurs ont proposé une méthode basée sur la pondération des différents indices sensoriels tels que la réverbération, l'intensité et les différences binaurales. Les pondérations sont déterminées pour chaque utilisateur par une méthode qu'il a défini. Les auteurs ont affirmé que cette méthode est flexible et elle est adaptable aux individus et au stimulus sonore utilisé. Cette méthode doit être adaptée à chaque utilisateur ce qui constitue son inconvénient majeur.

L'idée de modifier les différences interaurales pour améliorer la perception de profondeur a été déjà proposée et discutée par d'autres auteurs. Filipanits [1994] a proposé que les différences interaurales (ITD et ILD) d'une HRTF puissent être modifiées pour améliorer la simulation de la profondeur de sources sonores virtuelles. Sa proposition consiste à éliminer les ITD et ILD inhérentes à la HRTF utilisée et à recalculer ces différences pour les appliquer lors du traitement du signal sonore.

Fukuda et al. [2003] ont proposé une méthode pour simuler la profondeur avec la HRTF. Cette méthode est basée sur les modifications des ILDs. Ces auteurs ont réalisé une expérimentation pour comparer des HRTFs réelles avec des HRTFs à distance fixe et des ILD modifiées. Ils ont trouvé que leur méthode est efficace pour simuler la profondeur de sources sonores.

Kan et al. [2009] ont proposé l'utilisation d'une fonction pour simuler l'effet de changement de la distance de la source sonore sur les indices interaurales d'intensité. Les auteurs ont réalisé une expérimentation pour comparer leur proposition avec l'indice d'intensité. Leurs résultats montrent que l'utilisation des différences interaurales avec l'indice d'intensité permet d'améliorer la perception de la profondeur.

Les auteurs de ces approches n'ont pas effectué d'évaluation de performance basée sur des critères objectifs tels que le temps d'accomplissement de la tâche et l'erreur de localisation.

La simulation de la profondeur de sources sonore peut avoir des applications potentielles dans les interfaces homme machine comme le guidage auditif et pour améliorer le réalisme de la simulation acoustique. Par exemple, Loomis et al. [1999] ont développé un système de guidage basé sur l'audio pour lequel l'information de profondeur est importante.

6.3. Approche proposée : méthode de Différences Interaurales de Temps Compensées

6.3.1. Méthode de Différences Interaurales de Temps Compensées (CITD)

Dans cette section, nous décrivons la technique proposée pour la simulation de la profondeur de sources sonores virtuelles. Cette méthode est basée sur la compensation des ITDs d'une HRTF à distance fixe. La méthode est appelée Différences Interaurales de Temps Compensées (CITD : *Compensated Interaural Time Differences*).

La compensation réalisée consiste à garder les ITD inhérentes de la HRTF à rayon fixe et à introduire des retards de temps supplémentaires. Ces retards supplémentaires sont déterminés en fonction de la différence entre les distances suivantes : la distance entre les oreilles de l'auditeur et la HRTF ; et la distance entre les oreilles de l'auditeur et la source sonore. Cette méthode est différente de celle proposée par Filipanits [1994] car ce dernier ne tient pas compte des ITD inhérentes à la HRTF dans son modèle.

Pour le calcul des CITD, nous utilisons un modèle simplifié de la tête de l'auditeur. Dans ce modèle les positions initiales des oreilles gauche et droite sont définies par les coordonnées $(-0.13, 0, 0)$ [mètres] et $(0.13, 0, 0)$ [mètres] respectivement. Ces positions représentent les entrées aux oreilles internes. Les mouvements de tête effectués par l'utilisateur sont également appliqués au modèle virtuel de la tête.

Le modèle de la méthode de CITD est décrit ci-après. On calcule la distance entre la position à laquelle la HRTF utilisée a été mesurée et les oreilles gauche et droite du modèle de la tête de l'auditeur : Q_L et Q_R respectivement. Ensuite, on détermine les distances entre la position de la source sonore virtuelle et les oreilles gauche et droite de l'auditeur : S_L et S_R respectivement. La Figure 6.2 représente une illustration de ces distances.

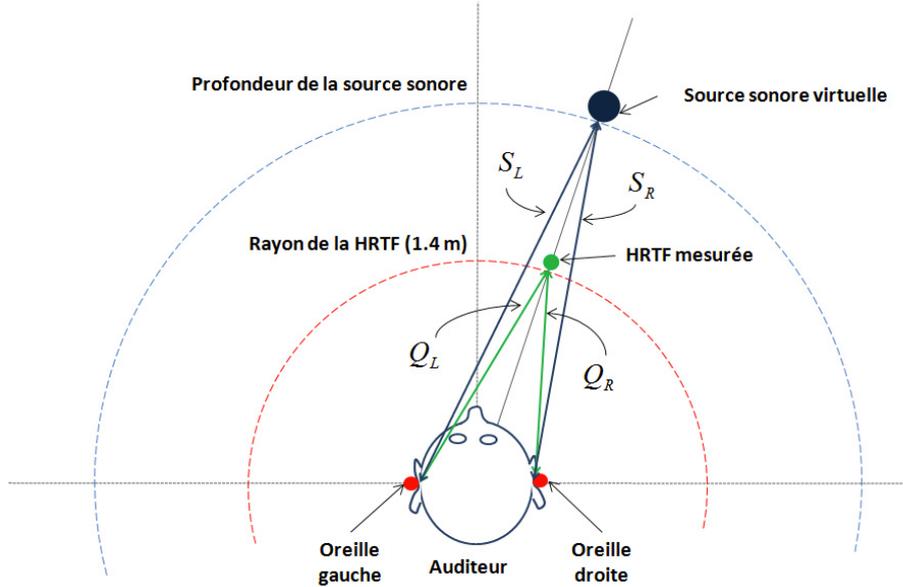


Figure 6.2. Représentation des distances comprises entre les oreillers de l'auditeur et la HRTF utilisée et des distances comprises entre les oreilles de l'auditeur et la source sonore.

Ensuite, deux indicateurs sont définis pour mesurer la différence des mesures pour les oreilles gauche et droite. Ainsi, Δ_L est la différence absolue (en mètres) des distances qui séparent l'oreille gauche des coordonnées de la HRTF utilisée et de la source sonore virtuelle. Ceci est défini comme suit :

$$\Delta_L = |S_L - Q_L| \quad (6.2)$$

où S_L [mètres] est la distance entre l'oreille gauche et la source sonore, et

Q_L [mètres] est la distance entre l'oreille gauche et la HRTF utilisée.

De la même manière Δ_R est la différence absolue pour l'oreille droite et elle est définie comme suit :

$$\Delta_R = |S_R - Q_R| \quad (6.3)$$

où S_R [mètres] est la distance entre l'oreille droite et la source sonore, et

Q_R [mètres] est la distance entre l'oreille droite et la HRTF utilisée.

La valeur principale de la méthode est le retard CITD qui représente le décalage supplémentaire à appliquer aux HRIRs utilisées. Le retard CITD est défini en fonction du retard de base ℓ . Ce dernier est mesuré en échantillons et il correspond au décalage de zéros (*zero shift*) à appliquer à l'HRIR droite. Il est défini comme suit :

$$\ell = k \frac{\Delta_L - \Delta_R}{v} f \quad (6.4)$$

où f [hertz] est la fréquence d'échantillonnage du stimulus sonore,

v [$\frac{m}{s}$] est la vitesse de propagation du son, et

k [échantillons] est un coefficient d'ajustement qui permet d'adapter le modèle aux dimensions de l'environnement virtuel.

La valeur de v est fixée à 346 [$\frac{m}{s}$] qui correspond à la vitesse du son à une température ambiante de 25°C . Le stimulus sonore utilisé dans cette thèse était échantillonné à $48,000$ [hertz], pour cela le paramètre f est fixé à cette valeur.

L'intervalle du retard CITD dépend de la longueur de l'entête des zéros des HRIR utilisées. Dans la base de données des HRTFs utilisées, les HRIRs ont une longueur de 128 échantillons desquels les premiers 50 échantillons forment une entête de zéros s . De cette manière, l'intervalle du retard CITD est de $[-50,50]$ pour l'implantation réalisée.

La valeur du coefficient k est déterminée comme suit :

$$k = (100[m]) \frac{s}{d} \quad (6.5)$$

où s [échantillons] est la longueur de l'entête de zéros des HRIRs (50 échantillons dans ce cas), et

d [mètres] est la distance maximale de la source sonore qui dépend des dimensions de l'environnement virtuel ainsi que des besoins spécifiques de l'application.

Pour cette implantation, la distance maximale est fixée à 5 mètres. Le paramètre d représente également la distance maximale à partir de laquelle la technique CITD n'a aucun effet.

Finalement le retard CITD est défini comme suit :

$$\begin{aligned} \text{si } |\ell| \leq s & \quad \text{CITD} = \ell \\ \text{si } |\ell| > s & \quad \text{CITD} = \frac{|\ell|}{\ell} s \end{aligned} \quad (6.6)$$

où ℓ est le retard de base et s est la longueur de l'entête de zéros des HRIRs utilisées.

Si le retard CITD est positif, il est appliqué à la HRIR gauche. En revanche, si le retard CITD est négatif, il est appliqué à la HRIR droite.

L'application d'un retard à une des HRIRs consiste à enlever un nombre de points égal à la valeur absolue du retard CITD au début de la fonction HRIR appliquée à l'autre oreille. Par exemple, si le retard est appliqué à la HRIR gauche, la HRIR droite sera décalée vers la gauche de la valeur du retard CITD calculé. De cette manière, l'application du retard permet de compenser les HRIRs de la HRTF utilisée en fonction des variations de la profondeur. La Figure 6.3 montre l'application du retard CITD aux HRIRs en fonction de son signe.

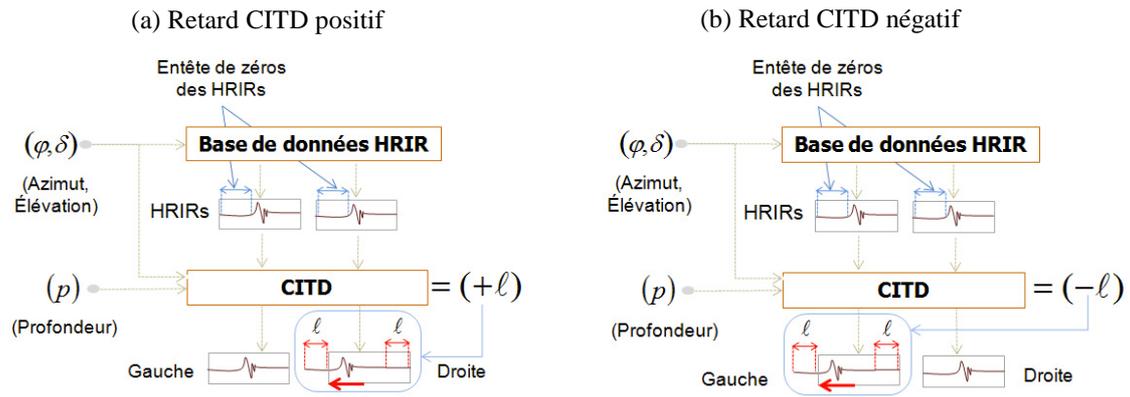


Figure 6.3. Applications du retard CITD aux HRIRs en fonction de son signe : (a) positif et (b) négatif.

La Figure 6.4 montre les valeurs du retard CITD pour des sources sonores à différentes distances et avec différentes provenances spatiales. Dans cette figure, on constate que le retard CITD est plus important pour les sources sonores avec un azimut différent de 0° , 90° , 180° et -90° .

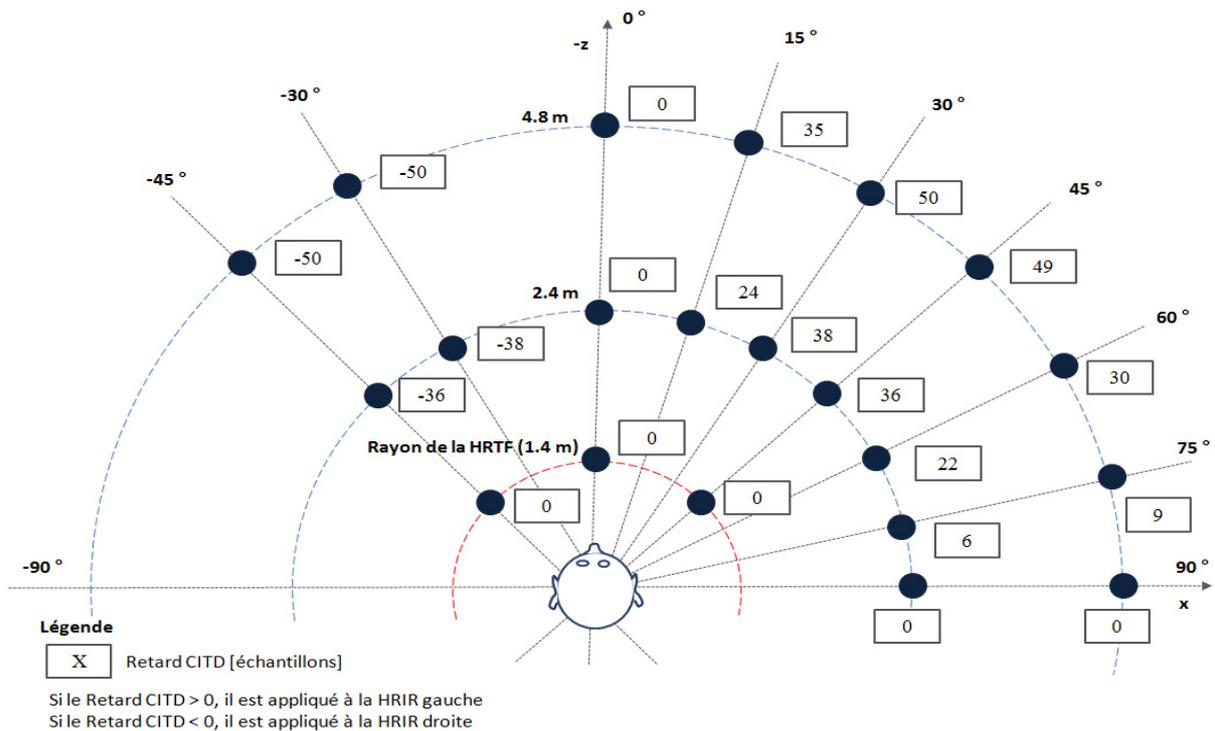


Figure 6.4. Exemples des valeurs de CITD pour différentes profondeurs et localisations de sources sonores.

La Figure 6.5 montre les deux méthodes de simulation de profondeur utilisées dans l'étude présentée dans ce chapitre. Les deux méthodes sont basées sur l'utilisation de la technique de la HRTF pour la

directionalisation des sources sonores. La première méthode (Figure 6.5a) permet de simuler la profondeur avec la variation de l'intensité de la source sonore en fonction de la distance entre la source et l'auditeur. La deuxième méthode (Figure 6.5b) se base également sur l'indice d'intensité mais avec l'incorporation du retard CITD. Dans cette figure, on constate la différence entre les deux méthodes ainsi que l'étape à laquelle le retard CITD est inséré.

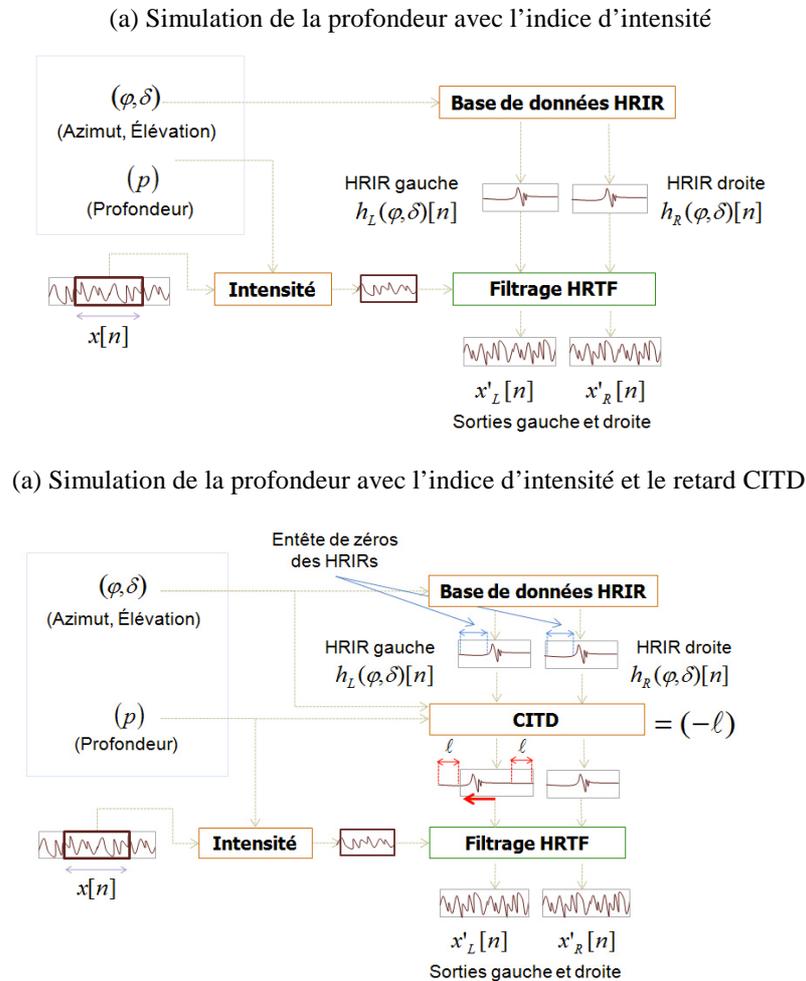


Figure 6.5. Méthodes de simulation de la profondeur de sources sonores spatiales basées sur (a) l'indice d'intensité et sur (b) la combinaison de l'indice d'intensité et le retard CITD.

Cette méthode permet de compenser les ITDs en fonction du changement de la profondeur de la source sonore. Néanmoins, elle ne prend pas en compte les variations des ILDs qui sont dues à l'effet d'occlusion de la tête. Les variations des ILDs pour les sources sonores situées dans le champ intermédiaire (1-7 mètres) sont négligeables car elles commencent à avoir des effets perceptibles à partir des changements de profondeur supérieurs à 20 m (Goldstein [1999]).

6.3.2. Simulation de la profondeur basée sur l'indice d'intensité

Dans cette partie, on décrit la méthode de simulation de la profondeur basée sur l'indice d'intensité. L'atténuation de l'intensité de la source sonore en fonction de sa profondeur est définie par l'équation (6.1). L'amplitude d'une source sonore A à une distance r est déterminée comme suit :

$$A = A_0 \cdot 10^{\frac{\Gamma}{10}} \quad (6.7)$$

où A_0 est l'amplitude originale du son à la source,
 Γ est l'atténuation du son en decibels (dB) à une distance r , et
 A est mesurée en [*watts / mètres²*].

La Figure 6.6 représente la courbe de la perte de puissance en fonction de la profondeur de la source sonore.

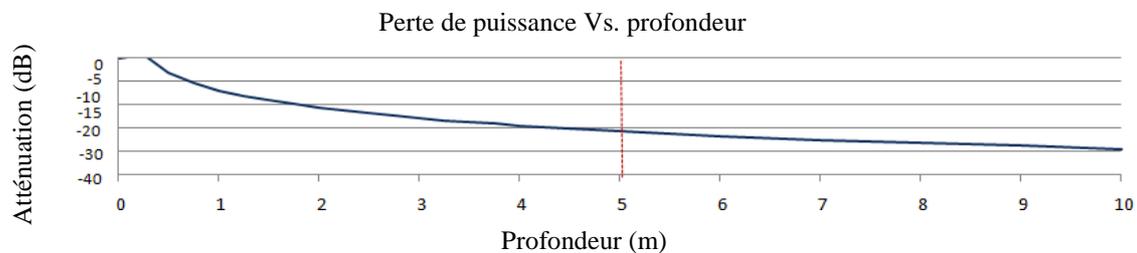


Figure 6.6. Courbe de la perte de puissance de la source sonore en fonction de sa profondeur pour la méthode de l'indice d'intensité. La ligne rouge représente la distance de 5 mètres à partir de laquelle la technique de CITD n'a aucun effet.

6.4. Étude expérimentale

Dans cette section, on décrit l'étude expérimentale réalisée pour valider l'approche proposée. Cette étude est séparée en deux parties : l'évaluation de la performance de l'utilisateur et l'analyse du retour d'information de l'utilisateur. Les deux parties se basent sur la réalisation de tâches différentes. L'objectif est de déterminer l'effet de la technique proposée dans ce chapitre sur la performance de l'utilisateur dans la localisation en profondeur de sources sonores spatiales fixes ; et de déterminer l'effet de la technique sur la perception de la profondeur de sources sonores qui se déplacent autour de l'auditeur.

6.4.1. Évaluation de la performance de l'utilisateur

Dans cette étude, la tâche consiste à identifier la profondeur d'un ensemble de sources sonores présentées dans deux conditions expérimentales. Le sujet utilise l'interface graphique montrée dans la Figure 6.7 pour interagir avec l'environnement virtuel. Il se sert du clavier pour indiquer la distance de la profondeur entre

1 et 7 mètres. Cette méthode de réponse est connue sous le nom de pondération directe (*direct scaling*). Le son est diffusé de manière continue jusqu'à l'obtention de la réponse de l'utilisateur.

La première condition expérimentale (C1 : modalité 1 ou Normale) se réfère à la technique basée uniquement sur l'indice d'intensité. La deuxième condition (C2 : modalité 2 ou CITD) se réfère à la technique qui combine l'indice d'intensité et l'approche de CITD. Les conditions expérimentales sont présentées aux sujets dans un ordre aléatoire.

Tableau 6.1. Essais expérimentales : Localisation et profondeur des sources sonores à présenter à l'utilisateur

Essai	Azimut (φ_0)	Élévation (δ_0)	Profondeur (p)
P1	15	0	1
P2	15	0	2
P3	15	0	3
P4	15	0	4
P5	15	0	5
P6	15	0	6
P7	15	0	7
1	15	0	7
2	15	0	4
3	15	0	2
4	15	0	1
5	15	0	6
6	15	0	3
7	15	0	5
8	45	0	5
9	45	0	7
10	45	0	3
11	45	0	2
12	45	0	4
13	45	0	1
14	45	0	6
15	60	0	7
16	60	0	4
17	60	0	2
18	60	0	1
19	60	0	6
20	60	0	3
21	60	0	5
22	30	0	5
23	30	0	7
24	30	0	3
25	30	0	2
26	30	0	4
27	30	0	1
28	30	0	6

La Tableau 6.1 montre les essais préliminaires (P1-P7) et les essais à analyser (1-28). Dans ces essais les sources sonores sont situées dans le plan horizontal afin de réduire le nombre total d'essais. Les sept essais préliminaires permettent aux sujets de se familiariser avec la tâche et la méthode de pondération directe. Dans les essais préliminaires, la réponse correcte est marquée par une flèche en bleu qui n'est pas affichée dans les

autres essais (Figure 6.7a et Figure 6.7b). Les essais préliminaires (P1-P7) ne sont pas pris en compte dans l'analyse statistique et ils ont été effectués selon l'ordre présenté dans le Tableau 6.1. Les essais traités dans l'analyse statistique ont été présentés au sujet dans un ordre aléatoire.

Le stimulus sonore utilisé dans cette expérience est identique à celui des études précédentes. La plateforme expérimentale utilisée dans cette étude est identique à celle décrite dans les chapitres précédents. Le suivi de mouvement de la tête n'est pas pris en compte dans cette étude expérimentale.

(a) Essais préliminaires



(b) Essais d'évaluation



Figure 6.7. Interface graphique pendant (a) les essais préliminaires et (b) les essais d'évaluation.

Seize sujets âgés entre 20 et 30 ans (14 sujets de sexe masculin et 2 sujets de sexe féminin) ont participé à ces expérimentations. Ils sont tous droitiers et ont déclaré ne pas avoir de maladies ni visuelles, ni auditives ni psychomotrices. Ils ont tous un niveau d'études supérieur sans expérience préalable en environnement virtuel. Chaque sujet passe 40 minutes environ pour réaliser la tâche dans les deux conditions expérimentales avec une pause de trois minutes à la fin de chaque série d'essais.

Les variables expérimentales objectives mesurées sont le temps d'accomplissement de la tâche (en secondes) et l'erreur absolue de localisation en profondeur (en mètres). La Figure 6.8a montre un sujet en train de réaliser l'expérience.

(a) Évaluation de la performance



(b) Évaluation du retour d'informations de l'utilisateur



Figure 6.8. Expérimentations de (a) l'évaluation de la performance de l'utilisateur et (b) de l'évaluation du retour d'informations de l'utilisateur.

6.4.2. Retour d'informations de l'utilisateur

La tâche consiste à écouter une source sonore qui se déplace en suivant des trajectoires prédéfinies. Le mouvement de la source sonore de long de chaque trajectoire est répété trois fois dans chaque condition expérimentale de manière alternée. Après chaque déplacement de la source sonore, l'utilisateur répond à une question sur la comparaison de la sensation de profondeur ressentie dans les deux conditions expérimentales (c.f. Annexe : 10.2).

La Figure 6.8b montre un utilisateur en train de réaliser la tâche de l'évaluation de la perception de profondeur. Quinze sujets (13 de sexe masculin et 2 de sexe féminin) ont participé à cette étude de perception de profondeur. Ils ont tous participé à la première expérience.

La Figure 6.9 représente les trajectoires suivies par les sources sonores dans cette expérience. Les trajectoires ont été choisies en considérant qu'elles couvrent une partie importante du volume de travail. Le mouvement des sources sonores est limité au plan horizontal.

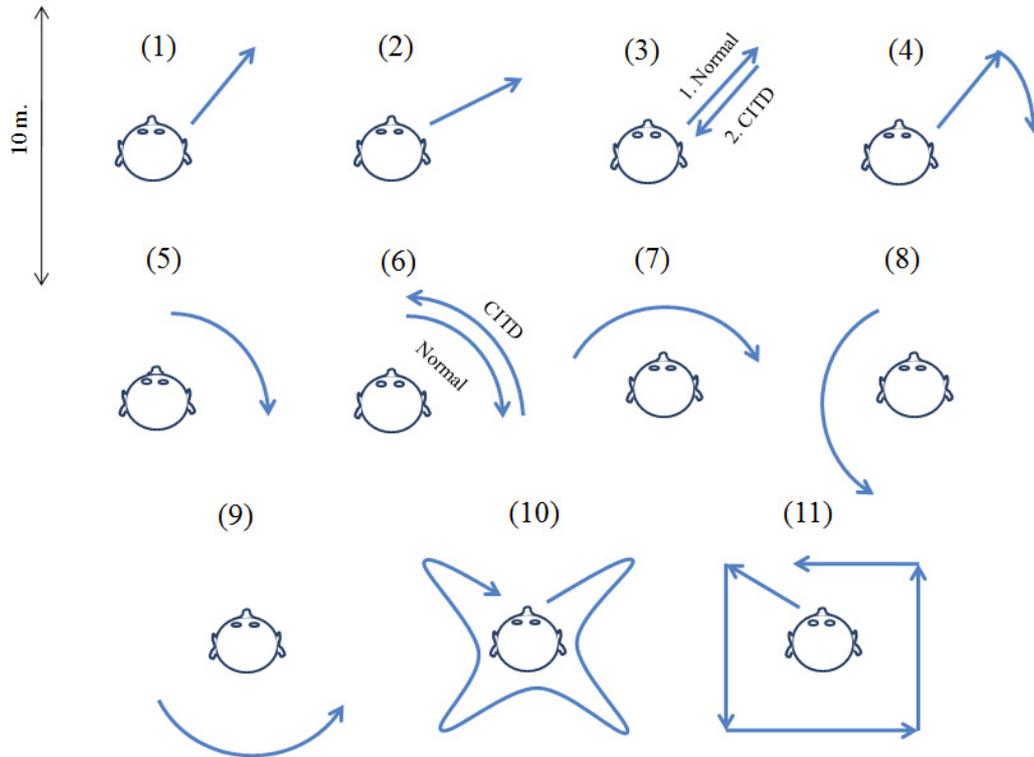


Figure 6.9. Les trajectoires suivies par les sources sonores dans l'expérience de la perception de la profondeur.

Le modèle des questions posées pour chaque trajectoire est décrit ci-dessous. Nous avons introduit une dernière question pour permettre à l'utilisateur de donner son appréciation sur la comparaison du réalisme des deux modalités évaluées.

Modèle de question pour les trajectoires : Comment vous jugez la qualité de la perception de profondeur dans la modalité 2 (CITD) comparée à la modalité 1 (Normale) pour la trajectoire X ?

- 3 Normale est beaucoup mieux
- 2 Normale est mieux
- 1 Normale est un peu mieux
- 0 CITD est aussi bien que Normale
- 1 CITD est un peu mieux
- 2 CITD est mieux
- 3 CITD est beaucoup mieux

où $X=1,2,3,\dots,11$.

Modèle de question finale (Evaluation globale): Comment vous jugez la qualité du réalisme de la modalité 2 (CITD) comparée à la modalité 1 (Normale) pour l'ensemble des trajectoires ?

- 3 Normale est beaucoup mieux
- 2 Normale est mieux
- 1 Normale est un peu mieux
- 0 CITD est aussi bien que Normale

- 1 CITD est un peu mieux
- 2 CITD est mieux
- 3 CITD est beaucoup mieux

Nous avons utilisé le modèle de Catégorie de Notation en Dégradation (DCR) recommandée par l'ITU-T [1996] pour représenter les réponses des sujets. Au total, chaque sujet répondra à douze questions qui correspondent à onze trajectoires plus la question finale.

6.5. Résultats expérimentaux et analyse de données

6.5.1. Évaluation de performance de l'utilisateur

Dans cette sous-section, on présente et on analyse les résultats de l'évaluation de performance. On décrit les résultats statistiques. Ensuite, on présente les tests pour déterminer la nature des distributions fréquentielles. Finalement, on montre la signification statistique des facteurs expérimentaux pris en compte dans cette étude.

6.5.1.1. Caractérisation des techniques évaluées

Le Tableau 6.2 montre les résultats de l'analyse statistique des variables mesurées. Le temps d'accomplissement de la tâche est une variable continue et l'erreur absolue de localisation est une variable discrète. De ce fait, la caractérisation, la représentation et l'analyse des variables sont effectuées de manières différentes. Le temps d'accomplissement de la tâche se caractérise statistiquement par la moyenne, la médiane, l'écart-type, les quartiles Q1 et Q3 et l'intervalle interquartile. L'erreur absolue est caractérisée par la moyenne, la médiane, l'écart-type, l'intervalle de confiance, l'asymétrie et le kurtosis.

Tableau 6.2. Mesures statistiques des données obtenues

Variable	Condition	N ^{bre}	Moyenne	Médiane	Écart type	Quartiles		
						Inf.	Sup.	Inter.
Temps d'accomplissement (s)	Toutes les conditions	896	7.05	6.04	4.08	4.16	8.82	4.66
	1 Normale	448	7.08	5.76	4.38	4.00	8.70	4.70
	2 CITD	448	7.03	6.19	3.76	4.38	8.88	4.49

Variable	Condition	N ^{bre}	Moyenne	Médiane	Écart type	IC SD ± 95%		Asymétrie	Kurtosis
						Sup.	Inf.		
Erreur absolue de localisation en profondeur (m)	Toutes les conditions	896	0.92	1.00	0.87	0.84	0.92	0.88	0.61
	1 Normale	448	1.00	1.00	0.91	0.86	0.97	0.86	0.50
	2 CITD	448	0.85	1.00	0.83	0.78	0.89	0.86	0.65

Seulement les valeurs absolues des erreurs de localisation ont été prises en compte.

La Figure 6.10 montre la moyenne du temps d'accomplissement de la tâche pour chaque condition expérimentale. Les statistiques centrales (moyenne et médiane) pour les deux conditions expérimentales sont proches. L'écart pour la condition CITD est légèrement inférieur à celle de la condition Normale. On constate que le nombre de valeurs extrêmes et atypiques est légèrement inférieur pour la condition CITD comparée à la condition Normale. La signification de ces différences sera déterminée par l'analyse statistique décrite ci-après.

La Figure 6.11 montre les moyennes de l'erreur absolue de localisation en profondeur pour chaque condition expérimentale. Il existe une différence non négligeable pour la moyenne de l'erreur absolue entre les conditions expérimentales. Les intervalles de confiance ne se superposent que marginalement. On constate une réduction de 15% environ pour la moyenne dans la condition CITD.

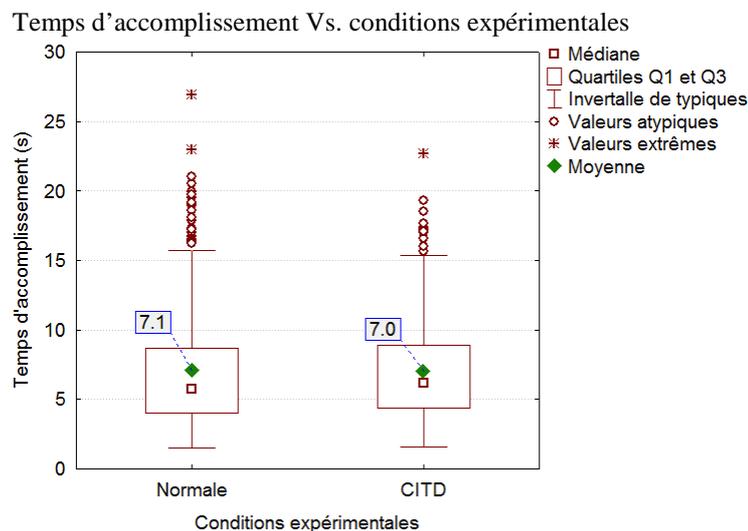


Figure 6.10. Temps d'accomplissement de la tâche pour les conditions expérimentales évaluées.

La Figure 6.12 montre la distribution de l'erreur non-absolue de localisation en profondeur. Le biais (*skewness*) observé autour de zéro pour les valeurs non absolues de l'erreur de localisation en profondeur sont négligeables (avec une médiane de 0.0 m et une moyenne de 0.3 m). Etant donné ce résultat et le fait que les valeurs absolues sont plus significatives pour analyser la valeur de l'erreur de localisation en profondeur, l'analyse effectuée prend en compte les valeurs absolues uniquement.

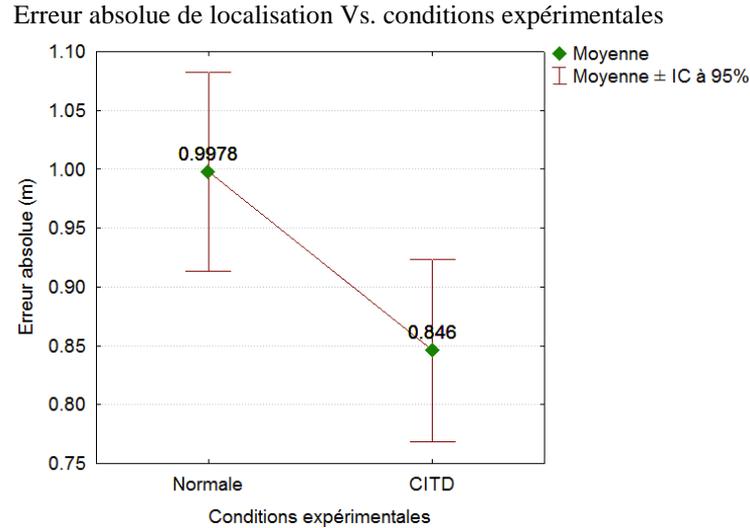


Figure 6.11. Erreur absolue de localisation en profondeur pour les conditions expérimentales évaluées.

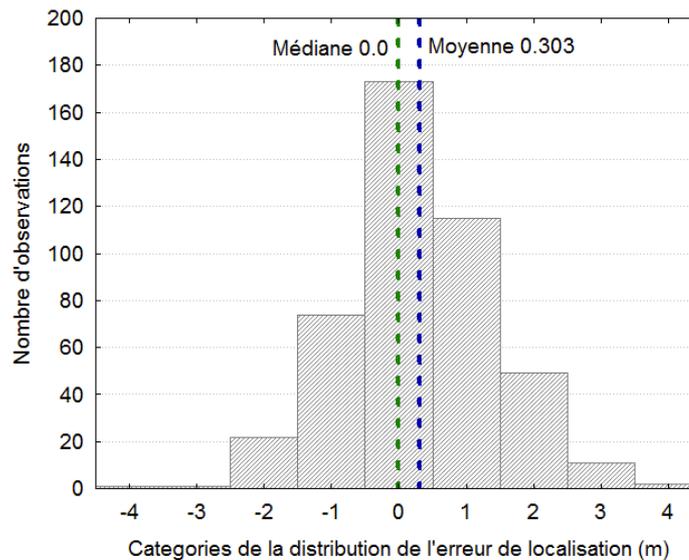


Figure 6.12. La distribution fréquentielle de l'erreur non absolue de localisation en profondeur.

La Figure 6.13a montre la performance individuelle en termes du temps d'accomplissement de la tâche dans chaque condition expérimentale. On considère qu'il n'est pas possible d'établir une différence systématique de la performance des individus entre les deux conditions expérimentales étudiées.

La Figure 6.13b montre la performance individuelle en termes de l'erreur absolue de localisation en profondeur pour chaque condition expérimentale. Nous avons constaté que dans la condition CITD, douze sujets ont obtenu une moyenne de l'erreur absolue inférieure (différence ≤ 0.15 m) ou égale à celle obtenue dans la condition Normale.

6.5.1.2. Distribution des données

Le Tableau 6.3 montre les tests de qualité d'ajustement pour déterminer le type de distribution des variables expérimentales. Ce test permet de connaître la structure de données et facilite le choix de la méthode d'analyse statistique à utiliser.

Le test montre que la variable de temps d'accomplissement de la tâche suit une distribution continue de type log-normale. La Figure 6.14a montre la distribution fréquentielle des données du temps d'accomplissement de la tâche ajustée à une fonction continue de type log-normale.

Le test montre également que la variable d'erreur absolue de localisation en profondeur suit une distribution discrète de type poisson. La Figure 6.14b montre la distribution fréquentielle des données de l'erreur absolue de localisation ajustée à une fonction discrète de type poisson.

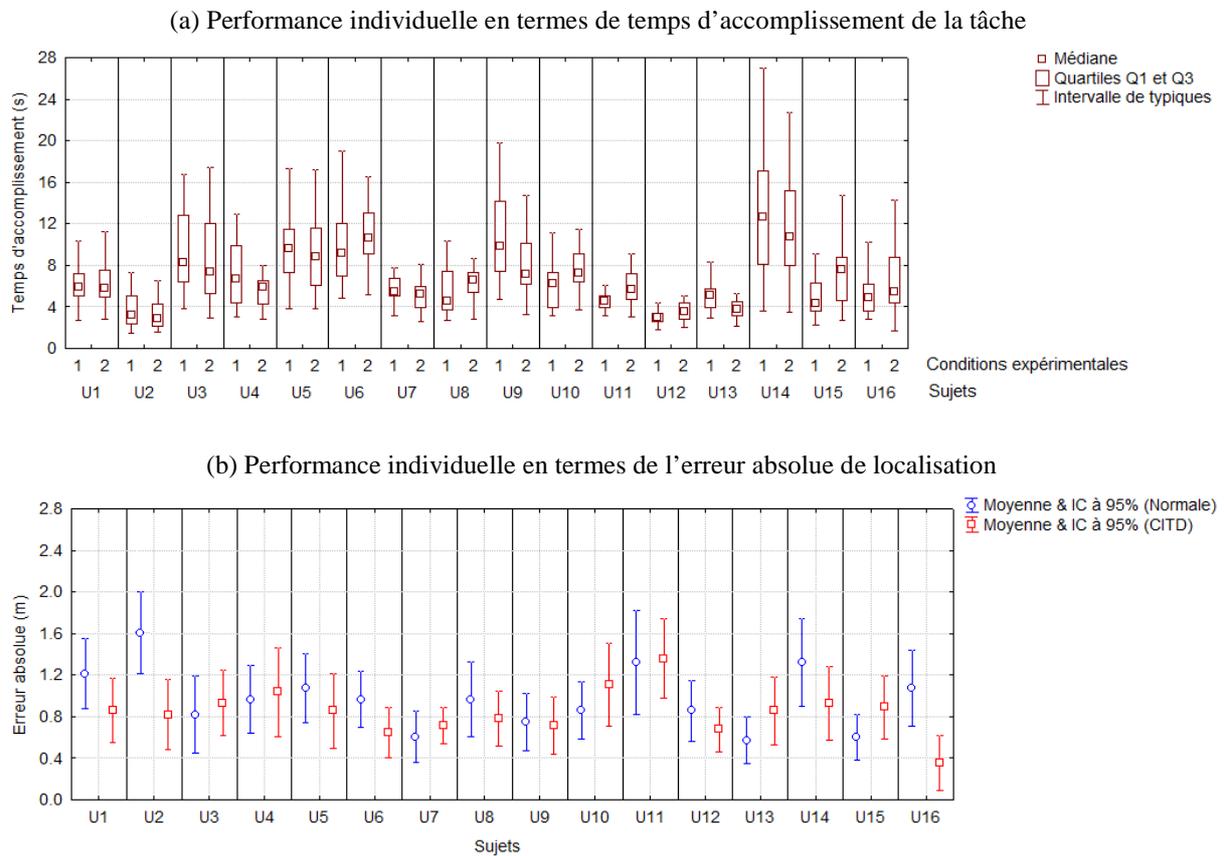


Figure 6.13. Performance individuelle : (a) temps d'accomplissement de la tâche et (b) erreur absolue de localisation. Les conditions expérimentales indiquées pour 1 et 2 représentent les modalités Normale et CITD respectivement.

Tableau 6.3. Tests de qualité d'ajustement

Fonction	Variable	Moyenne	Variance	Moyenne ajustée	Variance ajustée	Distance K-S	p	Ajustée (o/n) ?
Log-normale	Temps d'accomplissement (s)	7.05	16.64	1.81	0.28	0.027	>0.05	O

Fonction	Variable	Moyenne	Variance	Lambda	Distance K-S	p	Ajustée (o/n)??
Poisson	Erreur absolue de localisation en profondeur (m).	0.92	0.76	0.92	0.045	0.05 < p < 0.1	O

La méthode de comparaison utilisée est le test Kolmogorov-Smirnov (K-S) à un niveau de confiance de 95%. Uniquement la valeur absolue de l'erreur de localisation en profondeur a été prise en compte.

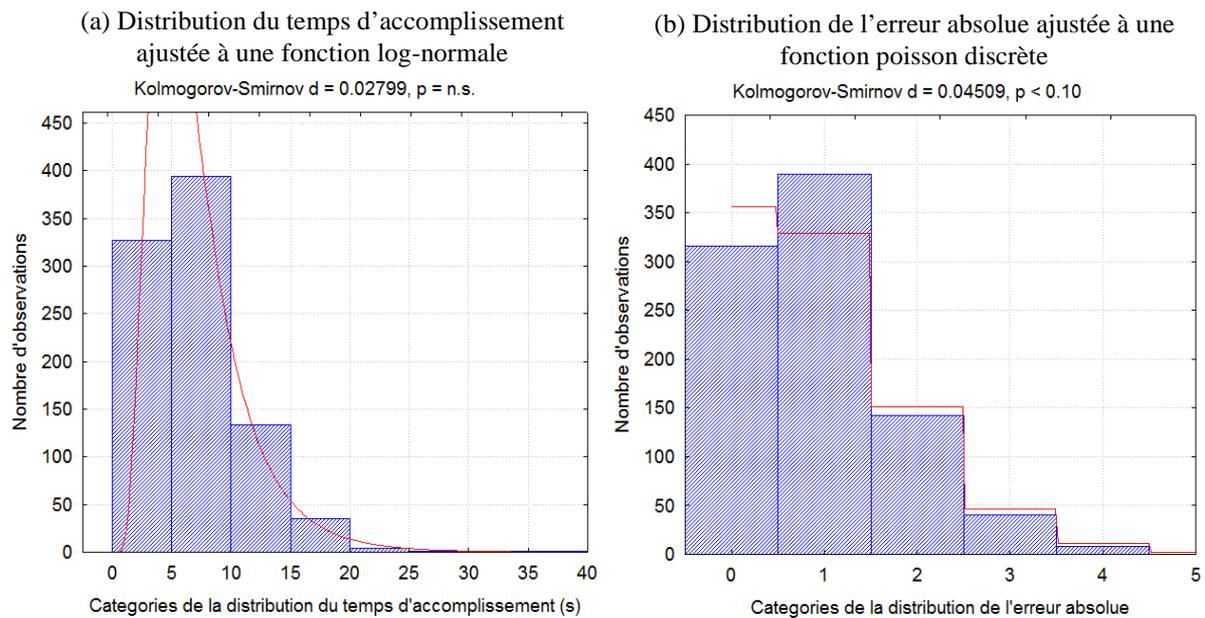


Figure 6.14. La distribution de fréquence (a) du temps d'accomplissement et (b) de l'erreur absolue de localisation en profondeur.

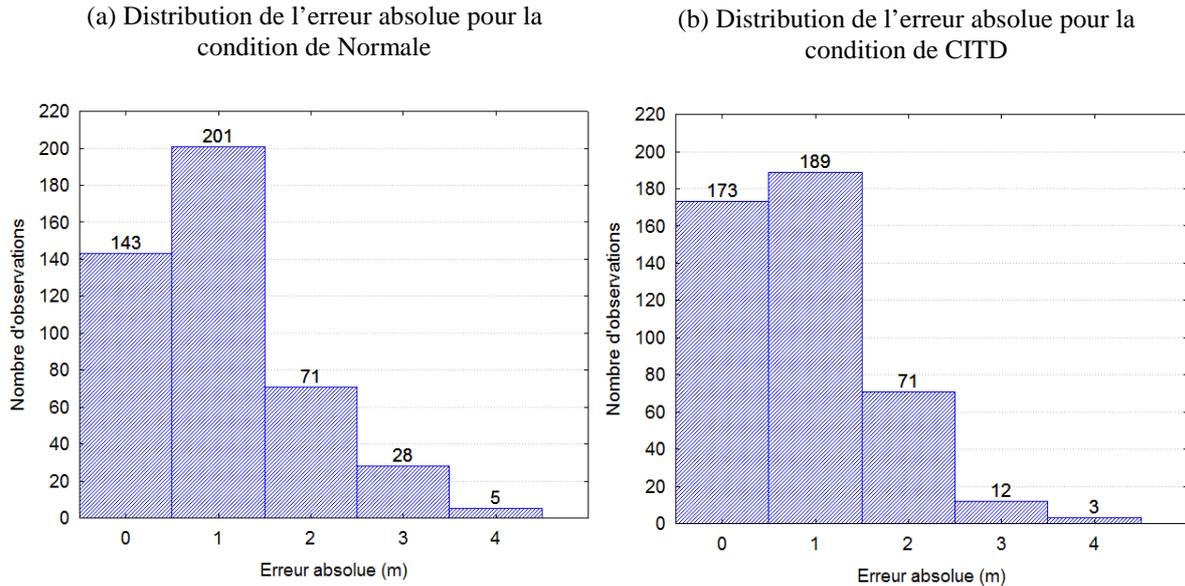


Figure 6.15. La distribution de fréquence de l'erreur absolue pour les échantillons correspondants aux conditions expérimentales (a) Normale et (b) CITD.

La Figure 6.15 montre les formes des distributions discrètes de la variable d'erreur absolue pour les données obtenues pour les deux conditions expérimentales : (a) Normale et (b) CITD. Cette figure permet de constater la différence entre les résultats des deux conditions expérimentales. Le nombre d'erreurs de valeur 0 (ce qui signifie une réponse exacte) augmente dans la condition CITD. Le résultat indique que la condition expérimentale CITD permet d'augmenter le nombre de réponses exactes et de réduire le nombre d'erreurs par rapport à la condition normale.

6.5.1.3. Analyse statistique

Nous avons utilisé la méthode d'analyse statistique des Equations d'Estimation Généralisées. Le Tableau 6.4 montre le résultat de l'analyse statistique pour le facteur : condition expérimentale.

Dans cette analyse, on constate que la condition expérimentale n'a pas d'effet significatif sur le temps d'accomplissement de la tâche ($p=0.865$). En revanche, l'effet de la condition expérimentale sur le temps d'accomplissement de la tâche est significatif ($p=0.017$). De ce fait, on constate que l'approche de CITD permet d'améliorer la précision sans modifier le temps nécessaire à la détermination de la profondeur de la source sonore spatiale.

Tableau 6.4. Analyse de l'effet des facteurs expérimentaux par l'Analyse de Mesures Répétées avec le modèle d'Equations Généralisées d'Estimation

Facteur expérimental	Variable	Modèle	Wald Chi-square (χ^2)	DDL	p	Significatif (o/n) ?
Conditions expérimentales	Temps d'accomplissement (s)	Log-normal	0.030	1	0.865	N
	Erreur absolue de localisation (m)	Poisson	0.598	1	0.017	O

L'estimation des effets est basée sur la mesure statistique Wald chi-square. Le test est de type III pour lequel l'ordre d'agrégation des facteurs n'est pas important. Le niveau de confiance est de 95%.

6.6. Analyse du retour d'informations de l'utilisateur

Dans cette partie, on présente les résultats obtenus pour l'analyse du retour d'informations de l'utilisateur sur la perception de la profondeur de sources sonore. La Figure 6.16 montre les résultats obtenus pour chaque trajectoire évaluée. Le résultat montre que pour toutes les trajectoires (sauf la sixième), la plupart des sujets préfèrent la sensation de profondeur fournie par l'approche CITD.

Pour l'évaluation générale, douze sujets sur quinze considèrent que le réalisme de l'approche CITD est mieux ou un peu mieux que celui de la condition Normale.

Ce résultat montre que les sujets préfèrent la sensation de profondeur fournie par l'approche CITD avec des sources sonores mobiles. Ainsi, nous pouvons conclure que l'effet de l'approche CITD dans la perception de la profondeur est significatif.

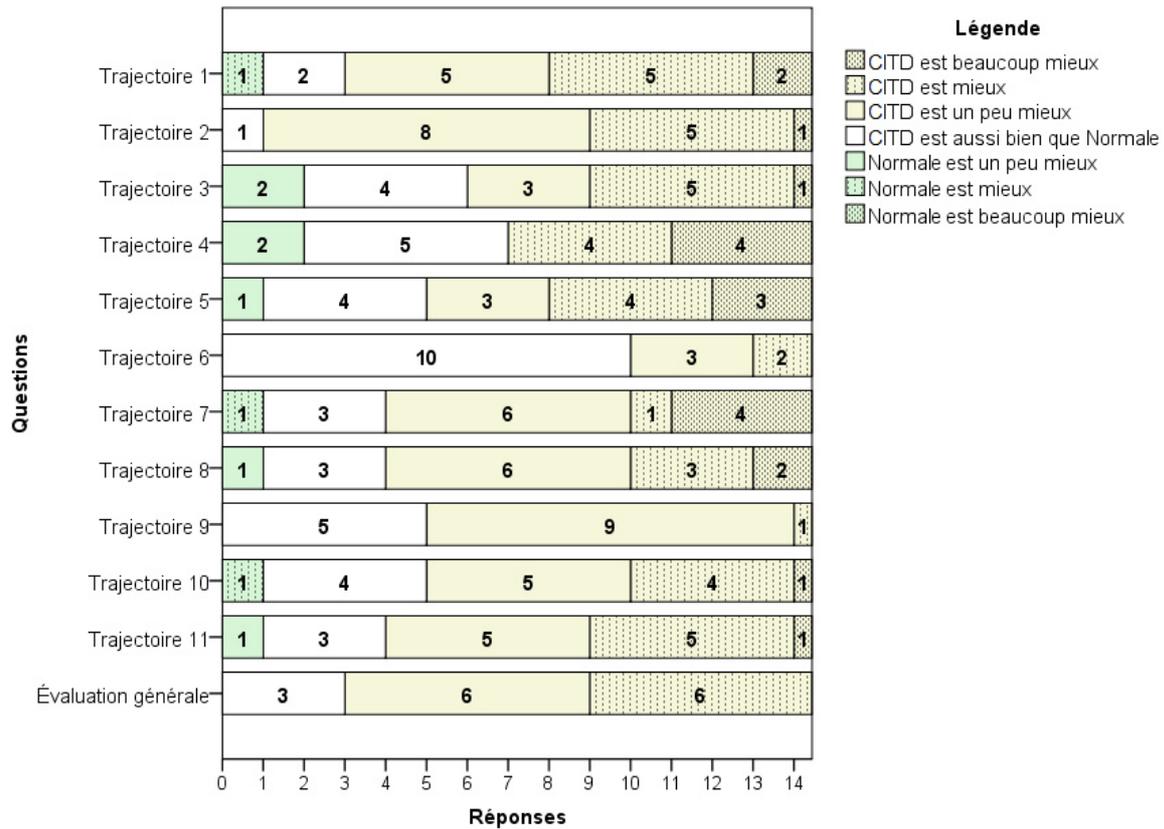


Figure 6.16. L'appréciation des sujets pour la perception de la profondeur dans les deux conditions expérimentales étudiées.

6.7. Conclusions

Dans les chapitres précédents, nous avons présenté des approches pour obtenir une localisation rapide et précise de la provenance des sources sonores. Néanmoins, l'amélioration de la localisation en profondeur de sources sonores n'avait pas été étudiée.

Plusieurs méthodes d'acoustique virtuelle n'utilisent que l'indice d'intensité pour simuler la profondeur de sources sonores (distance entre la source sonore et l'auditeur). Ceci implique que la profondeur sentie est moins fidèle. Dans le cadre de cette thèse, nous nous sommes intéressés à déterminer si l'agrégation des informations des différences interaurales de temps (ITDs) permet d'améliorer la perception de la profondeur. L'amélioration de la localisation en profondeur présente un intérêt dans les applications de réalité virtuelle développées dans cette thèse.

Une partie des techniques existantes pour l'amélioration de la perception de la profondeur nécessite l'adaptation à chaque individu. L'individualisation n'est pas une option très pratique dans les applications envisagées pour les raisons discutées dans les chapitres précédents.

Nous avons présenté une nouvelle approche pour simuler la profondeur à l'aide de la fonction HRTF basée sur la compensation des ITDs en fonction de la distance entre l'auditeur et la source sonore. Une étude de

validation expérimentale a été effectuée pour déterminer l'effet de cette approche sur la performance de localisation en profondeur et sur la perception de la profondeur.

L'évaluation de la performance révèle que l'approche CITD a un effet statistiquement significatif sur la précision de la localisation en profondeur avec une amélioration de 15% environ par rapport à l'utilisation de l'indice d'intensité uniquement. Il est à noter également que l'approche CITD n'a pas d'effet sur le temps d'accomplissement de la tâche.

L'analyse du retour d'informations des utilisateurs révèle que l'approche CITD est préférée par la plupart des sujets qui ont également considéré que cette approche est plus réaliste que celle basée sur l'indice d'intensité uniquement.

7. Application du Son 3D pour la simulation de chirurgie arthroscopique

Dans ce chapitre, on présente l'approche du Son 3D Enrichi qui consiste en la combinaison de la fonction HRTF, des indices sonores SASA et de la technique CITD. On présente également une étude expérimentale qui utilise l'approche du Son 3D Enrichi (S3E) en environnement virtuel dans une application de simulation de chirurgie arthroscopique. L'objectif de cette étude expérimentale est de déterminer l'effet du guidage auditif sur l'optimisation du geste de triangulation qui consiste en la manipulation d'un instrument et d'une caméra à l'intérieur du corps du patient.

7.1. La simulation de chirurgie arthroscopique

7.1.1. La simulation chirurgicale

La simulation chirurgicale consiste en l'utilisation des techniques de modélisation et d'interaction pour restituer les conditions de réalisation d'une chirurgie réelle. Elle a pour objectif la formation d'apprentis chirurgiens dans l'acquisition de gestes chirurgicaux complexes. La simulation chirurgicale permet aux apprentis chirurgiens d'acquérir de l'expérience et d'améliorer leurs compétences avant d'opérer des patients (Sutherland [2006]).

La chirurgie mini-invasive (MIS) regroupe les chirurgies thérapeutiques et diagnostiques effectuées à l'aide d'outils introduits par de petites incisions (García Berro et Toribio [2004]). La manipulation des instruments est réalisée à l'aide de l'endoscopie ou d'autres techniques de visualisation indirecte.

La chirurgie mini-invasive offre les avantages suivants : la réduction du saignement pendant l'intervention, l'apparition de cicatrices moins marquées, la réduction des risques d'infection et la réduction du coût et du temps d'hospitalisation. La chirurgie MIS utilise différentes techniques d'opération selon la partie du corps humain à traiter. Par exemple : la laparoscopie s'occupe de l'abdomen et l'arthroscopie concerne les grandes articulations comme les épaules et les genoux.

La chirurgie mini-invasive implique l'apprentissage et la réalisation de gestes plus complexes que ceux de la chirurgie classique. En effet, une opération mini-invasive se réalise avec une vue indirecte et limitée obtenue à l'aide d'une caméra (endoscope). La vue fournie par la caméra est indirecte parce que l'orientation de

la caméra ne correspond pas directement à celle de l'utilisateur. La vue est limitée parce que la composition de l'intérieur du corps humain complique souvent le positionnement de la camera pour couvrir adéquatement l'espace de travail.

Le chirurgien effectue le geste de triangulation pour positionner l'instrument chirurgical en ayant une vue indirecte et limitée. Cette situation complique la réalisation du geste de triangulation car l'information visuelle est souvent insuffisante pour exécuter la tâche et le sens des mouvements effectués par l'utilisateur ne correspondent pas directement à celui des mouvements visualisés à l'écran.

Il existe différents simulateurs de chirurgie arthroscopique. K. P Sherman et al. [2001] ont développé le système VE-KATS (Knee Arthroscopy Training System) et Moody et Waterworth [2004] ont développé le système WISHKATS (Warwick Imperial, Sheffield Haptic Knee Arthroscopy Training System). Ces deux systèmes sont utilisés pour entrainer des apprentis chirurgiens dans le geste de triangulation et dans le diagnostique de l'arthroscopie du genou. L'interaction est basée sur la restitution des sensations visuelles et haptiques.

Le système utilisé dans notre étude est décrit dans le travail de Sofía Bayona et al. [2006]. Ce système est utile pour entrainer des apprentis à la chirurgie arthroscopique de l'épaule et du genou. Il possède une architecture modulaire qui permet d'ajouter de nouveaux composants.

L'économie de mouvement consiste à minimiser les mouvements nécessaires et leurs amplitudes pour effectuer une tâche donnée. Dans le domaine de la chirurgie, l'économie de mouvement est un aspect important car il permet de réduire les risques inhérents à l'intervention.

Les avantages de la simulation chirurgicale dans l'entraînement d'apprentis chirurgiens ont été montrés par Gurusamy et al. [2009]. Clevin et Grantcharov [2008] ont montré que les techniques de simulation chirurgicale contribuent à l'économie de mouvement dans la chirurgie laparoscopique. Tuchschnid et al. [2007] ont proposé une méthode pour évaluer la performance des utilisateurs d'un système d'hystérocopie virtuelle. La mesure d'économie de mouvement proposée par les auteurs est calculée à partir de la longueur de la trajectoire parcourue par l'instrument.

Sofía Bayona et al. [2008] ont effectué une étude expérimentale sur l'économie de mouvement dans la simulation de chirurgie arthroscopique réalisée avec 94 apprentis chirurgiens. Les expérimentations de ce travail ont été réalisées sur le même simulateur que celui utilisé dans notre étude. Les chirurgiens considèrent que l'entraînement dans le geste de triangulation est très important pour la formation de futurs chirurgiens. Les auteurs ont conclu que le simulateur est plus bénéfique pour les chirurgiens novices que pour les experts.

Vazquez Buenos-Aires et al. [2003] ont proposé l'utilisation d'un dispositif d'affichage par la langue (TDU : *Tongue Display Unit*) pour le guidage du geste dans la simulation chirurgicale. L'inconvénient principal de cette approche est que le retour sensoriel se base sur l'utilisation d'un diapositif qui peut se révéler comme envahissant.

7.1.2. Le geste de triangulation et la manipulation des instruments

L'utilisateur manipule l'instrument et la caméra en les tenant par les pointes extérieures alors que l'opération principale s'effectue à l'intérieur du corps du patient avec les pointes localisées à l'intérieur du corps.

Les mouvements de débattement de l'instrument et de la caméra à l'intérieur du corps sont restreints à un cône de pénétration. L'extension de ce cône dépend d'un dispositif (ayant la forme d'un entonnoir) situé à

l'entrée des incisions pour faciliter l'insertion et la manipulation de l'instrument et de la caméra. Ce cône possède une ouverture maximale de 45° environ qui limite les mouvements de débattement angulaire à cette valeur.

La manipulation est décomposée en trois mouvements de base : le débattement latéral (Figure 7.1a), le débattement vertical (Figure 7.1b) et la pénétration (Figure 7.1c).

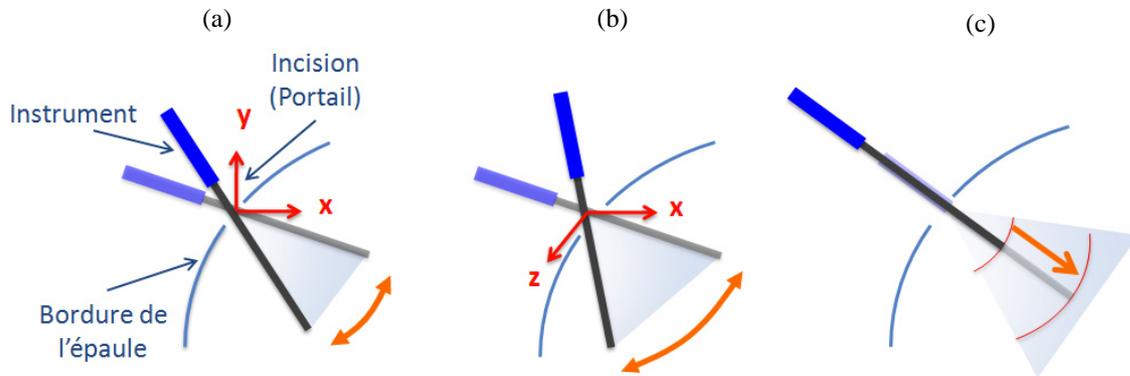


Figure 7.1. Mouvements de base du geste de triangulation : (a) latéral (b) vertical et (c) pénétration.

7.2. Le simulateur de chirurgie arthroscopique utilisé

Les expérimentations présentées dans ce chapitre ont été développées à l'aide d'un simulateur de chirurgie arthroscopique et d'un modèle d'épaule humaine. Le simulateur utilisé est l'Insight ArthroVR commercialisé par la société GMV Innovating Solutions (GMV [2009]).

Le simulateur est constitué de quatre modules principaux : le noyau (*kernel*) de simulation, l'interfaçage, le moteur physique, le rendu graphique et le rendu haptique. Les principales composantes matérielles de la plateforme expérimentale sont (Figure 7.2) :

- Un modèle en plastique d'une épaule humaine à l'échelle réelle. Ce modèle possède quatre incisions (portails).
- Deux dispositifs de retour haptique Phantom Omni (Sensable Technologies [2010]). Une extension métallique a été ajoutée à chaque dispositif pour représenter la longueur des instruments utilisés dans l'arthroscopie.
- Une base permettant d'ajuster les positions des interfaces haptiques et le modèle de l'épaule.
- Deux écrans LCD : un pour l'utilisateur et l'autre pour l'expérimentateur. L'écran de l'utilisateur montre l'image de la caméra (Figure 7.3a). Le second écran montre une vue générale de la scène virtuelle (Figure 7.3b).

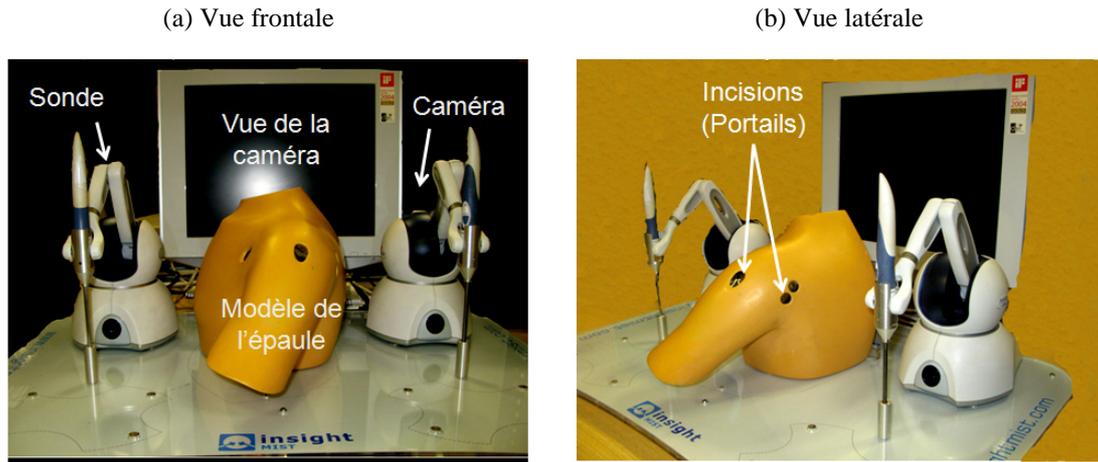


Figure 7.2. Simulateur de chirurgie arthroscopique utilisé.

L'image de la camera, qui correspond à la vision d'un arthroscopie réel, est une vue tridimensionnelle des principaux éléments à l'intérieur de l'épaule (les os, la peau, les muscles et les tendons). La scène a été créée à l'aide du logiciel Coin3D (Kongsberg Oil Gas Technologies [2010]). Dans la configuration utilisée du simulateur, l'interface haptique située dans la partie droite du simulateur est associée à la caméra et celle située dans la partie gauche est associée à l'instrument.

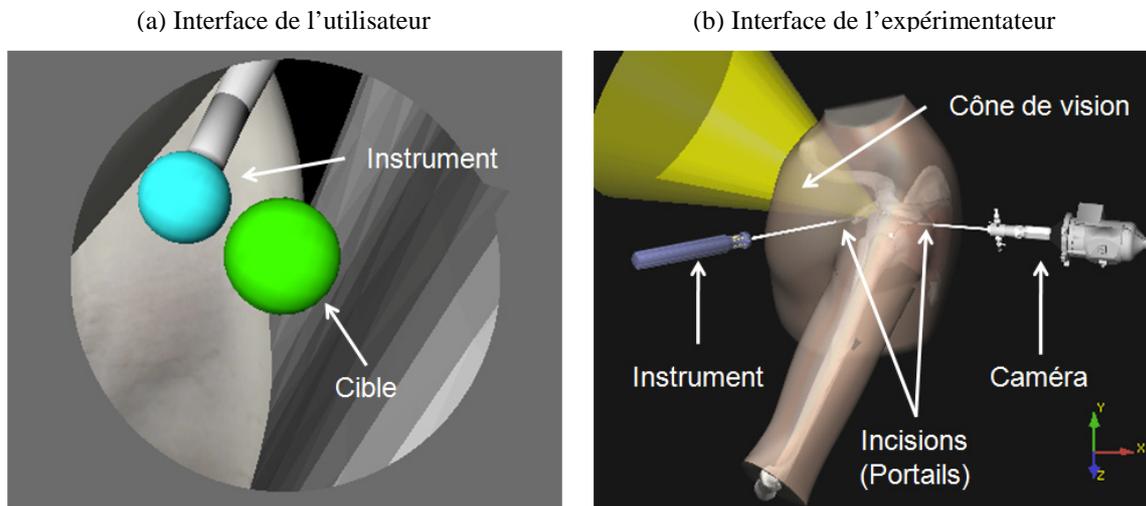


Figure 7.3. Interfaces graphiques du simulateur chirurgical.

Les modèles graphiques de l'intérieur et de l'extérieur de l'épaule utilisés dans le simulateur, sont montrés dans la Figure 7.4. Les unités dimensionnelles du simulateur sont les pixels. Le rapport entre les dimensions de l'environnement virtuel et celles du modèle physique de l'épaule est de 2 pixels par 1 centimètre.

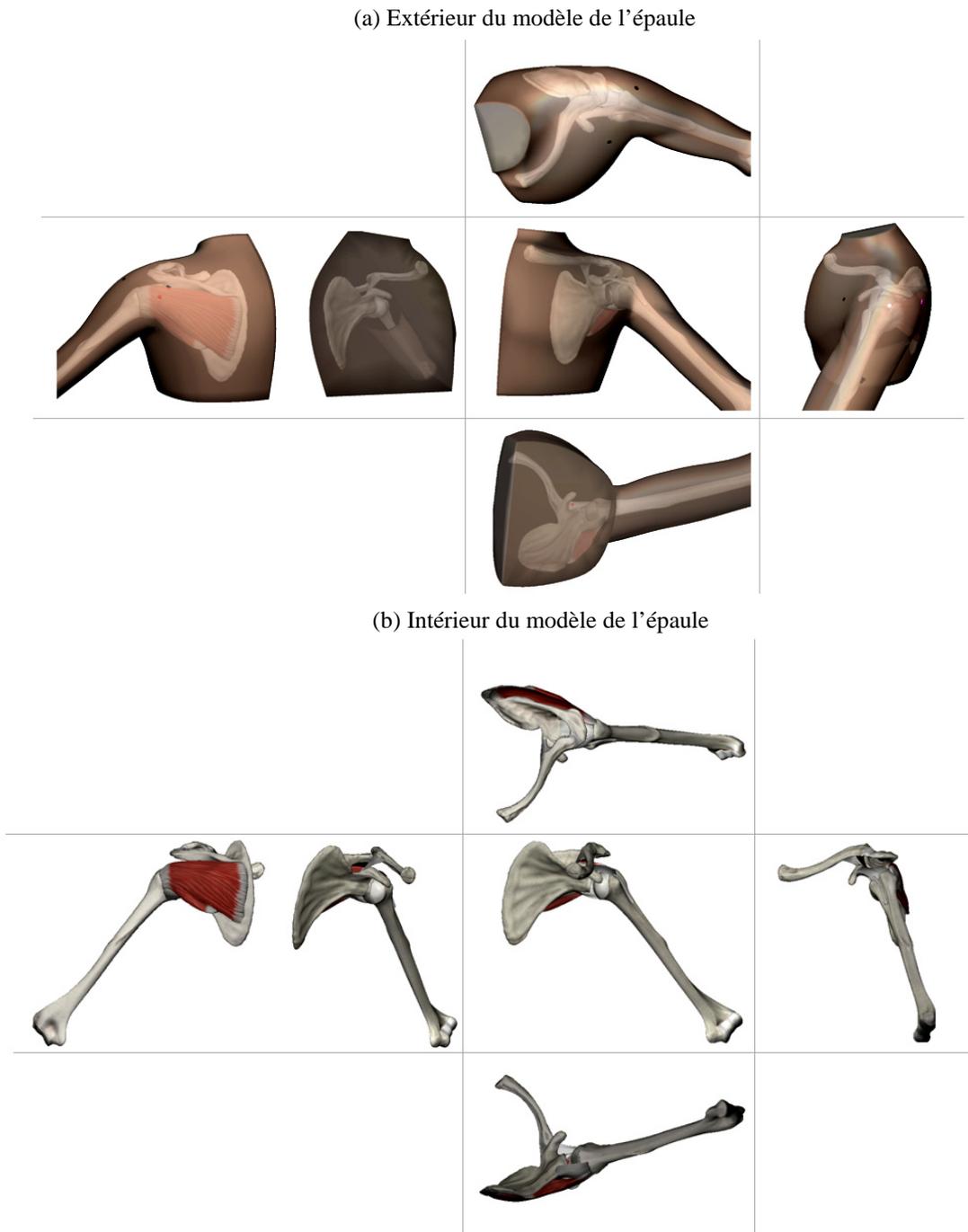


Figure 7.4. Modèles graphiques de l'épaule.

7.3. Approche : le Son 3D Enrichi pour guider le geste de triangulation

Dans cette section, on décrit l'utilisation du guidage auditif pour optimiser le geste de triangulation dans la simulation de chirurgie arthroscopique. On décrit l'approche proposée du Son 3D Enrichi pour le guidage

auditif. Ensuite, on présente la méthode utilisée pour appliquer cette approche. Enfin, on décrit les indices sonores SASA utilisés dans cette application.

7.3.1. Le Son 3D Enrichi

Le guidage auditif consiste à fournir à l'utilisateur des informations auditives pour l'assister dans la réalisation d'une tâche en environnement virtuel. Dans cette étude, on utilise l'approche du Son 3D Enrichi (S3E) pour assister l'utilisateur dans la réalisation du geste de triangulation en chirurgie virtuelle.

L'approche du Son 3D Enrichi consiste en la combinaison de la fonction HRTF, l'approche SASA et la technique de CITD. Le processus de traitement d'un stimulus sonore avec la technique S3E est montré dans la Figure 7.5.

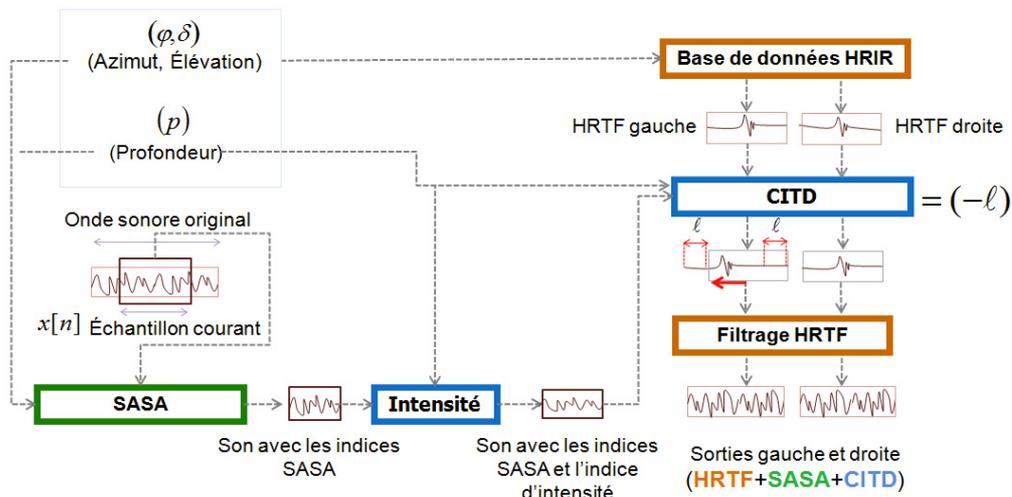


Figure 7.5. Schéma du traitement d'un stimulus sonore avec la combinaison de la HRTF, l'approche SASA et la technique CITD.

La HRTF permet la localisation rapide de source sonore. L'approche SASA améliore la précision de localisation de sources sonores. La technique CITD permet d'améliorer la restitution de la profondeur d'une source sonore. La combinaison de ces trois techniques permet de transmettre à l'utilisateur des indicateurs auditifs (*audio cues*) dans un contexte tridimensionnel (azimut, élévation et profondeur).

Le composant logiciel pour la génération du S3E a été intégré au simulateur grâce à l'architecture modulaire de ce dernier. Ce module est basé sur la plateforme expérimentale décrite dans la section 3.4.

7.3.2. Méthode d'intégration du Son 3D Enrichi en environnement virtuel

L'approche du S3E est conçue pour être utilisée dans différentes applications de manipulation d'objets en environnement virtuel. Dans cette partie, on présente la méthode d'intégration de l'approche du S3E dans une application particulière. Cette méthode comprend une série de tests empiriques pour s'assurer que l'effet obtenu est satisfaisant.

La méthode d'intégration est montrée dans la Figure 7.6. Elle est constituée des étapes suivantes :

- Etape 1. Spécification de la tâche : on doit identifier l'objectif recherché et déterminer le niveau de précision recherché pour la localisation de source sonore,
- Etape 2. Identification de critères d'évaluation de la performance de l'utilisateur dans la tâche étudiée (par exemple : la précision et le temps d'accomplissement de la tâche),
- Etape 3. Analyse de l'outil d'interaction : dans cette étape, on doit identifier les degrés de liberté de l'outil d'interaction et les variables qui représentent les états de ces degrés de liberté. Par exemple, l'allongement est un des degrés de liberté d'un bras mécanique ; la variable qui représente l'état de ce degré de liberté est la mesure de cet allongement,
- Etape 4. Génération du groupe d'indices SASA selon les procédures présentées dans la section 4.1.3. Dans cette étape, on doit prendre en compte l'objectif de la tâche et le niveau de précision requis,
 - Sous-étape 4a. Amélioration des indices SASA,
- Etape 5. Association des propriétés spatiales de la source sonore avec les degrés de liberté de l'outil d'interaction,
- Etape 6. Test empirique qui consiste à évaluer la procédure d'association effectuée dans l'étape 5. Le test est effectué avec un nombre limité d'utilisateurs (trois sujets dans notre étude),
- Etape 7. Analyse de la stimulation auditive obtenue : le groupe d'indices SASA et l'association effectuée dans l'étape 5 sont rejetés si le résultat obtenu n'est pas satisfaisant après six itérations. Dans ce cas, le processus est repris à l'étape 4 pour générer un autre groupe d'indices SASA. Si le concepteur constate une amélioration significative, le processus est repris à la sous-étape 4a. Le processus est achevé si aucune autre amélioration supplémentaire n'est possible.

La procédure présentée permet d'intégrer le guidage auditif basé sur le S3E pour la manipulation d'un objet en environnement virtuel. Cette procédure permet également de confirmer l'utilité du guidage auditif pour l'accomplissement de la tâche étudiée. L'étude de l'effet du guidage auditif sur la performance de l'utilisateur sera effectuée à travers une étude expérimentale.

7.3.3. Application du Son 3D Enrichi pour guider le geste de triangulation dans la simulation chirurgicale

Une tâche souvent utilisée dans l'entraînement au geste de triangulation en simulation chirurgicale consiste à visualiser et à toucher une cible localisée dans une position non connue à l'avance par l'utilisateur. Cette cible est représentée par une sphère.

La tâche décrite permet à l'apprenti chirurgien de développer les capacités psychomotrices suivantes : l'utilisation du retour visuel de la caméra, la navigation à l'intérieur du corps, la manipulation simultanée de la

caméra et de l'instrument, la localisation de la cible visuelle avec la caméra et l'introduction de l'instrument dans le cône de vision de la caméra. Dans cette étude, nous nous sommes intéressés à assister l'utilisateur dans l'étape qui consiste à l'introduction de l'instrument dans le cône de vision de la caméra à l'aide du guidage auditif.

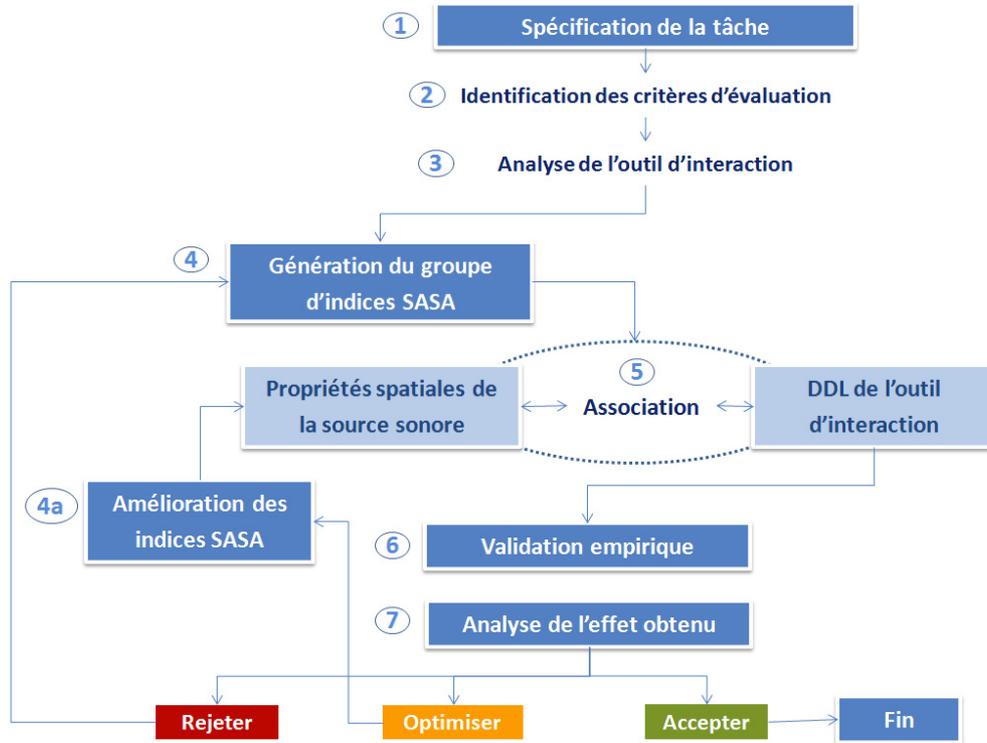


Figure 7.6. Schéma de la procédure d'intégration de l'approche de Son 3D Enrichi pour une tâche d'interaction 3D en environnement virtuel.

7.3.3.1. Approche pour toucher une cible virtuelle

Dans cette section, nous décrivons la procédure suivie pour la tâche qui consiste à toucher une cible virtuelle. Cette procédure est illustrée dans la Figure 7.7. L'utilisateur doit premièrement localiser la cible avec la caméra. Les utilisateurs novices nécessitent beaucoup de temps pour réaliser cette étape car ils ne connaissent pas l'environnement de navigation.

Après avoir localisé la cible par la camera, l'utilisateur introduit l'instrument à l'intérieur du modèle de l'épaule du patient. Ensuite, il procède au positionnement de l'instrument dans le cône de vision de la caméra. En général, pour les utilisateurs novices, cette étape nécessite plusieurs essais. Il est à noter qu'à chaque tentative pour mettre l'instrument dans le cône de vision de la caméra, l'opérateur risque de causer des dégâts dans les organes situés à proximité de la cible. Certains utilisateurs perdent de vue la cible dans cette étape, dans ce cas, on recommande à l'utilisateur de la rechercher.

Après avoir mis l'instrument et la cible dans le cône de vision de la caméra, l'opérateur procède au toucher de la cible. Pour certains utilisateurs novices, il est nécessaire de répéter plusieurs essais à cause du

manque de cohérence entre le retour visuel et les mouvements effectués par l'utilisateur. Dans cette étude, on demande à l'utilisateur de toucher la cible pendant deux secondes.

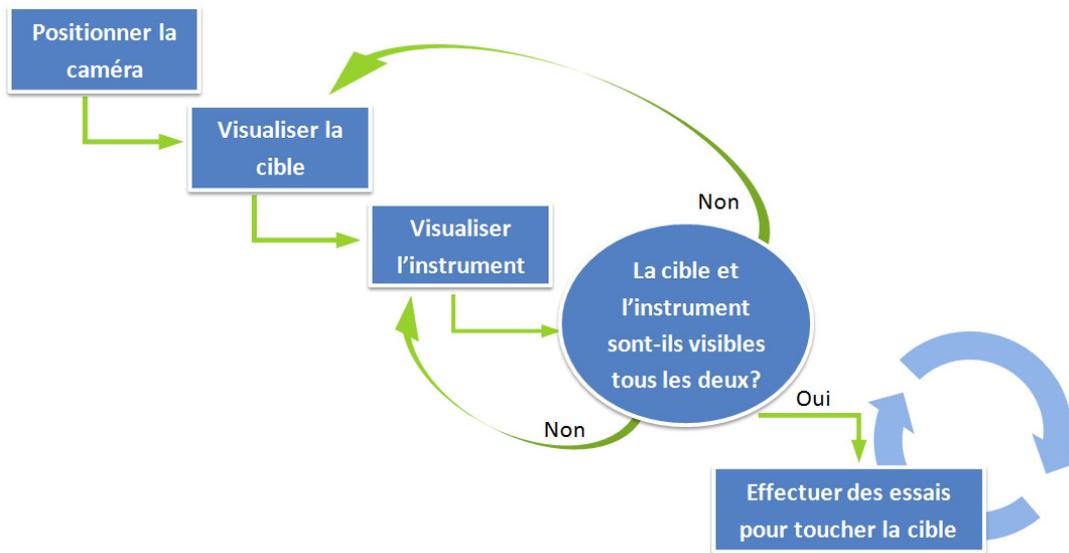


Figure 7.7. Procédure pour toucher la cible virtuelle dans la simulation chirurgicale.

Dans cette étude, on envisage d'utiliser une source sonore pour guider le mouvement de l'utilisateur dans la manipulation d'un instrument. La source sonore est positionnée de manière à être utile pour le guidage auditif du geste de triangulation. La rapidité et la précision de localisation (en azimut et en élévation) de source sonore est utile pour le guidage des mouvements de débattement de l'instrument. L'information de profondeur est utile pour le guidage du mouvement d'introduction (pénétration) de l'instrument dans le patient virtuel.

7.3.3.2. Association de la stimulation auditive avec l'outil d'interaction

Notre objectif est de communiquer des informations auditives pour lui faciliter la tâche qui consiste à positionner l'instrument dans le cône de vision de la caméra en optimisant le mouvement effectué. L'opérateur réalise la manipulation de l'instrument effectuant trois mouvements de base (Figure 7.1) : le débattement latéral, le débattement vertical et l'insertion de l'instrument. Ces mouvements de base correspondent aux degrés de liberté de l'outil d'interaction.

Avant de décrire l'association des propriétés spatiales de la source sonore avec les degrés de liberté de l'outil d'interaction, on présente le concept du point de référence de la caméra. Pour guider le geste, il faut déterminer la position de l'instrument par rapport au point de vue de la caméra. Le point de référence représente l'orientation de ce point de vue. Dans la solution implantée, le point de référence est situé sur l'axe de vision de la caméra dans la direction du cône de vision à une distance de 3.5 pixels de l'objectif (Figure 7.8). La source sonore utilisée pour le guidage auditif est positionnée en tenant en compte de la position de ce point de référence.

Dans les premiers essais réalisés lors du développement de la technique de guidage, nous avons positionné le point de référence au centre de l'objectif de la caméra. Avec cette solution, nous avons constaté que

l'utilisateur était guidé vers l'objectif de la caméra plutôt que vers le cône de vision. Les résultats n'étaient pas satisfaisants car l'utilisateur s'approchait par derrière ou du côté de l'objectif de la caméra sans réussir à mettre l'instrument dans le cône de vision. De ce fait, nous avons positionné le point de référence dans le cône de vision. Cette solution a permis à l'utilisateur qui réussit à mettre plus facilement l'instrument dans le cône de vision de la caméra.

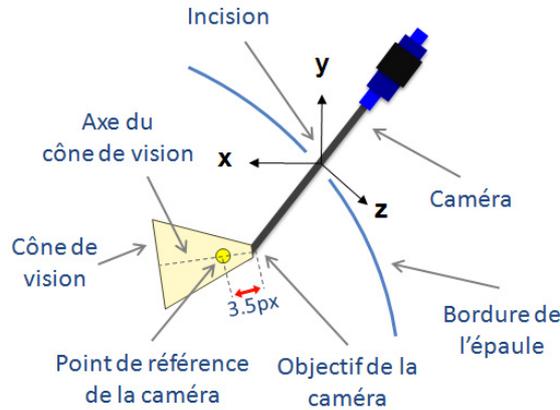


Figure 7.8. Position du point de référence de la caméra.

Il est à noter que l'utilisateur est assisté par la technique de guidage pour mettre l'instrument dans le cône de vision de la caméra. Ce processus est effectué indépendamment de la phase de la localisation de la cible virtuelle. Il est à noter également que la technique de guidage auditif ne connaît pas la position de la cible virtuelle générée de façon aléatoire. Par contre, la position et l'orientation de l'instrument et celles de la caméra doivent être connues par la technique de guidage.

Pour effectuer l'association des propriétés spatiales avec les degrés de liberté de l'outil d'interaction, on doit calculer les mouvements de débattement et de profondeur nécessaires pour amener la pointe de l'instrument au point de référence de la caméra.

Pour le calcul des débattements, on définit deux lignes droites de référence. La première ligne droite est comprise entre les coordonnées de l'incision P_1 et le point de référence de la caméra P_3 (Figure 7.9a). La deuxième ligne droite est comprise entre l'incision P_1 et la pointe de l'instrument P_2 . Le débattement latéral est l'angle α formé par les projections des lignes de référence sur le plan horizontal (Figure 7.9a). Le débattement vertical est la différence angulaire β comprise entre l'axe y et les lignes de référence illustrées dans la Figure 7.9b.

La Figure 7.9d montre l'association réalisée des propriétés spatiales avec les degrés de liberté de l'outil d'interaction. L'azimut φ de la source sonore est associé au débattement latéral α nécessaire pour amener la pointe de l'instrument au point de référence de la caméra. L'élévation δ est associée au débattement vertical β . La profondeur p est associée à la distance w comprise entre la pointe de l'instrument et le point de référence de la caméra.

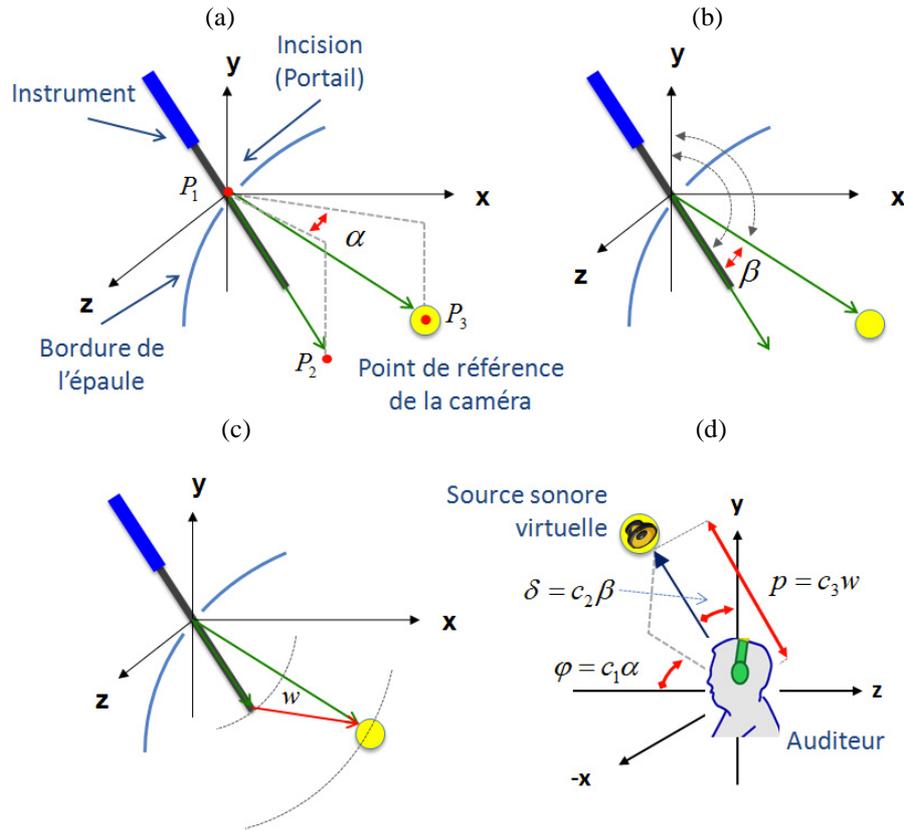


Figure 7.9. Association des propriétés spatiales de la source sonore avec les degrés de liberté de l'outil d'interaction.

L'azimut φ et l'élévation δ de la source sonore sont calculés come suit :

$$\varphi = c_1 \alpha \quad (7.1)$$

$$\delta = c_2 \beta \quad (7.2)$$

où α est le débattement latéral,

β est le débattement vertical,

c_1 et c_2 sont les coefficients de gain pour l'azimut et l'élévation respectivement.

Les coefficients de gain permettent d'amplifier ou d'atténuer le rapport entre les débattements et les angles qui déterminent la position de la source sonore. Dans cette étude, les coefficients de gain sont fixés de la manière suivante : $c_1 = c_2 = 2$.

L'amplification définie par les coefficients de gain angulaire (c_1 et c_2) est importante car l'amplitude du débattement angulaire de l'instrument est petite (quelques degrés) et elle est limitée par l'ouverture du cône de pénétration. Cette amplification permet de faciliter l'identification du débattement angulaire à effectuer et d'utiliser plus efficacement l'intervalle des valeurs des propriétés spatiales de la source sonore.

La profondeur p de la source sonore est représentée par :

$$p = c_3 w \quad (7.3)$$

où w est la distance (en pixels) entre la pointe de l'instrument et le point de référence de la caméra, et c_3 est le coefficient de gain de la profondeur de la source sonore.

Le coefficient de gain c_3 est calculé comme suit :

$$c_3 = \frac{d}{l} \quad (7.4)$$

où l (en pixels) est la distance maximale possible entre la pointe de l'instrument et le point de référence de la caméra à l'intérieur du modèle de l'épaule, et

d (en mètres) est le paramètre de la technique CITD qui représente la distance à laquelle de la technique n'a aucun effet.

Le coefficient de gain c_3 définit l'association de l'intervalle du paramètre d de la technique CITD avec la profondeur de la source sonore.

7.3.3.3. Approche du guidage auditif par le S3E pour toucher une cible virtuelle

Dans cette partie, on décrit la procédure suivie par l'utilisateur pour toucher une cible avec l'aide du guidage auditif apporté par l'approche S3E. La procédure est montrée dans la Figure 7.10. Il est à noter que cette procédure présente des différences importantes avec celle présentée dans la Figure 7.7.

Au début de la procédure, l'utilisateur localise la cible avec la caméra. Il est recommandé à l'utilisateur de positionner la cible au centre du cône de vision de la caméra. Ensuite, l'utilisateur introduit l'instrument à travers l'incision d'une distance de 3 centimètres. Il utilise l'information auditive générée avec l'approche S3E pour orienter l'instrument.

La manipulation de l'instrument est effectuée en trois étapes : le débattement latéral, le débattement vertical et l'introduction dans le modèle de l'épaule.

Le débattement latéral est effectué en suivant l'azimut de la source sonore diffusée jusqu'à ce que cette dernière se trouve dans le plan horizontal en face de l'utilisateur. Les indices SASA facilitent cette étape de la procédure. De ce fait, le débattement latéral revient à chercher la source sonore dans le plan horizontal.

Le débattement vertical est effectué en suivant l'élévation de la source sonore jusqu'à ce que cette dernière se trouve dans le plan horizontal (élévation égale à zéro). Les indices SASA facilitent cette étape de la procédure. Après la mise de l'instrument dans l'orientation recherchée, l'utilisateur introduit l'instrument dans l'épaule du patient virtuel sans modifier son orientation.

L'introduction de l'instrument dans le modèle de l'épaule du patient est effectuée en suivant la profondeur de la source sonore afin de connaître la distance à parcourir pour arriver au point de référence. L'instrument apparaît dans le cône de vision de la caméra si le guidage auditif est utilisé efficacement. L'orientation de l'instrument est incorrecte si la profondeur de la source sonore est perçue très proche par l'utilisateur mais l'instrument n'apparaît pas encore dans le cône de vision. Dans ce cas, il est recommandé à l'utilisateur de retirer l'instrument et de recommencer la procédure.

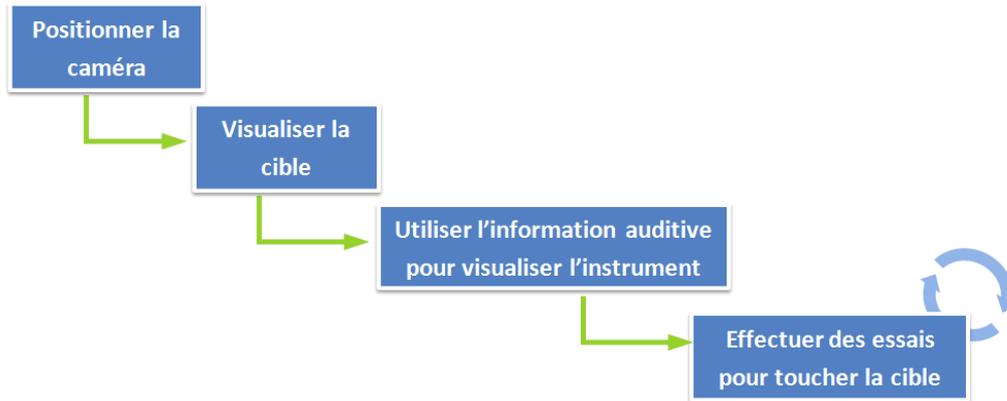


Figure 7.10. Procédure pour toucher une cible avec le guidage auditif fourni apporté par l'approche S3E.

Le guidage auditif est conçu pour limiter le mouvement de débattement réalisé avec des profondeurs importantes. Ainsi, l'instrument est orienté avant la phase d'introduction proprement dite. Lorsque l'instrument est introduit dans le modèle de l'épaule, le mouvement de débattement génère le déplacement d'une importante quantité de matière, ce qui risque de causer des dégâts secondaires à l'intervention chirurgicale.

Le guidage auditif permet de réduire le nombre de tentatives pour toucher la cible virtuelle. Dans la nouvelle procédure, l'instrument est dirigé vers la région où se trouve la cible car cette dernière est dans le centre du cône de vision de la caméra.

7.3.4. Groupe d'indices SASA utilisé

Dans cette partie, on décrit le groupe d'indices SASA utilisé dans cette étude. Nous avons utilisé le stimulus sonore présenté dans les chapitres précédents. Les propriétés spatiales du groupe d'indices SASA utilisés sont montrés dans le Tableau 7.1.

Tableau 7.1. Propriétés spatiales du groupe des indices SASA utilisé

Nom de la propriété	Description	Intervalle de la propriété
Élévation	Différence angulaire verticale	$[-90,90]^\circ$
Azimut	Différence angulaire horizontale	$[-180,180]^\circ$
Angle global	Différence angulaire globale	$[0,180]^\circ$
Profondeur	Distance entre l'auditeur et la source sonore	$[0,20]m$

D'après la méthode d'intégration de l'approche S3E décrite précédemment, il faut identifier le niveau de précision de localisation sonore nécessaire à l'accomplissement de la tâche étudiée. L'objectif principal du guidage auditif est de faciliter le positionnement de l'instrument dans le cône de vision de la caméra. Dans un premier temps, le guidage sert à orienter correctement l'instrument avant de l'introduire en profondeur. Pour

orienter correctement l'instrument, il est nécessaire que la précision de localisation de la source sonore utilisée pour guider l'utilisateur soit la meilleure possible.

Les indices SASA utilisés sont décrits dans le Tableau 7.2. Nous détaillons l'effet audio choisi et la propriété spatiale prédominante pour chaque indice.

Tableau 7.2. Indices sensoriels du groupe SASA pour la geste de triangulation

Indices	Effet audio	Paramètre de contrôle de l'effet audio	Intervalle du paramètre de l'effet audio	Propriété prédominante de l'indice
Nouvelle verticalité	Réverbération	Niveau	$[-10k, 2k]mdB$	Élévation
Horizontalité	Atténuation	Facteur	$[-40, 0]dB$	Azimut
Proximité angulaire	Filtre passe-bas	Fréquence de coupure	$[200, 22k]Hz$	Angle global

Les indices d'horizontalité et proximité angulaire utilisés et leur fonction d'association sont décrits dans le chapitre 4. Dans cette étude, il n'est pas nécessaire de résoudre le problème d'ambiguïté devant-derrrière de la localisation de sources sonores car l'ouverture maximale du cône de pénétration (45°) et les coefficients de gain angulaire pour le positionnement de la source sonore ($2x$) ne risquent pas de présenter la situation où la source sonore est perçue derrière l'auditeur. Pour cette raison, l'indice de frontalité, présenté dans le chapitre 4, n'est pas utilisé dans cette étude.

L'indice de verticalité présenté dans le chapitre 4 a été modifié dans cette étude. En principe, l'indice de verticalité se réfère à la distinction entre les sources sonores qui se trouvent au dessous et au dessus du plan horizontal. En revanche, le nouvel indice de verticalité permet de distinguer les sources sonores dont l'élévation est proche du plan horizontal (élévation entre -10 et 10 degrés) des sources sonores localisées plus haut ou plus bas. La forme de la fonction d'association du nouvel indice de Verticalité est montrée dans la Figure 7.11.

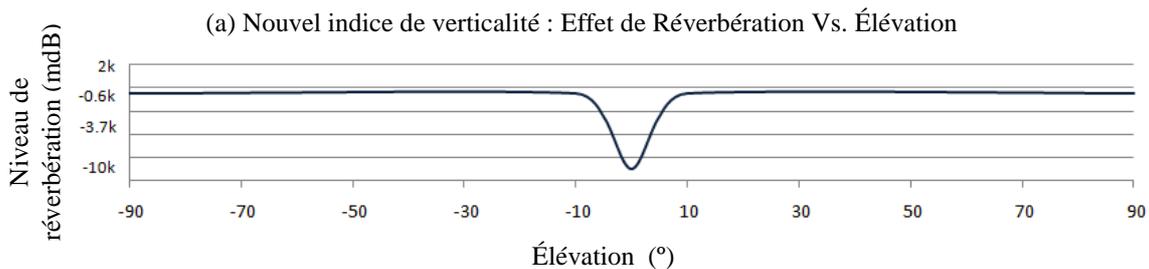


Figure 7.11. Fonction d'association du nouvel indice SASA de verticalité. L'axe vertical est tracé en échelle logarithmique.

Les indices SASA assistent l'utilisateur à orienter correctement l'instrument. L'utilisateur oriente l'instrument premièrement dans le plan horizontal (débattement latéral) et ensuite dans le plan vertical (débattement vertical). L'utilisateur arrête le mouvement de débattement latéral lorsque la source sonore se trouve centrée en face de lui (azimut égal à zéro). Il arrête le mouvement de débattement vertical lorsque la

source sonore se trouve à la hauteur du plan horizontal (élévation égale à zéro) et lorsque la réverbération est nulle. Après avoir orienté correctement l'instrument, l'utilisateur l'introduit dans le modèle de l'épaule sans modifier son orientation.

7.4. Étude expérimentale

L'objectif de cette étude est de déterminer l'effet du guidage auditif sur la performance de l'utilisateur dans la tâche qui consiste à toucher une cible dans la simulation de la chirurgie arthroscopique. La performance est mesurée en termes de temps d'accomplissement de la tâche et de l'économie de mouvement.

Les expérimentations décrites dans ce chapitre ont été réalisées avec les équipements du Groupe de Modélisation et Réalité Virtuelle (GMRV) de l'Université Rey Juan Carlos à Madrid en Espagne dans le cadre de collaboration scientifique avec l'Institut Image.

7.4.1. Evaluation de la performance

7.4.1.1. Conditions expérimentales

Les expérimentations ont été effectuées avec deux modalités d'opération et avec deux modalités de stimulation auditive. Les modalités d'opération concernent l'utilisation du simulateur avec une ou deux mains. Les modalités de stimulation auditive concernent l'utilisation d'un retour sonore basé sur l'approche du S3E ou l'absence de restitution auditive. Les quatre conditions expérimentales sont montrées dans le Tableau 7.3 et dans la Figure 7.12.

Tableau 7.3. Conditions expérimentales

Code	Modalité d'opération du simulateur	Modalité de stimulation auditive
1. UM+SS	- Une main (UM)	- Sans son (SS)
2. UM+S3E	- Une main (UM)	- Approche S3E (S3E)
3. DM+SS	- Deux mains (DM)	- Sans son (SS)
4. DM+S3E	- Deux mains (DM)	- Approche S3E (S3E)

Les expérimentations ont été réalisées en deux étapes : une expérience pour chaque modalité d'opération du simulateur. Chaque utilisateur effectue l'expérience avec une seule main en manipulant l'instrument uniquement. Avant la réalisation de l'expérience, la caméra est positionnée de manière que la cible est visible et centrée dans le cône de vision.

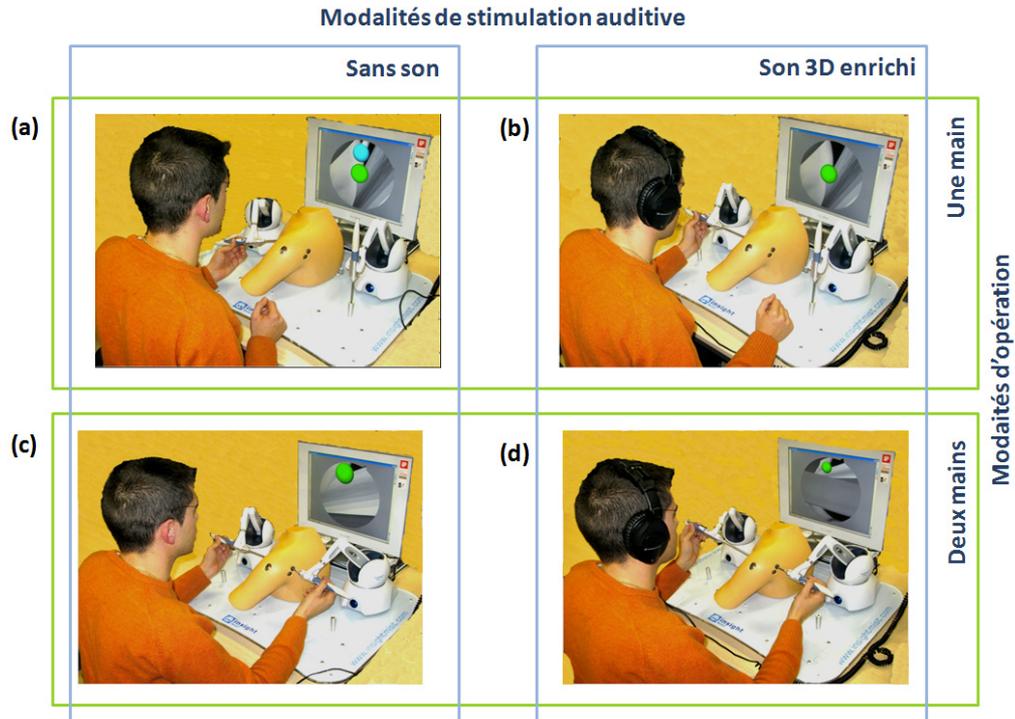


Figure 7.12. Les conditions expérimentales développées.

Ensuite, l'utilisateur exécute la manipulation avec deux mains. Dans cette modalité, l'utilisateur manipule la caméra avec la main droite pour visualiser la cible et il manipule l'instrument avec la main gauche pour le mettre dans le cône de vision de la caméra et toucher la cible.

D'après les pré-tests que nous avons effectués, nous avons constaté que la stratégie adoptée par l'utilisateur pour positionner la caméra a un effet important sur sa performance. Pour cela, nous avons décidé de réaliser l'expérience avec deux conditions opératoires : (1) l'utilisateur manipule l'instrument avec une seule main en ayant la caméra positionnée à l'avance et (2) l'utilisateur effectue l'opération avec deux mains (il manipule la caméra avec sa main droite et l'instrument avec sa main gauche). On étudie l'effet du retour auditif dans ces deux modalités d'opération.

7.4.1.2. Description de la tâche expérimentale

La tâche consiste à visualiser et à toucher une cible représentée par une sphère. Cette tâche s'effectue en différentes phases. Si l'opération est réalisée avec deux mains, l'utilisateur commence par introduire la caméra pour visualiser la cible. Ensuite, il introduit l'instrument pour le mettre dans le cône de vision de la caméra. Si le retour sonore est activé, l'utilisateur se sert des informations sonores pour orienter et introduire en profondeur l'instrument. Après avoir mis l'instrument dans le cône de vision de la caméra il doit toucher la cible pendant deux secondes. Au contact avec l'instrument, la cible change de couleur. Après deux secondes la sphère disparaît pour indiquer à l'utilisateur que la tâche est achevée.

L'approche de Son 3D Enrichi est expliquée à l'utilisateur avant de commencer l'expérience. Il réalise deux essais préliminaires pour chaque modalité de stimulation qui ne seront pas prises en compte pour l'analyse

des résultats. On recommande à l'utilisateur de retirer l'instrument et de recommencer la tâche en cas de situation de confusion lors de la manipulation de l'instrument à l'intérieur du modèle de l'épaule.

7.4.1.3. Essais expérimentaux

L'utilisateur réalise cinq essais de la tâche dans chaque condition expérimentale. Chaque essai est associé à une cible positionnée dans un emplacement différent. Les différentes positions des cibles sont montrées dans la Figure 7.13.

Il est à noter que l'accès aux cibles 1 et 2 est difficile. L'accès aux cibles 4 et 5 est relativement difficile et l'accès à la cible 3 est considéré facile. Les différentes localisations de la cible sont présentées dans un ordre aléatoire.

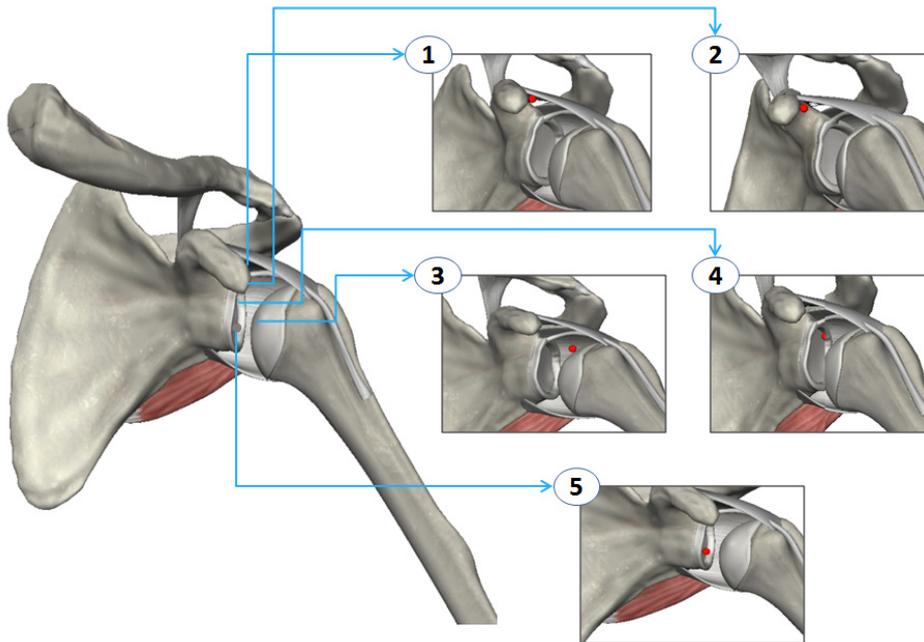


Figure 7.13. Différentes localisations de la cible.

7.4.1.4. Groupe de participants

Le groupe de sujets ayant participé à l'expérience avec une seule main est constitué par 16 personnes (13 hommes et 3 femmes). Le groupe de sujets qui ont réalisé l'expérience avec deux mains est également composé de 16 personnes (13 hommes et 3 femmes). Dix sujets du premier groupe ont également participé à la deuxième expérience.

Les sujets sont étudiants droitiers et ils ont déclaré qu'ils ont une vision et une audition normale. Chaque session dure 20 minutes environ avec une pause de deux minutes entre chaque les deux sessions. Tous les sujets ayant participé à cette expérience ont des connaissances en réalité virtuelle. La moitié de l'ensemble de participants ont des connaissances en simulation de chirurgie arthroscopique.

7.4.1.5. Évaluation objective

Les variables objectives prises en compte dans cette étude sont le temps d'accomplissement de la tâche, la longueur de la trajectoire réalisée par l'instrument et la distance moyenne de l'instrument par rapport à la trajectoire optimale. Les deux dernières mesures sont des indicateurs de l'économie de mouvement.

Le temps d'accomplissement de la tâche (en secondes) est la durée mesurée entre l'instant où l'instrument est inséré à travers l'incision et celui où l'instrument touche la cible. Cette durée correspond au temps au cours duquel le retour sonore est exploité par l'utilisateur. On présente ci-après les autres variables objectives prises en compte dans cette étude.

La trajectoire T parcourue par l'instrument à l'intérieur du modèle de l'épaule est représentée comme suit :

$$T = \langle (x_0, y_0, z_0), (x_1, y_1, z_1), \dots, (x_{n-1}, y_{n-1}, z_{n-1}) \rangle \quad (7.5)$$

où n est la longueur de T .

La longueur de la trajectoire effectuée $\|T\|$ est définie par :

$$\|T\| = \sum_{i=1}^{n-1} \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2 + (z_i - z_{i-1})^2} \quad (7.6)$$

où $\|T\|$ est mesurée en pixels.

La trajectoire optimale de l'instrument est définie par la ligne droite comprise entre l'incision (P_1) et le point de référence de la caméra (P_3). Pour mesurer la distance moyenne par rapport à cette trajectoire, nous avons défini un indicateur de distance moyenne basé sur le concept d'erreur instantanée.

L'erreur instantanée e_i (en pixels) est la distance minimale entre la pointe de l'instrument (P_2) et la ligne droite qui correspond à la trajectoire optimale à l'instant i . Elle est définie comme suit :

$$e_i = \frac{\left\| \overrightarrow{P_1 P_2} \right\| \left\| \overrightarrow{P_1 P_2} \cdot \overrightarrow{P_1 P_3} \right\|}{\left\| \overrightarrow{P_1 P_3} \right\|^2} \quad (7.7)$$

où P_1 est la position de la pointe de l'instrument à l'instant i ,

$\overrightarrow{P_1 P_2}$ est le vecteur formé entre les points P_1 et P_2 ,

$\overrightarrow{P_1 P_3}$ est le vecteur formé entre les points P_1 et P_3 ,

$\| \cdot \|$ représente l'amplitude d'un vecteur,

\cdot représente le produit scalaire point entre deux vecteurs.

La Figure 7.14 montre une représentation graphique de l'erreur instantanée.

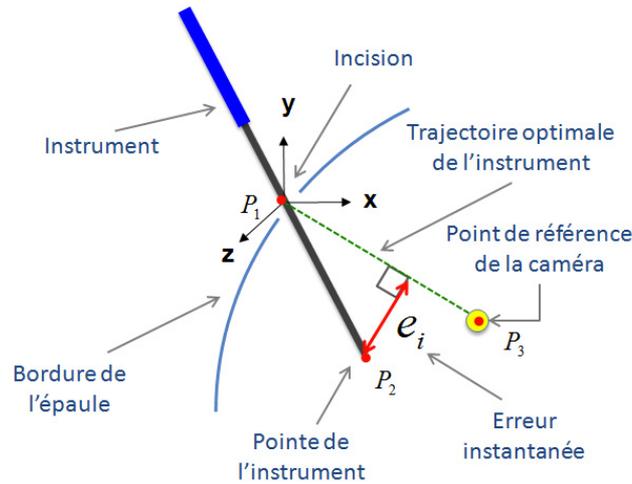


Figure 7.14. Erreur instantanée de la position de l'instrument par rapport la trajectoire optimale.

L'indicateur de la distance moyenne ID est défini comme suit :

$$ID = \frac{1}{n} \sum_{i=0}^{n-1} e_i \quad (7.8)$$

où e_i (en pixels) est l'erreur instantanée à l'instant i ,

n est la longueur de la trajectoire de l'instrument qui correspond au nombre d'erreurs instantanées mesurées et

ID est mesuré en pixels.

7.4.2. Retour d'informations de l'utilisateur

Pour recueillir le retour d'information des utilisateurs, on demande à chaque sujet de répondre à un questionnaire après la réalisation de l'expérience. Les questions posées portent sur leur appréciation sur la facilité d'utilisation de la technique d'interaction et l'intuitivité du retour sonore dans les conditions expérimentales étudiées.

Les questions ont été élaborées pour permettre au sujet de comparer les modalités de stimulation auditive pour chaque modalité d'opération. Le type de questions et de réponses possibles sont présentés ci-dessous.

Question 1 : Parmi les modalités de stimulation SS (sans son) et S3E (Son 3D Enrichi), quelle est la plus facile à utiliser?

- 3 SS est beaucoup plus facile à utiliser
- 2 SS est plus facile à utiliser
- 1 SS est un peu plus facile à utiliser
- 0 S3E est aussi facile à utiliser que SS
- 1 S3E est un peu plus facile à utiliser

2 S3E est plus facile à utiliser

3 S3E est beaucoup plus facile à utiliser

Question 2 : Parmi les modalités de stimulation SS (sans son) et S3E (Son 3D Enrichi), quelle est la plus intuitive?

-3 SS est beaucoup plus intuitive

-2 SS est plus intuitive

-1 SS est un peu plus intuitive

0 S3E est aussi intuitive que SS

1 S3E est un peu plus intuitive

2 S3E est plus intuitive

3 S3E est beaucoup plus intuitive

Pour définir les différentes réponses possibles, nous avons utilisé le modèle de la Catégorie de Notation en Dégradation (DCR : *Degradation Category Rating*) recommandée par l'ITU-T [1996]. Au total, chaque sujet répond à deux questions pour chaque modalité opératoire (c.f. Annexe : 10.3).

7.5. Résultats et analyse de données

7.5.1. Résultats de l'évaluation de performance

Dans cette partie, on présente l'analyse des données obtenues des expériences réalisées et on montre l'effet des facteurs expérimentaux sur la performance de l'utilisateur.

7.5.1.1. Conditions expérimentales évaluées

Le Tableau 7.4 montre les données obtenues dans les différentes conditions expérimentales.

La Figure 7.15 montre la moyenne de la variable de temps d'accomplissement de la tâche pour chaque condition expérimentale. Il est à noter que lorsque le simulateur est utilisé avec une seule main le retour sonore permet de réduire la dispersion de données. En revanche, lorsque le simulateur est utilisé avec deux mains, le retour sonore n'a pas d'effet sur la dispersion des données. La signification statistique de cette différence sera déterminée ultérieurement.

La Figure 7.16 montre la moyenne de la longueur de trajectoire réalisée dans les conditions expérimentales évaluées. On constate que le retour sonore permet de réduire la longueur de la trajectoire pour la tâche réalisée avec une main et deux mains. L'amélioration de la longueur moyenne de la trajectoire suivie par de l'instrument dans la tâche réalisée avec une main et deux mains est de 56% et 49% respectivement. Ceci implique que le retour sonore permet de réduire le mouvement nécessaire pour réaliser la tâche.

Tableau 7.4. Mesures statistiques des données obtenues

Condition	Variable	N ^{bre}	Moyenne	Médiane	Quartile			Écart type
					Inf.	Sup.	Inter.	
Toutes les conditions	Temps d'accomplissement (s)	320	19.48	14.26	5.62	23.59	17.97	23.36
	Longueur de trajectoire (px)	320	314.9	185.4	144.2	349.0	204.8	332.7
	Indicateur de distance moyenne (px)	320	6.31	5.34	3.12	9.09	5.96	3.94
1. UM+SS	Temps d'accomplissement (s)	80	19.39	18.56	1.39	28.00	26.61	26.98
	Longueur de trajectoire (px)	80	429.5	266.9	128.6	567.9	439.3	482.8
	Indicateur de distance moyenne (px)	80	9.10	9.02	5.24	13.12	7.88	4.31
2. UM+S3E	Temps d'accomplissement (s)	80	18.58	16.83	12.46	22.56	10.10	9.90
	Longueur de trajectoire (px)	80	191.2	163.9	146.0	203.7	57.7	82.7
	Indicateur de distance moyenne (px)	80	3.64	3.01	2.00	4.72	2.72	2.39
3. DM+SS	Temps d'accomplissement (s)	80	16.47	9.60	2.38	22.81	20.43	19.87
	Longueur de trajectoire (px)	80	395.0	248.1	145.7	495.2	349.4	364.6
	Indicateur de distance moyenne (px)	80	8.13	8.14	5.60	10.49	4.88	3.31
4. DM+S3E	Temps d'accomplissement (s)	80	23.47	14.65	9.72	23.51	13.79	30.92
	Longueur de trajectoire (px)	80	244.1	181.8	150.1	253.1	102.9	183.6
	Indicateur de distance moyenne (px)	80	4.37	3.73	2.61	5.74	3.13	2.27

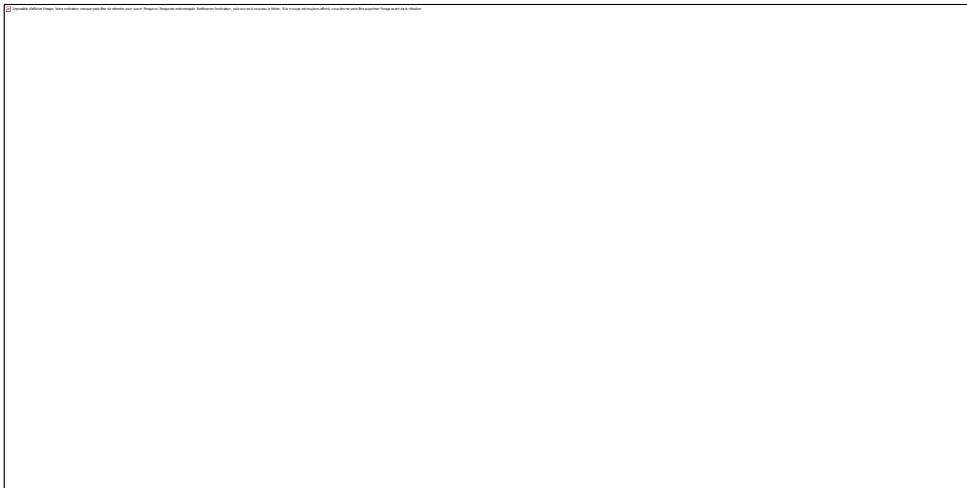


Figure 7.15. Temps d'accomplissement de la tâche pour les conditions expérimentales évaluées.

La Figure 7.17 montre la moyenne de l'indicateur de la distance moyenne pour les conditions expérimentales étudiées. On constate que le retour sonore permet de réduire l'indicateur de la distance moyenne pour la tâche réalisée avec une main et celle réalisée avec deux mains. L'amélioration de l'indicateur de la distance moyenne obtenue pour la tâche réalisée avec une main et avec deux mains est de 62% et de 46%

respectivement. Ceci permet de conclure que le mouvement de l'instrument est proche de la trajectoire optimale lorsque le retour sonore est utilisé dans l'expérience.

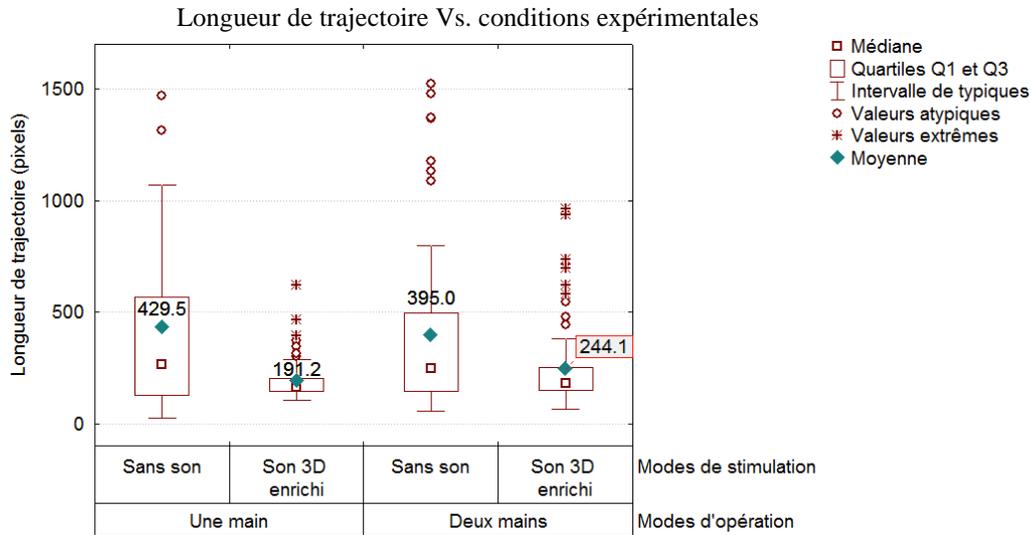


Figure 7.16. Longueur de la trajectoire suivie par l'instrument pour les conditions expérimentales évaluées.

La Figure 7.18a montre la performance des sujets en termes de temps d'accomplissement de la tâche pour les conditions expérimentales évaluées. On constate qu'il n'existe pas de différence significative dans la performance des sujets pour les conditions expérimentales évaluées.

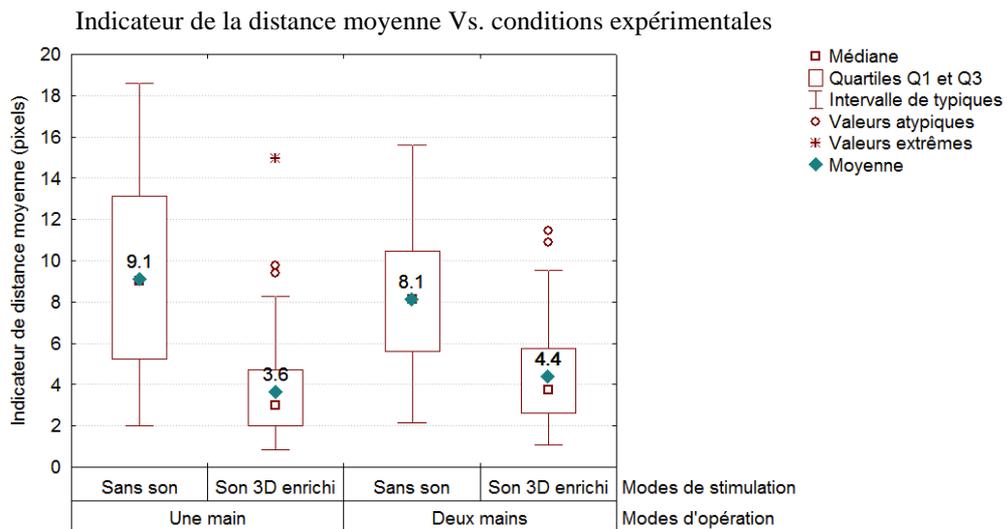
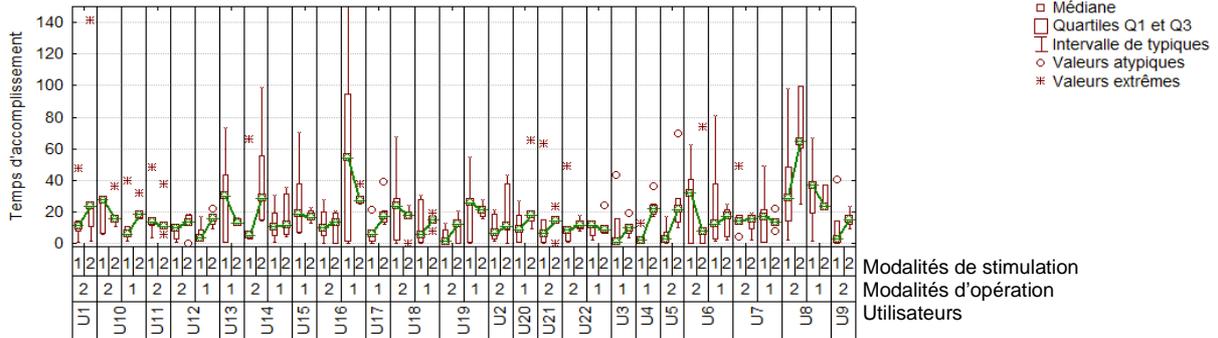


Figure 7.17. Indicateur de la distance moyenne pour les conditions expérimentales évaluées.

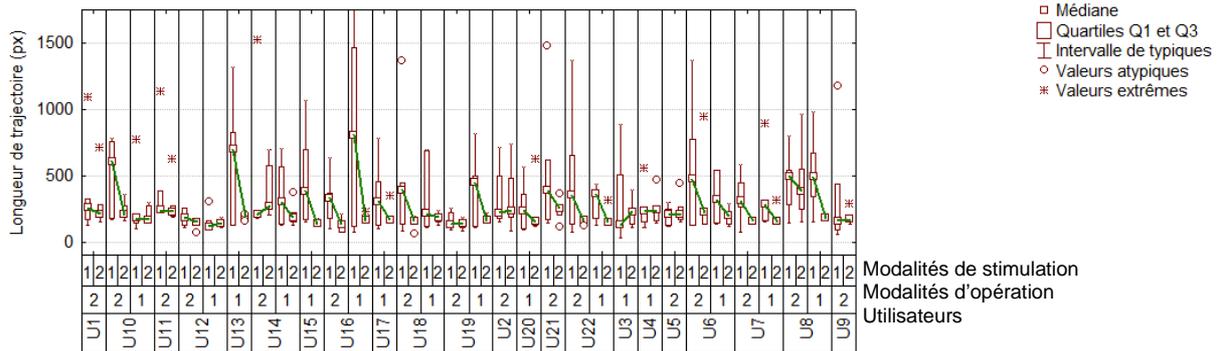
La Figure 7.18b montre la performance des sujets en termes de longueur de trajectoire pour les conditions expérimentales évaluées. Il est à noter que la longueur de la trajectoire est réduite dans douze cas sur seize lorsque le retour sonore est utilisé.

La Figure 7.18c montre la performance des sujets en termes d'indicateur de la distance moyenne pour les conditions expérimentales évaluées. L'indicateur de la distance moyenne est réduit dans quatorze cas sur seize lorsque le retour sonore est utilisé.

(a) Performance individuelle en termes de temps d'accomplissement de la tâche



(b) Performance individuelle en termes de longueur de la trajectoire



(c) Performance individuelle en termes d'indicateur de la distance moyenne

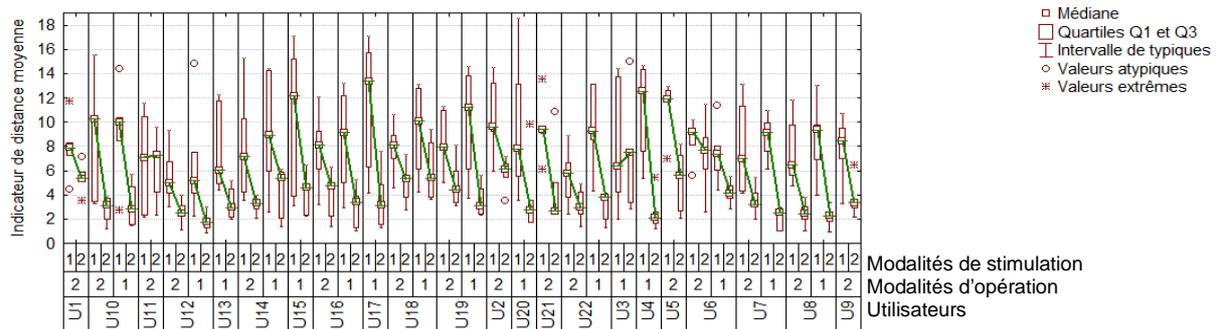


Figure 7.18. Performance individuelle pour (a) le temps d'accomplissement de la tâche, pour (b) la longueur de trajectoire et pour (c) l'indicateur de la distance moyenne. Les modalités d'opération du simulateur UM et DM sont représentées par 1 et 2 respectivement. Les modalités de stimulation auditive SS et S3E sont représentées par 1 et 2 respectivement.

7.5.1.2. Distribution de la fréquence des données

Le Tableau 7.5 montre les tests de qualité d'ajustement des variables expérimentales. Ces tests ont été effectués pour déterminer la distribution de la fréquence de ces variables. Toutes les mesures valides ont été prises en compte dans l'analyse.

Tableau 7.5. Tests de qualité d'ajustement

Fonction	Variable	Moyenne	Variance	Moyenne ajustée	Variance ajustée	K-S distance	p	Ajustée (o/n) ?
Normale	Temps d'accomplissement (s)	19.48	23.26	19.48	23.26	0.203	<0.01	N
	Longueur de trajectoire	314.9	1.1 ^{E5}	314.9	1.1 ^{E5}	0.192	<0.01	N
	Indicateur de distance moyenne	6.31	3.94	6.31	3.94	0.108	<0.01	N
Log-normale	Temps d'accomplissement (s)	19.48	23.26	19.47	545.49	0.071	0.05 <p< 0.1	O
	Indicateur de distance moyenne	6.31	3.94	1.72	15.51	0.053	>0.05	O

Fonction	Variable	Moyenne	Variance	Lambda	K-S distance	p	Ajustée (o/n) ?
Exponentielle	Longueur de trajectoire	314.9	1.1 ^{E5}	.003179	0.074	0.05 <p< 0.1	O

La méthode de comparaison utilisée est celle du test Kolmogorov-Smirnov (K-S) avec un niveau de confiance de 95%.

Le test K-S montre que le temps d'accomplissement de la tâche et la longueur de la trajectoire de mouvement suivent une distribution de fréquence de type log-normale. La Figure 7.19a et la Figure 7.19b montrent la distribution de fréquence du temps d'accomplissement de la tâche et de la longueur de la trajectoire de mouvement respectivement.

Le test K-S montre également que l'indicateur de la distance moyenne suit une distribution de fréquence de type exponentielle. La Figure 7.19c montre que la distribution de fréquence de la longueur de la trajectoire de mouvement ajustée à une fonction de type exponentielle.

7.5.1.3. Analyse statistique

La méthode d'analyse statistique utilisée est celle des Equations d'Estimation Généralisées. Le Tableau 7.6 montre le résultat de l'analyse statistique effectuée pour le facteur expérimental de modalités d'opération dans (*within*) les modalités de stimulation. Ces résultats montrent que ce facteur n'a pas d'effet significatif

($p=0.255$) sur le temps d'accomplissement de la tâche. En revanche, ce facteur a un effet significatif sur la longueur de la trajectoire ($p=0.000$) et sur l'indicateur de distance moyenne ($p=0.000$).

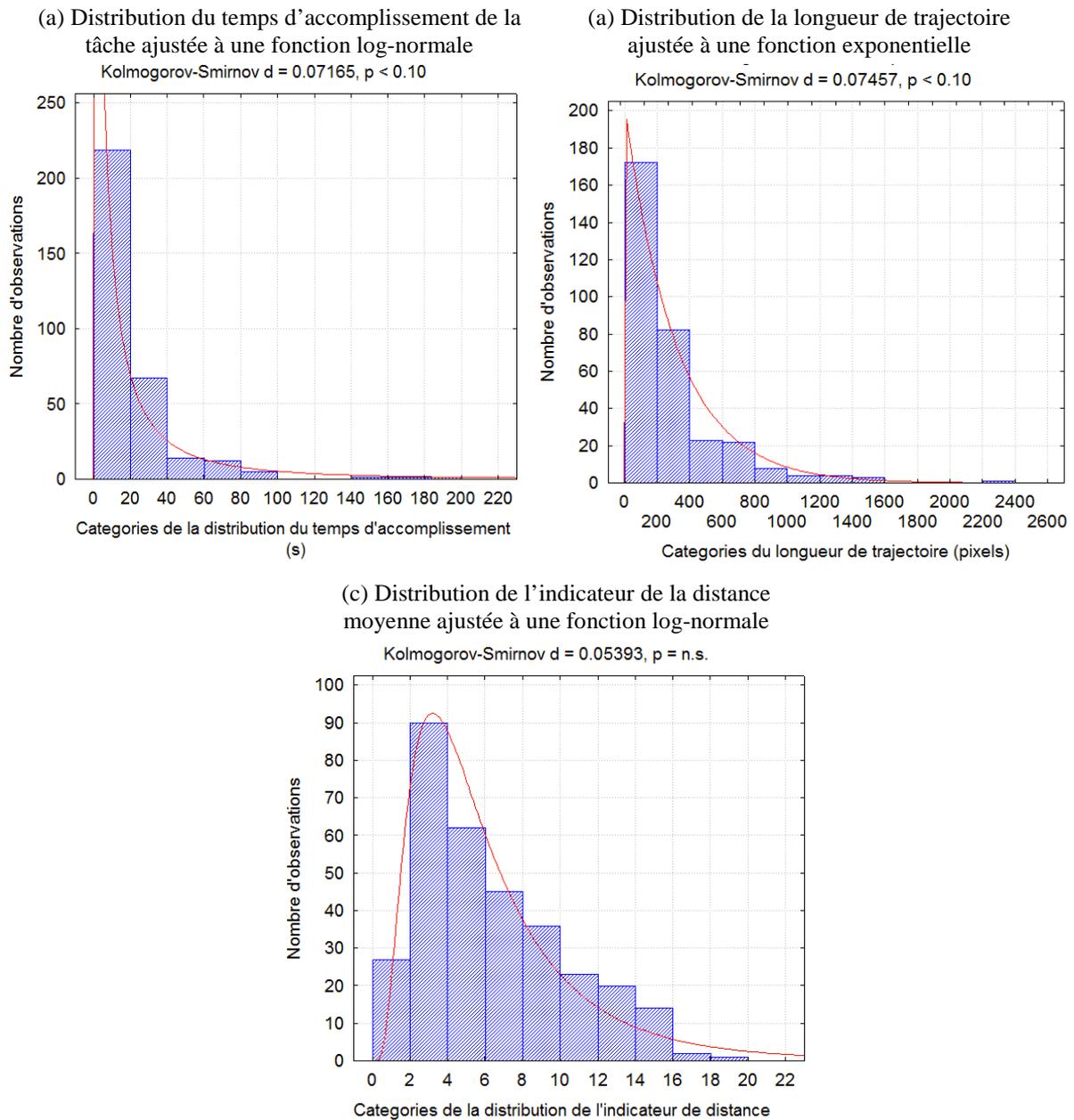


Figure 7.19. Distribution de fréquence pour : (a) le temps d'accomplissement de la tâche, (b) la longueur de trajectoire et (c) l'indicateur de la distance moyenne.

Le Tableau 7.7 montre le résultat de l'analyse statistique de la comparaison des modalités de stimulation. La méthode d'analyse utilisée est celle de l'Estimation des Moyennes Marginales.

D'après le résultat obtenu, le retour sonore n'a pas d'effet sur le temps d'accomplissement de la tâche dans les deux modalités d'opération ($p=0.801$ pour la tâche réalisée avec une seule main et $p=0.705$ pour celle réalisée avec deux mains). L'effet du retour sonore sur la longueur de la trajectoire est statistiquement significatif pour l'opération avec une main ($p=0.000$) et pour celle avec deux mains ($p=0.000$). L'effet du retour sonore sur l'indicateur de distance moyenne est statistiquement significatif pour l'opération réalisée avec une main ($p=0.000$) et pour celle réalisée avec deux mains ($p=0.000$).

Tableau 7.6. Analyse de l'effet des facteurs expérimentaux par l'Analyse de Mesures Répétées avec le modèle d'Equations Généralisées d'Estimation

Facteur expérimental	Variable	Modèle	Wald Chi-square (χ^2)	DDL	p	Significatif (o/n) ?
Modalités d'opération dans les modalités de stimulation	Temps d'accomplissement (s)	Log-normal	5.234	3	.255	N
	Longueur de trajectoire (px)	Exponentiel	97.419	3	.000	O
	Indicateur de distance moyenne	Log-normal	125.126	3	.000	O

L'estimation de l'effet est basée sur la mesure statistique Wald chi-square. Le test est de type III pour lequel l'ordre d'agrégation des facteurs n'est pas important. Le niveau de confiance est de 95%.

On constate que le retour sonore n'a pas d'effet sur le temps d'accomplissement de la tâche mais qu'il permet d'améliorer l'économie de mouvement définie sur la base de la longueur de la trajectoire suivie par l'instrument et de l'indicateur de distance moyenne. On peut conclure que le retour sonore permet à l'utilisateur d'améliorer sa performance en termes de ces deux critères mais il ne contribue pas à la réduction du temps d'accomplissement de la tâche.

Tableau 7.7. Estimation Généralisée de Moyennes Marginales par l'Analyse de Mesures Répétées avec le modèle d'Equations Généralisées d'Estimation

Facteurs (1)	(2)	Variable	Différence moyenne	Erreur standard	DDL	p	Wald IC 95%		Significatif (o/n) ?
							Inf.	Sup.	
UM+SS	UM+S3E	Temps d'accomplissement (s)	0.81	3.23	1	.801	-5.53	7.16	N
		Longueur de trajectoire (px)	238.3	54.4	1	.000	131.6	345.0	O
		Indicateur de distance moyenne	5.45	0.44	1	.000	4.58	6.33	O

DM+SS	DM+S3E	Temps d'accomplissement (s)	-6.99	3.92	1	.705	-14.69	0.70	N
		Longueur de trajectoire (px)	150.8	31.72	1	.000	88.66	213.0	O
		Indicateur de distance moyenne	3.75	0.38	1	.000	3.00	4.51	O

Comparaison des moyennes marginales pour les différentes conditions expérimentales basée sur l'échelle originale des valeurs des variables. Le niveau de confiance est de 95%.

7.5.2. Analyse du retour d'informations de l'utilisateur

Dans cette partie, on présente l'analyse du retour d'informations de l'utilisateur. Cette analyse permet de déterminer l'effet du retour sonore sur les critères de la facilité d'utilisation de la technique d'interaction et de l'intuitivité du retour sonore.

La Figure 7.20a présente le résultat de l'évaluation du critère de la facilité d'utilisation par rapport aux modalités d'opération du simulateur. On constate que tous les utilisateurs considèrent que le retour sonore est plus facile à utiliser. Le nombre de sujet qui considèrent que le retour sonore est beaucoup plus facile ou plus facile à utiliser diminue lorsque le simulateur est utilisé avec deux mains. Ce sentiment provient du fait que la manipulation à deux mains s'est avérée plus compliquée que celle réalisée avec une seule main en ayant la caméra positionnée à l'avance. Les sujets ont trouvé une difficulté à coordonner les gestes des deux mains pour manipuler la caméra et l'instrument chirurgical.

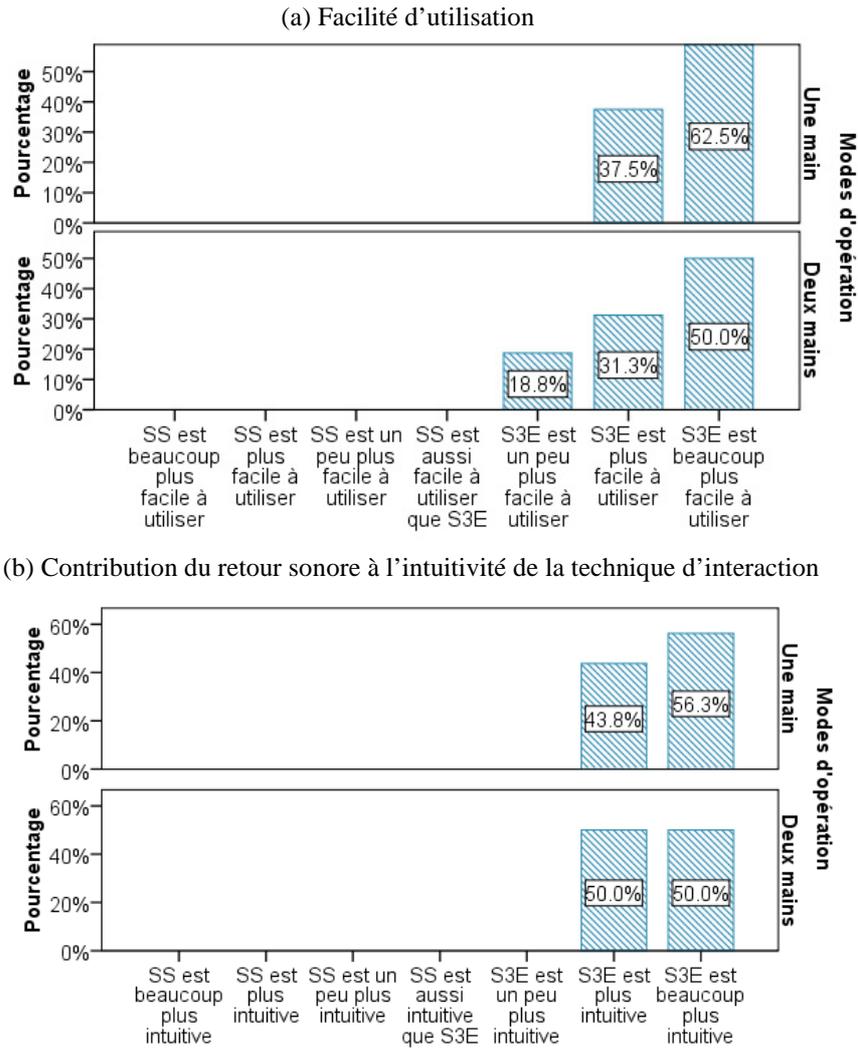


Figure 7.20. Appréciation des sujets de (a) la facilité d'utilisation et de (b) l'intuitivité.

La Figure 7.20b montre le résultat de l'évaluation de l'intuitivité par rapport aux différentes modalités d'opération du simulateur. Tous les utilisateurs considèrent que le retour sonore est plus intuitif. Le nombre d'utilisateurs qui considèrent que le retour sonore est beaucoup plus intuitif ou plus intuitif diminue lorsque le simulateur est utilisé avec deux mains.

7.6. Conclusions

Dans ce chapitre, nous avons présenté une étude sur l'intégration de l'approche de Son 3D Enrichi pour guider le geste de triangulation dans la simulation de chirurgie arthroscopique. Nous avons effectué cette étude expérimentale pour déterminer l'effet du guidage auditif sur la performance de l'utilisateur et sur la facilité

d'utilisation de la technique d'interaction et l'intuitivité du retour sonore utilisé. Notre objectif est de déterminer la contribution du guidage auditif à l'économie de mouvement du geste chirurgical. La tâche étudiée consiste à visualiser et à toucher une cible située dans une position inconnue à l'intérieur du modèle de l'épaule du patient.

Nous avons développé une méthode pour intégrer l'approche de guidage auditif pour la manipulation 3D en environnement virtuel et nous avons intégré cette approche dans la simulation de chirurgie arthroscopique. L'aspect important de cette méthode d'intégration est celui de l'association entre les degrés de liberté de l'outil d'interaction et les propriétés spatiales de la source sonore utilisée. Le retour sonore permet à l'utilisateur d'orienter correctement l'instrument afin de l'introduire dans le cône de vision de la caméra pour toucher la cible virtuelle.

Les résultats obtenus montrent que le retour sonore n'a pas d'effet significatif sur le temps d'accomplissement de la tâche. Par contre, il contribue à l'économie de mouvement en termes des critères de la longueur de trajectoire suivie par l'instrument et de la distance moyenne entre la trajectoire parcourue et la trajectoire optimale de l'instrument. On peut conclure que le retour sonore permet de réduire les risques sans augmenter le temps nécessaire pour réaliser la tâche. D'après l'évaluation des indicateurs de perception, les utilisateurs considèrent que le retour sonore facilite la tâche et la rend plus intuitive.

Dans ce chapitre, nous avons démontré la contribution du retour sonore à l'économie de mouvement pour assister la manipulation de l'instrument chirurgical. On considère que le guidage auditif présente des avantages dans la simulation de la chirurgie arthroscopique. L'assistance auditive pour la manipulation de l'instrument dans la simulation chirurgicale permet à l'opérateur de se concentrer sur la navigation à l'intérieur du modèle de l'épaule ainsi que sur l'opération de la caméra et sur l'acte chirurgical.

8. Conclusions et perspectives de recherche

8.1. Conclusions

8.1.1. Problématique scientifique traitée

Cette thèse traite de l'utilisation du son 3D pour améliorer l'interaction en environnement virtuel tridimensionnel. En général, les techniques d'interaction en environnement virtuel sont basées sur le retour visuel et le retour haptique. On considère que le potentiel du son 3D pour l'interaction n'a pas été suffisamment exploité dans le contexte de la réalité virtuelle.

Le guidage auditif est traditionnellement basé sur le son non spatial. Il ne permet pas de transmettre des informations spatiales précises. Les techniques de guidage auditif publiées dans la littérature sont limitées à la transmission d'informations non spatiales associées à des variables unidimensionnelles et au déclenchement d'évènements.

Dans ce travail, nous avons étudié les limites des techniques existantes du son 3D pour transmettre des informations spatiales précises. Les techniques existantes ne permettent pas la localisation précise de source sonore avec précision. Différents auteurs ont reporté une erreur de localisation globale comprise entre 10 et 25 degrés. Les techniques de son 3D qui permettent d'obtenir un bon niveau de précision de localisation de source sonore sont basées sur l'individualisation. Cette dernière engendre des problèmes d'ordre pratique.

Dans cette thèse, nous avons développé une technique de stimulation auditive basée sur le son 3D qui ne nécessite pas l'individualisation. L'objectif de communiquer des informations à l'utilisateur est de l'assister dans la réalisation d'une tâche interactive en environnement virtuel.

Dans cette thèse, nous avons étudié la simulation de la profondeur de sources sonores. D'après des travaux menés dans le domaine de l'écoute spatiale, l'identification de la profondeur d'une source sonore dépend principalement de l'indice sensoriel d'intensité et de la différence interaurale de temps perçues par l'auditeur. Par contre, la plupart des approches de simulation de la profondeur de source sonore spatialisées sont basées uniquement sur l'indice sensoriel d'intensité. Dans notre approche, nous nous sommes intéressés à l'amélioration de la sensation de profondeur fournie par l'indice d'intensité en ajoutant la restitution de la différence interaurale de temps.

8.1.2. Solutions développées

Afin d'obtenir une localisation précise de source sonore, nous avons développé une approche de stimulation auditive spatiale basée sur des indices sensoriels supplémentaires appelés SASA. Ces indices supplémentaires sont générés par l'application d'effets audio à un stimulus sonore. Les approches existantes basées sur la combinaison du son 3D et des informations supplémentaires se limitent au traitement des sources sonores localisées dans le plan horizontal et à l'utilisation simultanée de deux indices supplémentaires au maximum. En revanche, l'approche que nous avons proposée permet de combiner plusieurs indices supplémentaire et elle permet de créer des sources sonores situées dans différentes élévations.

L'approche d'indices SASA est une technique de stimulation auditive spatiale généralisée car elle ne nécessite pas l'individualisation. Nous avons conçu une méthode de génération et une méthode d'intégration d'indices SASA afin de faciliter sa mise en œuvre dans une application de réalité virtuelle.

Pour améliorer la rapidité de la localisation des sources sonores, nous avons combiné la technique de la HRTF et l'approche SASA. Les résultats obtenus montrent que l'intégration des deux méthodes permet de réduire le temps d'accomplissement de la tâche de localisation de source sonore.

Ce travail nous a conduits à développer une nouvelle approche de simulation de la profondeur de source sonore basée sur la combinaison de l'indice d'intensité et des différences binaurales de temps. L'objectif de cette approche est de compenser les différences interaurales de temps propres à la HRTF à rayon fixe utilisée pour restituer les différences interaurales de temps qui correspondent à la profondeur de la source sonore.

L'intégration du modèle combiné (HRTF + SASA) pour la représentation de la direction de provenance et la technique de CITD pour la simulation de la profondeur de source sonore, permet de transmettre des informations auditives spatiales en azimut, en élévation et en profondeur. Ces informations peuvent être utilisées pour le guidage auditif pour assister l'utilisateur dans l'exécution d'une tâche d'interaction 3D en environnement virtuel.

Nous avons réalisé une étude sur l'effet du guidage auditif spatial fourni par la combinaison des techniques HRTF, SASA et CITD ; sur l'économie de mouvement dans la simulation de la chirurgie arthroscopique. Cette étude intègre le retour visuel et le retour sonore pour assister le chirurgien dans la tâche de manipulation de l'instrument chirurgicale et la réalisation de l'opération.

8.1.3. Résultats obtenus

L'approche SASA permet d'obtenir une localisation de source sonore avec une erreur angulaire globale de 3° environ. Ce niveau de précision représente une contribution importante dans le domaine de l'intégration du son 3D en environnement virtuel. Notre approche permet une localisation plus précise avec une erreur de 2° environ en azimut et en élévation.

Le modèle combiné (HRTF + SASA) permet d'obtenir une localisation rapide et précise de source sonore. Le temps de localisation moyen pour le modèle combiné est de l'ordre de 14 secondes environ. L'étude comparative de ces deux techniques montre que les indices SASA contribuent à la précision et la HRTF contribue à la rapidité de localisation de source sonore. Le modèle combiné (HRTF + SASA) permet d'obtenir une performance de localisation de source sonore supérieure à celle reportée dans la littérature.

L'approche de CITD de la simulation de la profondeur de sources sonores permet d'améliorer de la précision de localisation en profondeur de 15% par rapport à l'utilisation de l'indice d'intensité uniquement. Le retour d'informations de l'utilisateur montre que les sujets considèrent que l'approche de CITD permet une représentation de la profondeur de sources sonores.

L'étude du guidage auditif dans la simulation de la chirurgie arthroscopique montre que les informations auditives permettent d'optimiser le mouvement nécessaire effectué dans la phase de triangulation mais il n'améliore pas le temps d'accomplissement de la tâche.

8.2. Perspectives de recherche

Dans la suite de nos travaux, nous envisageons d'étudier l'utilisation du retour sonore spatial pour la navigation assistée. La précision de l'approche de guidage auditif proposée dans cette thèse, peut apporter une aide importante dans la navigation en environnement non structuré « *unstructured environment* ».

Cette approche promet des applications potentielles dans le domaine de la substitution sensorielle (Bachy-Rita [1972]). Nous envisageons de l'exploiter dans la réalisation de tâches spécifiques dédiées aux utilisateurs non voyants et dans des applications pour lesquelles le coût d'un système de retour haptique n'est pas justifié.

Les alertes sonores spatiales utilisent des sources sonores pour transmettre un message à l'utilisateur. La précision de la directivité des alertes sonores spatiales est importante dans certaines applications de simulation de conduite (Ho; Spence (2005)). Nous envisageons d'appliquer notre approche de son spatialisé dans la simulation de conduite. Les indices SASA peuvent également être intégrés dans un simulateur de vol où les cibles volantes sont représentées par des sources sonores spatiales. Ces indices permettront d'améliorer la rapidité de l'identification de la provenance des cibles volantes.

8.3. Publications réalisées

8.3.1. Articles soumis pour la publication dans des revues internationales

- Vladimir Ortega-González, Samir Garbaya, Frédéric Merienne, « The approach of Artificial Spatial Auditory Sensation for precise sound source localization in virtual environment », soumis à l'International Journal of Human-Computer Studies.
- Vladimir Ortega-González, Samir Garbaya, Frédéric Merienne, « Comparative study of Head-Related Transfer Function and Artificial Spatial Auditory Sensation for sound source localization task in virtual environment », soumis au Journal Interacting with Computers.

8.3.2. Articles publiés dans des conférences internationales avec comité de lecture

- Vladimir Ortega-González, Samir Garbaya, Frédéric Merienne, “Reducing reversal errors in localizing the source of sound in virtual environment without head tracking”, Springer Lectures Notes in Computer Sciences, Proceedings of the 5th International Workshop on Haptic and Interaction Design HAID’ 10, Copenhagen, Danemark, Septembre 2010.
- Vladimir Ortega-González, Samir Garbaya, Frédéric Merienne, “Using 3D sound for providing 3D interaction in virtual environment”, ASME World Conference on Innovative Virtual Reality WINVR 2010, Ames, Iowa, États-Unis, Mai 2010, Identifiant ASME de l’article : WINVR10-3750.
- Loïc Corenthy, Vladimir Ortega-González, José M. Espadero-Guillermo, Samir Garbaya, “3D sound for simulation of arthroscopic surgery”, ASME World Conference on Innovative Virtual Reality WINVR 2010, Ames, Iowa, États-Unis, Mai 2010, Identifiant ASME de l’article : WINVR10-3745.
- Vladimir Ortega-Gonzalez, Samir Garbaya, Frédéric Merienne, “An approach for studying the effect of high-level spatial properties of 3D Audio in interactive systems ”, World Conference on Innovative Virtual Reality WINVR 2009, Chalon-sur-Saône, France, Février 2009, Identifiant ASME de l’article : WINVR09-756.

8.3.3. Poster

- Vladimir Ortega-Gonzalez, Samir Garbaya, Frédéric Merienne, “Experimentation with metaphors of 3D sound”, Ancillary proceedings of the 4th International Workshop on Haptic and Interaction Design HAID’ 09, Dresden, Allemagne, Septembre, 2009.

9. Références

- Algazi, V. R., R. O Duda, D. M Thompson, et C. Avendano. 2001. The CIPIC HRTF database. Dans *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the*, 99–102. doi:10.1109/ASPAA.2001.969552.
- Bach-y-Rita, P. 1972. *Brain mechanisms in sensory substitution*. USA: Academic Press.
- Bayona, Sofía, José Manuel Fernández-Arroyo, Isaac Martín, et Pilar Bayona. 2008. Assessment study of insightArthroVR arthroscopy virtual training simulator: face, content, and construct validities. *Journal of Robotic Surgery* 2, no. 3 (Septembre): 151–158.
- Bayona, Sofía, Marcos García, César Mendoza, et José M Fernández. 2006. Shoulder Arthroscopy Training System with Force Feedback. Dans *MEDIVIS '06: Proceedings of the International Conference on Medical Information Visualisation–BioMedical Visualisation*, 71–76. Washington, DC, USA: IEEE Computer Society. doi:http://dx.doi.org/10.1109/MEDIVIS.2006.23.
- Bech, Soren, et Nick Zakarov. 2006. *Perceptual Audio Evaluation: Theory, Model and Application*. England: Willey.
- Begault, Durand. 2005. *3D-Sound for Virtual Reality and Multimedia*. USA: AP Professional.
- Begault, Durand R, Elizabeth M Wenzel, et Mark R Anderson. 2001. Direct Comparison of the Impact of Head Tracking, Reverberation, and Individualized Head-Related Transfer Functions on the Spatial Perception of a Virtual Speech Source. *Journal of the Audio Engineering Society* 49, no. 10: 916.
- Berkhout, A. J. 1988. A Holographic Approach to Acoustic Control. *The Journal of the Audio Engineering Society*, no. 36: 977–995.
- Blattner, Meera M, Denise A Sumikawa, et Robert M Greenberg. 1989. Earcons and icons: their structure and common design principles. *Hum.-Comput. Interact.* 4, no. 1: 11–44. doi:http://dx.doi.org/10.1207/s15327051hci0401_1.
- Blauert, Jens. 1997. *Spatial Hearing: The Psychophysics of Human Sound Localization*. USA: MIT Press.
- Booth, D.A., et R.P.J. Freeman. 1993. Discriminative feature integration by individuals. *Acta Psychologica* 84, no. 1: 1 - 16. doi:DOI: 10.1016/0001-6918(93)90068-3.
- Brewster 2003. Non-speech auditory output. Dans *The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications*, éd. J. A Jacko et A. Sears, Chap. 12, 220–239. Hillsdale, NJ, USA: L. Erlbaum Associates Inc.
- Bronkhorst, A. W, J. A Veltman, et L. van Breda. 1996. Application of a three-dimensional auditory display in a flight task. *Human factors* 38, no. 1 (Mars): 23–33.
- Brown, C. P, et R. O Duda. 1997. An efficient HRTF model for 3-D sound. Dans *Applications of Signal Processing to Audio and Acoustics, 1997. 1997 IEEE ASSP Workshop on*, 4 pp.+.
- Burgess, David A. 1992. Techniques for low cost spatial audio. Dans *UIST '92: Proceedings of the 5th annual ACM symposium on User interface software and technology*, 53–59. New York, NY, USA: ACM. doi:http://doi.acm.org/10.1145/142621.142628.

-
- Carlile, Simon, Philip Leong, et Stephanie Hyams. 1997. The nature and distribution of errors in sound localization by human listeners. *Hearing Research* 114, no. 1-2: 179 - 196. doi:DOI: 10.1016/S0378-5955(97)00161-5.
- Clevin, Lotte, et Teodor P Grantcharov. 2008. Does box model training improve surgical dexterity and economy of movement during virtual reality laparoscopy? A randomised trial. *Acta Obstet Gynecol Scand* 87, no. 1: 99-103.
- Cohen, Michael, et Elizabeth M Wenzel. 1995. The design of multidimensional sound interfaces. Dans *Virtual environments and advanced interface design*, éd. Woodrow Barfield et Thomas A. Furness, 291–346. New York, NY, USA: Oxford University Press, Inc.
- Cooke, M., et Lu Yan-Chen. 2007. Active binaural distance estimation for dynamic sources. Dans . Antwerp, Belgium.
- Cooley, James W., et John W. Tukey. 1965. An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*, no. 19: 297-301.
- Creative-Labs. 1999. *Environmental Audio Extensions: EAX 2.0*. USA: Creative Technology Limited.
- Dallal, Gerard. 2001. *The Little Handbook of Statistical Practice*. Biostatistics Unit Jean Mayer USDA Human Nutrition Research Center on Aging, Tufts University. <http://www.tufts.edu/~gdallal/LHSP.HTM>.
- Delahoussaye, K.A. 2010. *USB HID API Library*. <http://www.usb.org/>.
- Doerr, Kai-Ume, Holger Rademacher, Silke Huesgen, et Wolfgang Kubbat. 2007. Evaluation of a Low-Cost 3D Sound System for Immersive Virtual reality Training Systems. *IEEE Transactions on Visualization and Computer Graphics* 13, no. 2.
- Duda, R O. 1993. Modeling Head Related Transfer Functions. Dans *Proc. 27th Asilomar conf. on Signal, Systems and Computers, Asilomar, CA*, 457–461.
- Duda, R.O. 2000. *3D audio for HCI*. USA: Department of Electrical Engineering, San Jose State University.
- Durlach, N. I., B. G. Shinn-Cunningham, et R. M. Held. 1993. Supernormal auditory localization. I. General background. *Presence* 2, no. 2: 89–103.
- Filipanits, Frank. 1994. Design and implementation of an auralization system with a spectrum-based temporal. USA: University of Miami, Mai.
- FMOD Technologies. 2005. *FMOD Official Documentation*.
- Foster, Scott H., et Elizabeth M. Wenzel. 1992. The Convolvotron: Real-time demonstration of reverberant virtual acoustic environments. *The Journal of the Acoustical Society of America* 92, no. 4: 2376-2376. doi:10.1121/1.404833.
- Fukuda, Takuya, Toshiharu Horiuchi, Haruhide Hokari, et Shoji Shimada. 2003. Relative distance perception by manipulating the ILD of HRTFs. *Acoustical Science and Technology* 24, no. 5: 325-326.
- García Berro, Myriam, et Concha Toribio. 2004. *Ciencias de la Salud, El Futuro de la Cirugía Mínimamente Invasiva. Tendencias tecnológicas a medio y largo plazo*. Prospective report. Madrid, Spain: Fundación OPTI,FENIN, Novembre.
- Gardner, Bill, et K. Martin. 1994. *HRTF Measurements of a KEMAR Dummy-Head Microphone*. USA: MIT Media Lab Perceptual Computing.
- Gardner, W. 2001. *3D Audio and Acoustic Environment Modeling*. Janvier.
- Gaver, W. 1986. Auditory icons: using sound in computer interfaces. *Human Computer Interaction* 2, no. 2: 167–177. doi:http://dx.doi.org/10.1207/s15327051hci0202_3.
- . 1989. The SonicFinder: an interface that uses auditory icons. *Human Computer Interaction* 4, no. 1: 67–94. doi:http://dx.doi.org/10.1207/s15327051hci0401_3.
- Gerzon, M. A. 1973. Periphony: With-Height Sound Reproduction. *Journal of the Audio Engineering Society* 21, no. 1 (Février): 2-10.
- GMV. 2009. *Insight user's guide*.
- Goldstein, E. Bruce. 1999. *Sensation and Perception*. USA: Brooks/Cole Publishing, 6th édition.
- Grohn, M. 2002. Localization of a moving virtual sound source in a virtual room, the effect of a distracting
-

-
- auditory stimulus. Dans *Proceedings of the 8th International Conference on Auditory Display (ICAD2002)*, éd. R. Nakatsu et H. Kawahara. Kyoto, Japan: Advanced Telecommunications Research Institute (ATR), Kyoto, Japan. Proceedings/2002/Groehn2002.pdf.
- Grohn, M., T. Lokki, et T. Takala. 2002. Static and dynamic sound source localization in a virtual room. Dans *Proc. AES 22nd Int. Conf. on Virtual, Synthetic and Entertainment Audio*, 15–17.
- Grohn, M., et T. Takala. 1994. MagicMikes - Multiple Aerial Probes for Sonification of Spatial Databases. Dans *Int. Conf. on Auditory Displays (ICAD'94)*, 108. Santa Fe, USA.
- . 1995. MagicMikes - Method for Spatial Sonification. Dans *IS&T/SPIE Symposium on Electronic Imaging: Science & Technology*, 108. San Jose, USA.
- Gurusamy, KS, R Aggarwal, L Palanivelu, et BR Davidson. 2009. Virtual reality training for surgical trainees in laparoscopic surgery. *Cochrane Database Syst Rev.* 21, no. 1 (Janvier).
- Hendrix, C., et W. Barfield. 1995. Presence in virtual environments as a function of visual and auditory cues. Dans *VRAIS '95: Proceedings of the Virtual Reality Annual International Symposium (VRAIS'95)*, 74. Washington, DC, USA: IEEE Computer Society.
- Ho, Cristy, et Charles Spence. 2005. Assessing the effectiveness of various auditory cues in capturing a driver's visual attention. *Journal of Experimental Psychology: Applied* 11, no. 3 (Septembre): 157-174.
- Huopaniemi, J., L. Savioja, et T. Takala. 1996. DIVA Virtual Audio Reality System. Dans *Proceedings of the 3rd International Conference on Auditory Display (ICAD 96)*. Palo Alto, California, USA.
- IBM. 2009. *SPSS Statistics 17.0*.
- ITU-T. 1996. *Recommendation P.800, Methods for Subjective Determination of Transmission Quality*.
- Jin, Craig, Anna Corderoy, Simon Carlile, et André van Schaik. 2004. Contrasting monaural and interaural spectral cues for human sound localization. *Journal of the Acoustical Society of America* 115, no. 6: 3124-3141. doi:10.1121/1.1736649.
- John, Kamen Guentchev, et John J. Weng. 1998. *Learning-Based Three Dimensional Sound Localization Using a Compact Non-Coplanar Array of Microphones*.
- Jot, Jean-Marc, Veronique Larcher, et Olivier Warusfel. 1995. Digital Signal Processing Issues in the Context of Binaural and Transaural Stereophony. Dans *AES: Convention of the Audio Engineering Society*. Paris, France, Fevrier. <http://articles.ircam.fr/index.php?Action=ShowArticle&IdArticle=52&ViewType=1>.
- Jot, Jean-Marc, et Olivier Warusfel. 1995. A Real-Time Spatial Sound Processor for Music and Virtual Reality Applications. Dans *ICMC: International Computer Music Conference*, 294-295. Banff, Canada, Septembre. <http://articles.ircam.fr/index.php?Action=ShowArticle&IdArticle=all&ViewType=1>.
- Kan, Alan, Craig Jin, et André van Schaik. 2009. A psychophysical evaluation of near-field head-related transfer functions synthesized using a distance variation function. *The Journal of the Acoustical Society of America* 125, no. 4: 2233-2242. doi:10.1121/1.3081395.
- Kan, Alan, Graeme Pope, Andre van Schaik, et Craig T. Jin. 2004. Mobile Spatial Audio Communication System. Dans *Proc of ICAD 04-Tenth Meeting of the International Conference on Auditory Display*. Sydney, Australia.
- Kapralos, Bill, Michael R. M. Jenkin, et Evangelos E. Milios. 2008. Virtual Audio Systems. *Presence* 17, no. 6: 527-549.
- Kapralos, Bill, et Nathan Mekuz. 2007. Application of dimensionality reduction techniques to HRTFS for interactive virtual environments. Dans *ACE '07: Proceedings of the international conference on Advances in computer entertainment technology*, 256–257. New York, NY, USA: ACM. doi:<http://doi.acm.org/10.1145/1255047.1255112>.
- Kilgard, Mark J. 1996. *The OpenGL Utility Toolkit (GLUT) ver. 3.7*.
- Knudsen, Eric, et Masakazu Konishi. 1979. Mechanisms of sound localization in the barn owl. *Journal of Comparative Physiology* 133, no. 1: 13–21.
- Kongsberg Oil Gas Technologies. 2010. *Coin3D*. <http://www.coin3d.org/>.
- Kramer, G. 1992. *An introduction to auditory display*. Éd. G. Kramer. SFI studies in the sciences of complexity. Addison Wesley Longman. Proceedings/1992/Kramer1992a.pdf.
-

-
- Kyriakakis, C., P. Tsakalides, et T. Holman. 1999. Surrounded by sound. *IEEE Signal Processing Magazine* 16, no. 1: 55–66.
- Langendijk, E. H. A. 2002. Contribution of spectral cues to human sound localization. *Acoustical Society of America Journal* 112 (Octobre): 1583+. doi:10.1121/1.1501901.
- Larsson, Pontus, Daniel Vastfjall, et Mendel Kleiner. 2008. Effects of auditory information consistency and room acoustic cues on presence in virtual environments. *Acoustical Science and Technology* 29, no. 2: 191-194.
- Lokki, Tapio, et Matti Grohn. 2005. Navigation with Auditory Cues in a Virtual Environment. *IEEE MultiMedia* 12, no. 2: 80–86. doi:http://dx.doi.org/10.1109/MMUL.2005.33.
- Loomis, JM, R Klatzky, et R Golledge. 1999. Auditory Distance Perception in Real, Virtual and Mixed Environments. Dans *Mixed Reality: Merging Real and Virtual Worlds*. Tokyo: Y. Ohta and H. Tamura.
- MacDonald, Justin, et Phuong K. Tran. 2008. The Effect of Head-Related Transfer Function Measurement Methodology on Localization Performance in Spatial Audio Interfaces. *Human factors* 50, no. 2: 256-263.
- Makous, James C, et John C Middlebrooks. 1990. Two-dimensional sound localization by human listeners. *The Journal of the Acoustical Society of America* 87, no. 5: 2188–2200.
- Marentakis, Georgios, et Stephen A. Brewster. 2004. A study on gestural interaction with a 3D audio display. *Lecture Notes in Computer Science*: 180-191.
- . 2005. A comparison of feedback cues for enhancing pointing efficiency in interaction with spatial audio displays. Dans *MobileHCI '05: Proceedings of the 7th international conference on Human computer interaction with mobile devices & services*, 55–62. New York, NY, USA: ACM. doi:http://doi.acm.org/10.1145/1085777.1085787.
- Martens, William L. 2003. Perceptual evaluation of filters controlling source direction: Customized and generalized HRTFs for binaural synthesis. *Acoustical Science and Technology* 24, no. 5: 220–232.
- McKinley, R., W. R. D'Angelo, M. W. Hass, D.R. Perrot, W.T. Nelson, L.J. Hettinger, et B.J. Brickman. 1995. An initial study of the effects of 3-dimensional auditory cueing on visual target detection. Dans *39th Human Factors and Ergonomics Society Annual Meeting*, 119–123. USA.
- Ménelas, Bob, Lorenzo Pincinali, Brian F.G. Katz, Patrick Bourdot, et Mehdi Ammi. 2009. Haptic Audio Guidance for Target Selection in a Virtual Environment. Dans *HAID '09: Ancillary Proceedings of the 4th International Haptic and Auditory Interaction Design Workshop*, 12-13. Technical University of Dresden.
- Mercier, Dany-Jack. 2010. *Cahiers de Mathématiques du Supérieur*. Vol. 1. France: Publibook. http://megamaths.perso.neuf.fr/exgeo/cms1.html.
- Microsoft. 2003. *Microsoft Visual Studio 7.0*. http://www.microsoft.com/.
- Middlebrooks, John C., et David M. Green. 1991. Sound Localization by Human Listeners. *Annual Review of Psychology* 42, no. 1: 135–159.
- Moody, Louise, et A. Waterworth. 2004. A Flexible Virtual Reality Tutorial for the Training and Assessment of Arthroscopic Skills. Dans *Medicine Meets Virtual Reality 12. Building a Better You: The Next Tools for Medical Education, Diagnosis and Care.*, éd. J.D. Westwood, \bf 98:244–246. Studies in Health Technology and Informatics. Nieuwe Hemweg 6B, 1013 BG Amsterdam, The Netherlands: IOS Press, Janvier.
- Music, Royalty Free. 2009. *Sound Rangers sound library*.
- Naef, Martin, Oliver Staadt, et Markus Gross. 2002. Spatialized audio rendering for immersive virtual environments. Dans *VRST '02: Proceedings of the ACM symposium on Virtual reality software and technology*, 65–72. New York, NY, USA: ACM. doi:http://doi.acm.org/10.1145/585740.585752.
- Nintendo. 2008. *Wii Operations Manual*. http://www.nintendo.com/.
- Norusis, Marija J. 2009. *SPSS Statistics 17.0 Advanced Statistical Procedures Companion*. USA: Prentice Hall.
- Novint Technologies Inc. 2007. *Novint Falcon User Manual*. http://home.novint.com/.
- Oldfield, S.R., et S.P.A. Parker. 1984. Acuity of sound localisation: a topography of auditory space. II. Pinna
-

-
- cues absent. *Perception* 19, no. 5: 601-607.
- Oppenheim, Alan V, Ronald W Schafer, et John R Buck. 1999. *Discrete-time signal processing (2nd ed.)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc.
- Peek, Brian. 2009. *Wiiyourself! Native C++ Wiimote Library ver. 1.01*. <http://wiiyourself.gl.tter.org/>.
- Pralong, D., et Simon. Carlile. 1994. Measuring the human head-related transfer functions: construction and calibration of a miniature in ear recording system. *Journal of the Acoustical Society of America* 95, no. 6: 3435-3444.
- Proakis, John, et Dimitris G. Manolakis. 2007. *Digital Signal Processing, Principles, Algorithms and Applications*. Fourth Edition. Prentice Hall.
- Pulkki, Ville. 2001. Spatial sound generation and perception by amplitude panning techniques. PhD in Technology, Helsinki, Finland: Department of Electrical and Communications Engineering, Helsinki University of Technology.
- Roads, Curtis. 1996. *The Computer Music Tutorial*. Cambridge, MA, USA: MIT Press.
- Rocchesso, Davide. 2003. *The sounding object*. Italy: Mondo Estremo.
- Rocchesso, Davide, Roberto Bresin, et Mikael Fernstrom. 2003. Sounding Objects. *IEEE MultiMedia* 10, no. 2: 42-52. doi:<http://doi.ieeecomputersociety.org/10.1109/MMUL.2003.1195160>.
- Sánchez, Jaime, et Mauricio Sáenz. 2005. 3D sound interactive environments for problem solving. Dans *Assets '05: Proceedings of the 7th international ACM SIGACCESS conference on Computers and accessibility*, 173-179. New York, NY, USA: ACM. doi:<http://doi.acm.org/10.1145/1090785.1090817>.
- Savioja, Lauri, Jyri Huopaniemi, Tapio Lokki, et Ritta Vaananen. 1999. Creating Interactive Virtual Acoustic Environments. *Journal of the Audio Engineering Society* 47, no. 9 (Septembre): 675-705.
- Segal, Mark, et Kurt Akeley. 2006. *The OpenGL Graphics System: A Specification Version 2.1*.
- Sensible Technologies. 2010. *Phantom Omni*. <http://www.sensible.com/>.
- Shelley, Simon, Miguel Alonso, Jacqueline Hollowood, Michael Pettitt, Sarah Sharples, Dik Hermes, et Armin Kohlrausch. 2009. Interactive Sonification of Curve Shape and Curvature Data. Dans *Haptic and Audio Interaction Design*, éd. M. Altinsoy, Ute Jekosch, et Stephen Brewster, 5763:51-60. Lecture Notes in Computer Science. Springer Berlin / Heidelberg. http://dx.doi.org/10.1007/978-3-642-04076-4_6.
- Sherman, K. P, J. W Ward, D. P Wills, V. J Sherman, et A. M Mohsen. 2001. Surgical trainee assessment using a VE knee arthroscopy training system (VE-KATS): experimental results. Dans *Medicine Meets Virtual Reality 2001. Outer space, Inner space, Virtual space*, éd. Westwood J.D. et al, 81:465-470. Studies in Health Technology and Informatics. Nieuwe Hemweg 6B, 1013 BG Amsterdam, The Netherlands: IOS Press.
- Shinn-Cunningham, Barbara G., Nathaniel I. Durlach, et Richard M. Held. 1998a. Adapting to supernormal auditory localization cues. I. Bias and resolution. *Journal of the Acoustical Society of America* 103, no. 5: 3656-3666.
- . 1998b. Adapting to supernormal auditory localization cues. II. Constraints on adaptation of mean response. *Journal of the Acoustical Society of America* 103, no. 6: 3667-3676.
- Silzle, Andreas, Holger Strauss, et Pedro Novo. 2004. IKA-SIM: A System to Generate Auditory Virtual Environments. Dans *Proc. 116th AES Convention*.
- Smith, Steven W. 1999. *Digital Signal Processing, A practical guide for engineers and scientists*. Newnes.
- Spiegel, Murray R., John J. Schiller, et R. Alu Srinivasan. 2000. *Schaum Outline of Probability and Statistics*. 2 éd. McGraw-Hill.
- StatSoft. 2009. *Statistica 9.0*. <http://www.statsoft.com/>.
- Sutherland, LM. 2006. *Surgical simulation: a systematic review*. Report. Adelaide, South Australia: ASERNIP-S, Août.
- Takala, T., R. Hanninen, V. Valimaki, L. Savioja, J. Huopaniemi, T. Huotilainen, et M. Karjalainen. 1996. An integrated system for virtual audio reality. Dans *100th Audio Engineering Society (AES) Convention*. Copenhagen, Denmark.
-

-
- Tuchschmid, S, M Bajka, D Bachofen, G Szekely, et M Harders. 2007. Objective surgical performance assessment for virtual hysteroscopy. *Stud Health Technol Inform* 1, no. 125: 473-478.
- Valin, J-M, F. Michaud, J. Rouat, et D. Létourneau. 2003. Robust sound source localization using a microphone array on a mobile robot. Dans *Proceedings International Conference on Intelligent Robots and Systems*, 1228–1233.
- Vazquez Buenos-Aires, J.O., Y. Payan, et J. Demongeot. 2003. Electro-stimulation of the tongue as a passive surgical guiding system. Dans , 638-643.
- Vazquez-Alvarez, Yolanda, et Stephen Brewster. 2009. Investigating background & foreground interactions using spatial audio cues. Dans *CHI EA '09: Proceedings of the 27th international conference extended abstracts on Human factors in computing systems*, 3823–3828. New York, NY, USA: ACM. doi:<http://doi.acm.org/10.1145/1520340.1520578>.
- Vorlander, Michael. 2007. *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality (RWTH edition)*. Springer.
- Walker, Bruce N, et Jeffrey Lindsay. 2006. Navigation Performance With a Virtual Auditory Display: Effects of Beacon Sound, Capture Radius, and Practice. *Hum Factors* 48, no. 2 (Janvier): 265–278. doi:10.1518/001872006777724507.
- Warusfel, Olivier. 2003. *The LISTEN HRTF database*.
- Wenzel, E. M. 1994. Spatial Sound and Sonification. Dans *Auditory Display: Sonification, Audification, and Auditory Interfaces*, XVIII:127-150. G. Kramer.
- Wenzel, E. M., M. Arruda, D. Kistler, et F. Wightman. 1993. Localization using non-individualized head-related transfer functions. *Journal of the Acoustical Society of America*.
- Wightman, Frederic L, et Doris J Kistler. 1989a. Headphone simulation of free-field listening. I: Stimulus synthesis. *Journal of the Acoustical Society of America* 85, no. 2: 858–867.
- . 1989b. Headphone simulation of freefield listening II: psychophysical validation. *J. Acoust. Soc. Amer* 85, no. 2: 868–878.
- Wu, Jiann-Rong, Cha-Dong Duh, Ming Ouhyoung, et Jei-Tun Wu. 1997. Head motion and latency compensation on localization of 3D sound in virtual reality. Dans *VRST '97: Proceedings of the ACM symposium on Virtual reality software and technology*, 15–20. New York, NY, USA: ACM. doi:<http://doi.acm.org/10.1145/261135.261140>.
- Xsens. 2009. *MTi and MTx User Manual and Technical Documentation*.
- Zahorik, Pavel. 2002a. Assessing auditory distance perception using virtual acoustics. *The Journal of the Acoustical Society of America* 111, no. 4: 1832-1846. doi:10.1121/1.1458027.
- . 2002b. Auditory display of sound source distance. Dans *Proceedings of the 8th International Conference on Auditory Display (ICAD2002)*, éd. R. Nakatsu et H. Kawahara. Kyoto, Japan: Advanced Telecommunications Research Institute (ATR), Kyoto, Japan. Proceedings/2002/Zahorik2002.pdf.
- Zahorik, Pavel, C. Tam, K. Wang, P. Bangayan, et V. Sundareswaran. 2001. Localization Accuracy In 3-D Sound Displays: The Role Of Visual-Feedback Training. Dans *Proceedings of the Advanced Displays and Interactive Displays Federal Laboratory Consortium*.
-

10. Annexe : Questionnaires utilisés

10.1. Analyse du retour d'informations de l'utilisateur : Chapitre 5

Date : ___/___/___

Heure de début : ___h___

Heure de fin : ___h___

Données du participant

Prénom: _____	Nom: _____	
Age: _____	Féminin / Masculin	Droitier / Gaucher
Maladies visuelles ou auditives : _____		
Téléphone: _____	E-mail: _____	

Instructions: Sélectionner une seule réponse pour chaque question

- Parmi les conditions SASA et HRTF, quelle est la condition la plus réaliste?**
 - 3 HRTF est beaucoup plus réaliste
 - 2 HRTF est plus réaliste
 - 1 HRTF est un peu plus réaliste
 - 0 SASA est aussi réaliste que HRTF
 - 1 SASA est un peu plus réaliste
 - 2 SASA est plus réaliste
 - 3 SASA est beaucoup plus réaliste

- Parmi les conditions SASA et HRTF, quelle est la condition la plus facile à utiliser?**
 - 3 HRTF est beaucoup plus facile à utiliser
 - 2 HRTF est plus facile à utiliser
 - 1 HRTF est un peu plus facile à utiliser
 - 0 SASA est aussi facile à utiliser que HRTF
 - 1 SASA est un peu plus facile à utiliser
 - 2 SASA est plus facile à utiliser
 - 3 SASA est beaucoup plus facile à utiliser

3. Parmi les conditions HRTF+SASA et HRTF, quelle est la condition la plus réaliste?

- 3 HRTF est beaucoup plus réaliste
- 2 HRTF est plus réaliste
- 1 HRTF est un peu plus réaliste
- 0 HRTF+SASA est aussi réaliste que HRTF
- 1 HRTF+SASA est un peu plus réaliste
- 2 HRTF+SASA est plus réaliste
- 3 HRTF+SASA est beaucoup plus réaliste

4. Parmi les conditions HRTF+SASA et HRTF, quelle est la condition la plus facile à utiliser?

- 3 HRTF est beaucoup plus facile à utiliser
- 2 HRTF est plus facile à utiliser
- 1 HRTF est un peu plus facile à utiliser
- 0 HRTF+SASA est aussi facile à utiliser que HRTF
- 1 HRTF+SASA est un peu plus facile à utiliser
- 2 HRTF+SASA est plus facile à utiliser
- 3 HRTF+SASA est beaucoup plus facile à utiliser

5. Parmi les conditions HRTF+SASA et SASA, quelle est la condition la plus réaliste?

- 3 SASA est beaucoup plus réaliste
- 2 SASA est plus réaliste
- 1 SASA est un peu plus réaliste
- 0 HRTF+SASA est aussi réaliste que SASA
- 1 HRTF+SASA est un peu plus réaliste
- 2 HRTF+SASA est plus réaliste
- 3 HRTF+SASA est beaucoup plus réaliste

6. Parmi les conditions HRTF+SASA et SASA, quelle est la condition la plus facile à utiliser?

- 3 SASA est beaucoup plus facile à utiliser
- 2 SASA est plus facile à utiliser
- 1 SASA est un peu plus facile à utiliser
- 0 HRTF+SASA est aussi facile à utiliser que SASA
- 1 HRTF+SASA est un peu plus facile à utiliser
- 2 HRTF+SASA est plus facile à utiliser
- 3 HRTF+SASA est beaucoup plus facile à utiliser

Commentaires du participant

10.2. Analyse du retour d'informations de l'utilisateur : Chapitre 6

Date : __/__/____

Heure de début : __h__

Heure de fin : __h__

Données du participant

Prénom: _____	Nom: _____
Age: _____	Féminin / Masculin Droitier / Gaucher
Maladies visuelles ou auditives : _____	
Téléphone: _____	E-mail: _____

Instructions: Sélectionner une seule réponse pour chaque question

Comment vous jugez la qualité de la perception de profondeur dans la modalité 2 (CITD) comparée à la modalité 1 (Normale) pour.....

1. ...la trajectoire 1 ?

- 3 Normale est beaucoup mieux
- 2 Normale est mieux
- 1 Normale est un peu mieux
- 0 CITD est aussi bien que Normale
- 1 CITD est un peu mieux
- 2 CITD est mieux
- 3 CITD est beaucoup mieux

2. ...la trajectoire 2 ?

- 3 Normale est beaucoup mieux
- 2 Normale est mieux
- 1 Normale est un peu mieux
- 0 CITD est aussi bien que Normale
- 1 CITD est un peu mieux
- 2 CITD est mieux
- 3 CITD est beaucoup mieux

3. ...la trajectoire 3 ?

- 3 Normale est beaucoup mieux
- 2 Normale est mieux
- 1 Normale est un peu mieux
- 0 CITD est aussi bien que Normale
- 1 CITD est un peu mieux
- 2 CITD est mieux
- 3 CITD est beaucoup mieux

4. ...la trajectoire 4 ?

- 3 Normale est beaucoup mieux
- 2 Normale est mieux
- 1 Normale est un peu mieux

- 0 CITD est aussi bien que Normale
- 1 CITD est un peu mieux
- 2 CITD est mieux
- 3 CITD est beaucoup mieux

5. ...la trajectoire 5 ?

- 3 Normale est beaucoup mieux
- 2 Normale est mieux
- 1 Normale est un peu mieux
- 0 CITD est aussi bien que Normale
- 1 CITD est un peu mieux
- 2 CITD est mieux
- 3 CITD est beaucoup mieux

6. ...la trajectoire 6 ?

- 3 Normale est beaucoup mieux
- 2 Normale est mieux
- 1 Normale est un peu mieux
- 0 CITD est aussi bien que Normale
- 1 CITD est un peu mieux
- 2 CITD est mieux
- 3 CITD est beaucoup mieux

7. ...la trajectoire 7 ?

- 3 Normale est beaucoup mieux
- 2 Normale est mieux
- 1 Normale est un peu mieux
- 0 CITD est aussi bien que Normale
- 1 CITD est un peu mieux
- 2 CITD est mieux
- 3 CITD est beaucoup mieux

8. ...la trajectoire 8 ?

- 3 Normale est beaucoup mieux
- 2 Normale est mieux
- 1 Normale est un peu mieux
- 0 CITD est aussi bien que Normale
- 1 CITD est un peu mieux
- 2 CITD est mieux
- 3 CITD est beaucoup mieux

9. ...la trajectoire 9 ?

- 3 Normale est beaucoup mieux
- 2 Normale est mieux
- 1 Normale est un peu mieux
- 0 CITD est aussi bien que Normale
- 1 CITD est un peu mieux
- 2 CITD est mieux
- 3 CITD est beaucoup mieux

10. ...la trajectoire 10 ?

- 3 Normale est beaucoup mieux
- 2 Normale est mieux
- 1 Normale est un peu mieux
- 0 CITD est aussi bien que Normale
- 1 CITD est un peu mieux
- 2 CITD est mieux
- 3 CITD est beaucoup mieux

11. ...la trajectoire 11 ?

- 3 Normale est beaucoup mieux
- 2 Normale est mieux
- 1 Normale est un peu mieux
- 0 CITD est aussi bien que Normale
- 1 CITD est un peu mieux
- 2 CITD est mieux
- 3 CITD est beaucoup mieux

12. Comment vous jugez la qualité du réalisme de la modalité 2 (CITD) comparée à la modalité 1 (Normale) pour l'ensemble des trajectoires ?

- 3 Normale est beaucoup mieux
- 2 Normale est mieux
- 1 Normale est un peu mieux
- 0 CITD est aussi bien que Normale
- 1 CITD est un peu mieux
- 2 CITD est mieux
- 3 CITD est beaucoup mieux

Commentaires du participant

10.3. Analyse du retour d'informations de l'utilisateur : Chapitre 7

Date : ___/___/___

Heure au début : ___h___

Heure à la fin : ___h___

Données du participant

Prénom: _____	Nom: _____
Age: _____	Féminin / Masculin Droitier / Gaucher
Maladies visuelles ou auditives : _____	
Téléphone: _____	E-mail: _____

Modalité d'opération évaluée : _____ avec une main ou _____ avec deux mains

Instructions: Sélectionner une seule réponse pour chaque question

1. Parmi les modalités de stimulation SS (sans son) et S3E (Son 3D Enrichi), quelle est la plus facile à utiliser?

- 3 SS est beaucoup plus facile à utiliser
- 2 SS est plus facile à utiliser
- 1 SS est un peu plus facile à utiliser
- 0 S3E est aussi facile à utiliser que SS
- 1 S3E est un peu plus facile à utiliser
- 2 S3E est plus facile à utiliser
- 3 S3E est beaucoup plus facile à utiliser

2. Parmi les modalités de stimulation SS (sans son) et S3E (Son 3D Enrichi), quelle est la plus intuitive?

- 3 SS est beaucoup plus intuitive
- 2 SS est plus intuitive
- 1 SS est un peu plus intuitive
- 0 S3E est aussi intuitive que SS
- 1 S3E est un peu plus intuitive
- 2 S3E est plus intuitive
- 3 S3E est beaucoup plus intuitive

Commentaires du participant

11. Glossaire

Angle global	L'angle entre le vecteur de l'orientation de la tête de l'auditeur et le vecteur de la position de la source sonore
Arthroscopie	Type de chirurgie mini-invasive (MIS) spécialisée dans les grandes articulations telles que le genou et l'épaule
Auralisation	Ensemble des techniques de rendus sonores basées sur des données numériques
Azimut	L'angle entre les projections sur le plan horizontal des vecteur suivants : (1) Le vecteur de l'orientation de tête de l'auditeur et (2) le vecteur formé par la source sonore et la tête de l'auditeur
CITD	<i>Compensated Interaural Time Differences</i> : une approche qui consiste à restituer la profondeur d'une source sonore basée sur la compensation des ITDs d'une fonction HRTF mesurée à une profondeur fixe en fonction de la distance souhaitée entre l'auditeur et la source sonore virtuelle
Convolution	Méthode mathématique permettant de combiner deux signaux. Elle est utilisée pour appliquer une fonction HRTF à un stimulus sonore afin d'obtenir une source sonore virtuelle
Directionalisation	Ce terme représente le fait d'associer à un stimulus sonore, les propriétés spatiales de directivité d'une position donnée dans l'espace
Directivité	La direction de provenance d'une source sonore dans l'espace, elle est définie par les propriétés spatiales d'azimut et d'élévation
Élévation	L'angle entre le plan horizontal et le vecteur de l'orientation de la tête de l'auditeur
Erreur inverse	Elle se produit lorsque l'auditeur estime que la source sonore se trouve dans un hémisphère opposé à celui de sa position réelle. Les erreurs inverses sont

	les suivantes : haut/bas, devant/derrière et gauche/droite
Généralisation	La généralisation de la spatialisation sonore implique que le processus de traitement ne dépend pas des caractéristiques individuelles de l'auditeur
HRIR	<i>Head-Related Impulse Response</i> : La réponse impulsionnelle de l'anatomie de l'auditeur représente la différence entre le son émis depuis une source située dans une position donnée et celui qui arrive aux oreilles de l'auditeur
HRTF	<i>Head-Related Transfer Function</i> : Les fonctions de transfert de tête sont des ensembles de transformations mathématiques qui peuvent être appliquées à un signal sonore mono. Les signaux gauches et droit résultant sont les mêmes que les signaux qu'un auditeur perçoit en écoutant un son qui provient d'un emplacement dans le monde réel
ILD	Interaural Level Difference : Ce terme se réfère à la différence d'intensité des sons qui arrivent aux oreilles de l'auditeur et le son émis à la source
Individualisation	L'individualisation de la spatialisation sonore implique l'adaptation du processus de traitement aux caractéristiques individuelles de l'auditeur
ITD	<i>Interaural Time Difference</i> : Ce terme se réfère à la différence de temps entre l'émission du son à la source et son arrivée aux oreilles de l'auditeur
Localisation sonore	L'identification des propriétés spatiales d'une source sonore réalisée par un auditeur
MIS	<i>Mini-Invasive Surgery</i> : La chirurgie mini-invasive désigne un type d'intervention dans laquelle les instruments chirurgicaux sont introduits dans le corps au travers de petites incisions
Occlusion visuelle	L'absence de retour visuel dans la réalisation d'une tâche
Profondeur	La distance entre la source sonore et l'auditeur
Propriétés spatiales	Les propriétés spatiales d'une source sonore définissent sa position dans l'espace : l'azimut, l'élévation, la profondeur
Restitution binaurale	Rendu sonore réalisé à deux canaux pour être restitué directement aux oreilles de l'utilisateur
SASA	L'approche des Sensations Artificielles Spatialisées Auditives consiste en la restitution des propriétés spatiales d'une source sonore à travers l'application des effets sonores à un stimulus donné
Source sonore	Ce terme se réfère à la fois à un émetteur de son, à sa position dans l'espace

et au son émis

Spatialisation

Ce terme représente l'association d'un stimulus sonore avec les propriétés spatiales d'azimut, d'élévation et de profondeur d'une position donnée dans l'espace

Étude de son 3D pour une interaction audio-visuelle en environnement virtuel enrichi

La plupart des applications qui intègrent le son 3D en environnement virtuel sont limitées à la simulation acoustique. L'objectif de cette thèse est d'étudier l'apport du son 3D dans l'interaction en environnement virtuel. Nous avons développé une technique de stimulation sonore appelée : Sensation Artificielle Spatialisée Auditiv (SASA), basée sur la création d'effets sonores. Cette technique restitue des sensations de spatialisation de son permettant la localisation précise (en azimut et en élévation) de sources sonores. Afin d'améliorer le temps de localisation de sources sonores, nous avons développé un modèle qui intègre la fonction HRTF (*Head-Related Transfer Function*) avec la technique SASA. Pour simuler la profondeur de sources sonores, nous avons développé une technique basée sur la combinaison entre l'indice d'intensité et les différences interaurales de temps. Les techniques développées dans le cadre de cette thèse ont été utilisées pour le guidage auditif afin de fournir des informations spatiales en azimut, en élévation et en profondeur. Nous avons appliqué cette technique dans la simulation chirurgicale pour assister l'opérateur dans la phase de triangulation. Nous avons validé cette approche en termes de performance de manipulation et du retour d'informations de l'utilisateur. Les résultats obtenus des travaux de cette thèse sont prometteurs pour l'utilisation de son 3D dans l'interaction en environnement virtuel.

Mots-clés : Son 3D, HRTF, Localisation de sources sonores, Interfaces sonores, Environnement Virtuel

Study of 3D sound for audio-visual interaction in enriched virtual environment

Most of the applications integrating 3D sound in virtual environment are limited to acoustic simulation. The objective of this thesis is to study the added value of 3D sound in the interaction with virtual environment. We developed an audio stimulation technique called "Artificial Spatial Audio Sensation: (ASAS)" based on the creation of audio effects. This technique conveys the sensations of spatialized sound that allows accurate sound source localization in azimuth and in elevation. In order to improve the localization time we also developed a model that integrates Head-Related Transfer Function (HRTF) with the ASAS technique. For the simulation of the depth of spatial sound sources, we developed a technique based on a combined model that integrates the intensity cue and the interaural time differences. The techniques developed in this thesis were applied for audio cueing in order to convey accurate spatial information in azimuth, elevation and depth. The combined model was applied in surgical simulation to assist the operator in the execution of the triangulation gesture. We conducted an experimental validation of this approach in terms of the criteria of manipulation performance and user's appreciation. The results obtained in this research are promising for the use of 3D sound to enrich the interaction in virtual environment.

Keywords: · 3D sound, HRTF, Sound Source Localization, Sound Interfaces, Virtual Environment

