THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

École doctorale : ISCE - Ingénierie pour la Santé la Cognition et l'Environnement
Spécialité : CIA - Ingénierie de la Cognition, de l'interaction, de l'Apprentissage et de la création
Unité de recherche : Laboratoire de Psychologie et Neuro Cognition

Conception et évaluation de métaphores sonores pour atteindre une cible sans la vue

Design and evaluation of sound metaphors to reach a target without sight.

Présentée par :

Coline FONS

Direction de thèse :

Christian GRAFF MAITRE DE CONFERENCES HDR, UNIVERSITE GRENOBLE ALPES	Directeur de thèse
Denis PELLERIN PROFESSEUR DES UNIVERSITES EMERITE, GRENOBLE INP - UGA	Co-directeur de thèse
Sylvain HUET MAITRE DE CONFERENCES, GRENOBLE INP - UGA	Co-encadrant de thèse

Rapporteurs :

BOB-ANTOINE MENELAS
FULL PROFESSOR, UNIVERSITE DU QUEBEC A CHICOUTIMI
KATIA ROVIRA
PROFESSEURE DES UNIVERSITES, UNIVERSITE DE ROUEN NORMANDIE

Thèse soutenue publiquement le **24 novembre 2025**, devant le jury composé de :

MICHEL GUERRAZ, PROFESSEUR DES UNIVERSITES, UNIVERSITE DE CHAMBERY	Président
CHRISTIAN GRAFF, MAITRE DE CONFERENCES HDR, UNIVERSITE GRENOBLE ALPES	Directeur de thèse
DENIS PELLERIN, PROFESSEUR DES UNIVERSITES EMERITE, GRENOBLE INP - UGA	Co-directeur de thèse
BOB-ANTOINE MENELAS, FULL PROFESSOR, UNIVERSITE DU QUEBEC A CHICOUTIMI	Rapporteur
KATIA ROVIRA, PROFESSEURE DES UNIVERSITES, UNIVERSITE DE ROUEN NORMANDIE	Rapporteure
TIFANIE BOUCHARA, MAITRESSE DE CONFERENCES, UNIVERSITE PARIS- SACLAY	Examinatrice

Invités :

SYLVAIN HUET
MAITRE DE CONFERENCES, GRENOBLE INP - UGA

Résumé

Les personnes déficientes visuelles rencontrent des difficultés dans les activités nécessitant un repérage spatial, qui ont des conséquences sociales et psychologiques importantes : dépendance accrue, fatigue cognitive liée à l'effort d'interprétation de l'environnement, isolement, etc. De nombreuses activités de loisirs et sportives, restent peu accessibles faute de dispositifs adaptés, ce qui souligne la nécessité de développer des outils favorisant l'autonomie et l'inclusion.

Dans ce contexte, des dispositifs de substitution sensorielle ont été développés afin de transformer l'information visuelle en information auditive ou tactile. Les premières approches visaient à restituer l'ensemble de la scène visuelle, mais étaient trop coûteuses cognitivement, et ne reproduisaient pas la vision, mais proposaient un nouveau mode de perception : on parle aujourd'hui de suppléance perceptive.

Plutôt que de maximiser la quantité d'informations, ces dispositifs doivent assister une tâche spécifique. Cette thèse s'intéresse au guidage vers une cible à l'aide de la sonification (l'utilisation de sons non verbaux pour transmettre des informations). Une métaphore sonore est la conversion d'un paramètre spatial (e.g. distance à la cible) en un paramètre sonore (e.g. intensité). Un enjeu central est de trouver un compromis entre précision et simplicité. Un guidage trop précis peut être difficile à comprendre, un guidage trop simple peut être inefficace.

L'objectif de cette thèse est de concevoir et évaluer des métaphores sonores de guidage afin d'identifier celles qui offrent les meilleures performances tout en respectant les capacités cognitives des utilisateurs. Nos travaux répondent à plusieurs limites de la littérature : le manque de comparaisons systématiques, la surcharge cognitive de certains dispositifs et l'absence de métriques d'évaluation standardisées.

Trois séries d'expériences ont été menées sous la forme d'un jeu virtuel, avec des participants voyants, puis des participants déficients visuels. Les participants devaient atteindre des cibles invisibles uniquement guidés par des sons, tandis que leur position était suivie par capture de mouvement. Les sonifications ont été comparées deux à deux, en allant des plus simples aux plus précises, et évaluées à partir des performances (vitesse et précision) et des retours subjectifs.

La première série, portant sur l'atteinte de cibles en 3D avec la main, montre qu'il est plus efficace d'encoder séparément chaque dimension spatiale que de convertir un écart directionnel global. Les performances augmentent encore quand les informations sont présentées de manière séquentielle. Ce guidage s'est même révélé plus rapide qu'un guidage vocal humain, qui adopte naturellement une stratégie séquentielle.

La deuxième série visait à proposer une métrique standardisée basée sur la loi de Fitts, reliant

le temps de mouvement à la taille et à la distance de la cible. Les résultats confirment l'intérêt de la présentation séquentielle et permettent de comparer nos travaux à ceux d'autres équipes.

La troisième série a appliqué ces métaphores à une tâche de tir en 2D. Comparées à la sonification traditionnelle du biathlon, les sonifications proposées ont permis des tirs plus rapides et plus précis. Bien que des différences de préférences apparaissent entre participants voyants et déficients visuels, la présentation séquentielle s'est révélée bénéfique pour tous.

Cette thèse montre qu'un guidage sonore peut être rendu plus efficace et plus ergonomique en dissociant les dimensions spatiales, à la fois dans leur encodage sonore et dans leur présentation temporelle. À terme, ces sonifications pourraient être intégrées à des applications mobiles pour faciliter des tâches quotidiennes, et permettre le développement de nouvelles pratiques de handisport. Plus largement, ces résultats contribuent à la conception de dispositifs de suppléance perceptive centrés sur la tâche et sur les capacités cognitives des utilisateurs.

Abstract

People with visual impairments face difficulties in activities that require spatial orientation, which have significant social and psychological consequences : increased dependence on others, cognitive fatigue due to the effort required to interpret the environment, social isolation, etc. Many leisure and sports activities remain poorly accessible due to the lack of adapted devices, highlighting the need to develop tools that promote autonomy and inclusion.

In this context, sensory substitution devices have been developed to transform visual information into auditory or tactile information. Early approaches aimed to reproduce the entire visual scene, but they proved to be too cognitively demanding and did not reproduce vision itself, instead offering a new mode of perception. This approach is now referred to as sensory substitution or perceptual supplementation.

Rather than maximizing the amount of information transmitted, such devices should assist a specific task. This thesis focuses on guiding a user toward a target using sonification, defined as the use of non-verbal sounds to convey information. A sound metaphor consists in converting a spatial parameter (e.g., distance to the target) into a sound parameter (e.g., intensity). A central challenge is to find a balance between precision and simplicity : overly precise guidance can be difficult to understand, while overly simple guidance can be ineffective.

The objective of this thesis is to design and evaluate sound metaphors for guidance in order to identify those that provide the best performance while respecting users' cognitive capacities. Our work addresses several limitations of the existing literature, including the lack of systematic comparisons, the cognitive overload induced by some devices, and the absence of standardized evaluation metrics.

Three series of experiments were conducted in the form of a virtual game, first with sighted participants and then with visually impaired participants. Participants were required to reach invisible targets guided solely by sounds, while their position was tracked using motion capture. Sonifications were compared pairwise, from the simplest to the most precise, and evaluated based on performance measures (speed and accuracy) and subjective feedback.

The first series, which focused on reaching 3D targets with the hand, shows that encoding each spatial dimension separately is more effective than converting a global directional error. Performance further improves when sound information is presented sequentially. This type of guidance was even faster than human verbal guidance, which naturally follows a sequential strategy.

The second series aimed to propose a standardized metric based on Fitts' law, which relates

movement time to target size and distance. The results confirm the benefits of sequential presentation and enable comparison with work conducted by other research groups. The third series applied these sound metaphors to a 2D shooting task. Compared with the traditional biathlon sonification, the proposed sonifications led to faster and more accurate shots. Although preference differences emerged between sighted and visually impaired participants, sequential presentation proved beneficial for all.

This thesis demonstrates that sound-based guidance can be made more effective and more ergonomic by dissociating spatial dimensions, both in their sound encoding and in their temporal presentation. In the long term, these sonifications could be integrated into mobile applications to facilitate everyday tasks and support the development of new adapted sports practices. More broadly, these results contribute to the design of perceptual supplementation devices that are task-oriented and aligned with users' cognitive capacities.

Remerciements

Je tiens à adresser mes premiers remerciements à mes encadrants, Christian, Sylvain et Denis, qui m'ont accordé leur confiance, leur accompagnement, leur bienveillance et leurs conseils pendant ces trois années de thèse. Je les remercie tout particulièrement pour l'impact positif qu'ont eu sur moi leurs encouragements et leurs félicitations pour mon travail dès mes premières semaines de stage de master avec eux, qui m'ont fait prendre confiance en moi et ont façonné la chercheuse que je suis aujourd'hui. Leurs retours positifs, aussi bien que leurs critiques constructives, m'ont été précieux et je m'estime chanceuse d'avoir eu un encadrement d'aussi bonne qualité, autant d'un point de vue scientifique qu'humain.

Je remercie Katia Rovira et Bob-Antoine Ménélas de m'avoir fait l'honneur d'évaluer mon travail de thèse en tant que rapporteurs, et Tifanie Bouchara et Michel Guerraz en tant qu'examinateurs. Je remercie également Alan Chauvin et Olivier Aycard qui ont accepté de faire partie de mon Comité de Suivi de Thèse et pour leurs retours constructifs.

Je remercie les personnes avec qui j'ai eu l'opportunité de collaborer dans le cadre de cette thèse, notamment Allan et Solange avec qui j'ai travaillé sur le guidage d'une personne par la voix dans le cadre du CDP-Boot, ainsi que Silvain Gerber qui a participé aux analyses statistiques. De même, je remercie tous les membres du projet Sam-Guide pour ces échanges au cours de ces trois années.

Je tiens à remercier toutes les personnes qui m'ont fait découvrir le tir sportif, notamment Hervé Sacquet qui nous a fait tester le système de tir à la carabine avec guidage sonore utilisé au biathlon paralympique, Marc Le Goff, Jean Grenet et Jean Ostian qui nous ont permis de faire tester notre propre système de tir à leurs étudiants lors de leurs cours de tir, et Mihaela Smerea qui nous a permis de le faire tester à des étudiants en IUT Carrières sociales lors d'un cours d'initiation au handisport.

J'en profite pour remercier tous ces étudiants qui se sont prêté au jeu, ainsi que toutes les autres personnes, que ce soit les étudiants en psychologie, les collègues, les personnels administratifs, la famille et les amis qui ont participé à nos études. Je remercie tout particulièrement les personnes déficientes visuelles qui se sont déplacées jusqu'au laboratoire pour venir tester notre dispositif et ont donné de leur temps non seulement pour l'expérience mais pour discuter avec moi et les autres membres du projet de leurs besoins et leurs envies. Je remercie l'association Valentin Haüy pour avoir diffusé notre annoncé et fait passer le mot à leurs membres.

Je remercie tous les stagiaires de licence et de master qui se sont joints à nous pour des durées plus ou moins longues, et ont tous fourni du travail de qualité, en particulier Sophie Diogo, Loui-

son Barbé, Eva Incorvaia et Chloé Héry que j'ai eu la chance de recruter et d'encadrer personnellement et qui ont participé aux passations expérimentales et aux analyses, mais également Julien Caffarel et Mialy Rakotonjatovo qui ont travaillé sur des aspects plus techniques, la conception du système de tir utilisable hors laboratoire.

Je remercie également les équipes administratives, ainsi que les équipes techniques des plateformes du LPNC et du GIPSA-Lab qui ont participé au bon déroulement de ma thèse.

Pour leur écoute, leur soutien, leur foi en ma capacité à réaliser ce que j'entreprends, leur bienveillance, leurs retours constructifs et leur esprit critique, je remercie mes sœurs Sarah et Marie, ma mère Isabelle et mes amies Emma et Emilie. Rien n'aurait été possible sans vous. Enfin, pour toutes ces raisons, pour avoir toujours cru en moi et pour être venu m'encourager à ma soutenance, je remercie mon grand-père, Jean. Papy, tu as fait mon bonheur.

Table des matières

1 Introduction **1**

1.1 Contexte et enjeux : le guidage sonore comme outil d'assistance et de loisir . . . 2

1.2 Limites dans la littérature : validation, complexité et standardisation des DSS . . 3

1.3 Questions de recherche : concevoir, évaluer et adapter le guidage sonore 3

1.4 Méthodologie : environnement virtuel et système de capture de mouvement . . 4

1.5 But des trois séries expérimentales 5

1.6 Trois tâches d'atteinte de cible pour trois séries expérimentales 6

1.6.1 Série expérimentale A : Optimisation et validation empirique de sonifications pour le guidage du corps et de la main vers des cibles 3D 6

1.6.2 Série expérimentale B : Validation empirique de sonifications avec la loi de Fitts pour le guidage de la main vers des cibles 3D 8

1.6.3 Série expérimentale C : Application et validation empirique des sonifications améliorées en 3D pour le guidage de la visée 2D d'une cible au pistolet 8

1.7 Contributions 9

1.8 Plan du manuscrit 9

2 État de l'art **11**

2.1 La substitution sensorielle 12

2.1.1 La substitution sensorielle vision-vers-toucher 12

2.1.2 La substitution sensorielle vision-vers-audition 13

2.1.3 Limites des dispositifs de substitution sensorielle 14

2.1.4 Synthèse 15

2.2 La perception auditive 15

2.2.1 Perception des paramètres sonores 15

2.2.2 Localisation d'une source sonore 16

2.2.3 Correspondances cross-modales . . . 18
2.2.4 Synthèse . . . 18
2.3 Dispositifs de substitution sensorielle appliqués au guidage vers une cible . . . 19
2.3.1 Perception . . . 19
2.3.2 Guidage . . . 20
2.3.3 Évaluation . . . 24
2.3.4 Synthèse . . . 25
2.4 Sonification . . . 26
2.4.1 Conditions pour une sonification de données 3D précise . . . 26
2.4.2 Défis liés à la charge cognitives . . . 27
2.4.3 Équilibre entre précision et simplicité . . . 27
2.4.4 Rôle des conventions culturelles et des correspondances incarnées . . . 27
2.4.5 Synthèse . . . 28
2.5 La loi de Fitts . . . 28
2.5.1 Première formulation de la loi de Fitts . . . 28
2.5.2 Autres formulations de la loi de Fitts . . . 30
2.5.3 La loi de Fitts adaptée à des tâches en 3D . . . 31
2.5.4 La loi de Fitts adaptée à l'atteinte de cible sans la vue . . . 31
2.5.5 Évaluer un système de guidage sonore en utilisant les paramètres de la loi de Fitts. . . . 33
2.5.6 Synthèse . . . 34

3 Série expérimentale A : Optimisation et validation empirique de sonifications pour le guidage du corps et de la main vers des cibles 3D 35

3.1 Introduction . . . 36
3.2 Étude A1 : Comparaison de sonifications explicites vs parcimonieuses pour l'atteinte d'une cible avec la main sur une tâche en 3D . . . 38
3.2.1 Introduction . . . 38
3.2.2 Matériel et Méthode . . . 39
3.2.3 Résultats . . . 45
3.2.4 Discussion . . . 47

3.2.5 Conclusion . . . 49
3.2.6 Résumé de l'étude A1 . . . 49
3.3 Étude A2 : Comparaison de métaphores sonores pour l'atteinte d'une cible avec la main sur une dimension. . . . 51
3.3.1 Introduction . . . 51
3.3.2 Matériel et Méthodes . . . 54
3.3.3 Résultats . . . 58
3.3.4 Discussion . . . 63
3.3.5 Conclusion . . . 65
3.3.6 Résumé de l'étude A2 . . . 65
3.4 Étude A3 : Comparaison d'un mode de sonification séquentiel à un mode de sonification parallèle et à un guidage vocal sur une tâche d'atteinte de cible 3D avec la main . . . 67
3.4.1 Introduction . . . 67
3.4.2 Matériel et Méthodes . . . 69
3.4.3 Résultats . . . 76
3.4.4 Discussion . . . 81
3.4.5 Conclusion . . . 83
3.4.6 Résumé de l'étude A3 . . . 84

4 Série expérimentale B : Validation empirique de sonifications avec la loi de Fitts pour le guidage de la main vers des cibles 3D 87

4.1 Introduction . . . 88
4.2 Étude B1 : Comparaison de sonifications parcimonieuses et explicites avec la loi de Fitts, sur une tâche d'atteinte de cible 3D avec la main . . . 90
4.2.1 Introduction . . . 90
4.2.2 Matériel et Méthodes . . . 92
4.2.3 Résultats . . . 97
4.2.4 Discussion . . . 99
4.2.5 Conclusion . . . 101
4.2.6 Résumé de l'étude B1 . . . 102

4.3 Étude B2 : Comparaison d'un mode de sonification séquentiel vs parallèle avec la loi de Fitts, sur une tâche d'atteinte de cible 3D avec la main simplifiée 103

4.3.1 Introduction 103

4.3.2 Matériel et Méthodes 106

4.3.3 Résultats 108

4.3.4 Discussion 109

4.3.5 Conclusion 110

4.3.6 Résumé de l'étude B2 110

5 Série expérimentale C : Application et validation empirique des sonifications améliorées en 3D pour le guidage de la visée 2D d'une cible au pistolet 115

5.1 Introduction 116

5.2 Etude C1 : Comparaison de sonifications pour l'atteinte d'une cible 2D au tir au pistolet, sans contrainte de temps 119

5.2.1 Introduction 119

5.2.2 Matériel et Méthode 120

5.2.3 Résultats 124

5.2.4 Discussion 127

5.2.5 Conclusion 128

5.2.6 Résumé de l'étude C1 129

5.3 Etude C2 : Comparaison de sonifications pour l'atteinte d'une cible 2D au tir au pistolet, avec contrainte de temps 131

5.3.1 Introduction 131

5.3.2 Matériel et Méthode 132

5.3.3 Analyses 134

5.3.4 Résultats 137

5.3.5 Discussion 147

5.3.6 Conclusion 149

5.3.7 Résumé de l'étude C2 150

6 Discussion générale 153

6.1 Rappel du contexte et des principaux résultats 154
6.2 Mise en perspective par rapport à l'état de l'art 155
6.2.1 Dissociation des dimensions 155
6.2.2 Présentation séquentielle des dimensions 156
6.2.3 Simplification des métaphores sonores 158
6.2.4 Évaluation des métaphores 159
6.3 Recrutement de participants déficients visuels 160
6.4 Perspectives 161
6.4.1 Perspectives théoriques 161
6.4.2 Perspectives pratiques 162
6.5 Conclusion 164

Bibliographie **165**

7 Annexes **177**

7.1 Annexe 1 : Questionnaire sur l'ergonomie 177
7.2 Annexe 2 : Liste des tableaux 180
7.3 Annexe 3 : Liste des figures 183
7.4 Annexe 4 : Liste des abréviations 191
7.5 Annexe 5 : Publications 192
7.5.1 Conférences 192
7.5.2 Revue 192
7.5.3 Soumis à PLOS One 192

1. Introduction

Sommaire

1.1 Contexte et enjeux : le guidage sonore comme outil d'assistance et de loisir . . . 2

1.2 Limites dans la littérature : validation, complexité et standardisation des DSS . . . 3

1.3 Questions de recherche : concevoir, évaluer et adapter le guidage sonore 3

1.4 Méthodologie : environnement virtuel et système de capture de mouvement . . . 4

1.5 But des trois séries expérimentales . . . 5

1.6 Trois tâches d'atteinte de cible pour trois séries expérimentales . . . 6

1.6.1 Série expérimentale A : Optimisation et validation empirique de sonifications pour le guidage du corps et de la main vers des cibles 3D . . . 6

1.6.2 Série expérimentale B : Validation empirique de sonifications avec la loi de Fitts pour le guidage de la main vers des cibles 3D . . . 8

1.6.3 Série expérimentale C : Application et validation empirique des sonifications améliorées en 3D pour le guidage de la visée 2D d'une cible au pistolet . . . 8

1.7 Contributions . . . 9

1.8 Plan du manuscrit . . . 9

1.1 Contexte et enjeux : le guidage sonore comme outil d'assistance et de loisir

Les personnes déficientes visuelles rencontrent des défis majeurs dans les tâches spatiales du quotidien, qu'il s'agisse de se déplacer dans un environnement inconnu, de manipuler des objets ou de participer à des activités de loisir et sportives. Ces difficultés dépassent les simples obstacles physiques et incluent également des barrières sociales et psychologiques, telles que la dépendance accrue à autrui, la fatigue cognitive liée au traitement de l'information et l'isolement social (Burton et al. 2021). Se déplacer dans des espaces inconnus ou encombrés représente un risque important de chute et nécessite souvent l'usage d'aides à la mobilité, comme la canne blanche ou le chien guide. De nombreuses activités de loisirs deviennent difficiles ou inaccessibles, et la pratique des sports traditionnels est souvent limitée par l'absence d'activités adaptées, des expériences négatives passées ou la crainte d'être perçu comme une charge (Hillan et al. 2023). Ces constats soulignent l'importance de développer des dispositifs d'assistance innovants pour favoriser l'autonomie et l'inclusion des personnes déficientes visuelles, que ce soit dans les tâches du quotidien ou dans des contextes ludiques et sportifs.

Pour favoriser l'autonomie, des dispositifs de substitution sensorielle (DSS) se développent depuis quelques décennies. Ceux-ci convertissent des informations normalement accessibles par la vue, en informations accessibles dans une autre modalité sensorielle, en général l'audition ou le toucher. Quand les données sont converties en sons, il s'agit de sonification, qui offre une capacité de transmission d'information beaucoup plus grande que le toucher, plus limité en nuances perçues. Alors que les approches cherchant à remplacer intégralement la vue ont montré leurs limites, les DSS qui se révèlent efficaces sont ceux qui se concentrent sur des tâches ciblées (lecture de texte, navigation, évitement d'obstacles). Notre travail s'inscrit dans cette perspective : nous étudions une tâche précise, l'atteinte de cible guidée par le son.

Cette thèse propose une contribution méthodologique et expérimentale : concevoir et évaluer des métaphores sonores pour atteindre une cible sans la vue. Les résultats obtenus sont pertinents pour des applications variées de la vie quotidienne, comme localiser un produit dans un rayon ou trouver un interphone.

Au delà de ce cadre général, la thèse est financée par le projet ANR SAM-Guide, dont l'un des objectifs est de rendre accessible le laser-run, une épreuve du pentathlon moderne combinant course et tir au pistolet laser, aux personnes privées de vue. Les sonifications développées et validées expérimentalement sur des tâches d'atteinte de cible 3D dans ce travail de thèse ont ainsi été adaptées et testées dans ce contexte de tir au pistolet laser, en comparaison avec la solution déjà utilisée dans une discipline voisine, le biathlon.

1.2 Limites dans la littérature : validation, complexité et standardisation des DSS

L'état de l'art révèle plusieurs défis majeurs pour les dispositifs de substitution sensorielle (DSS) :

1. Le manque de validation empirique systématique des sonifications : la plupart des DSS proposent des métaphores sonores ad hoc (e.g. conversion directe d'angles en fréquence) sans comparer rigoureusement leur efficacité relative ni leur intuitivité. Peu d'études testent différentes alternatives pour identifier les correspondances optimales entre dimensions spatiales et paramètres auditifs pour la tâche en jeu.
2. La surcharge cognitive induite par des systèmes trop complexes. Les travaux existants montrent que maximiser la quantité d'informations, comme dans the vOICe, nuit à la performance, tandis qu'une optimisation des paramètres sonores via la prise en compte des correspondances cross-modales (ex. hauteur sonore et hauteur spatiale) améliore l'efficacité et l'intuitivité (Vickers et al. 2014).
3. L'absence de standardisation de l'évaluation des dispositifs, qui rend difficile de comparer différents dispositifs dans la littérature. La loi de Fitts, qui permet de prédire le temps mis pour toucher une cible en fonction de la difficulté de la tâche, pourrait être une métrique standardisée permettant la comparaison, mais son application à des systèmes de guidage par sonification non spatialisée n'a pas encore été explorée.

1.3 Questions de recherche : concevoir, évaluer et adapter le guidage sonore

L'objectif de cette thèse est de concevoir et évaluer différentes méthodes de guidage sonore vers une cible pour déterminer lesquelles sont les plus efficaces pour une tâche donnée, tout en tenant compte des limites des capacités perceptives, des capacités d'interprétation et des capacités d'attention de l'utilisateur.

Dans ce cadre, nous posons la question suivante : Comment concevoir et évaluer des métaphores sonores pour atteindre une cible sans la vue ?

Trois sous-questions découlent de cette question générale :

1. Quelles métaphores sonores sont les plus efficaces et intuitives pour coder les trois dimensions spatiales sans dépasser les limites cognitives des utilisateurs ?
2. Comment évaluer rigoureusement ces DSS, en dépassant les biais méthodologiques identifiés dans la littérature ?
3. Les métaphores optimisées pour l'atteinte de cible 3D avec la main sont-elles transposables à la visée au pistolet, où la précision et la rapidité sont critiques ?

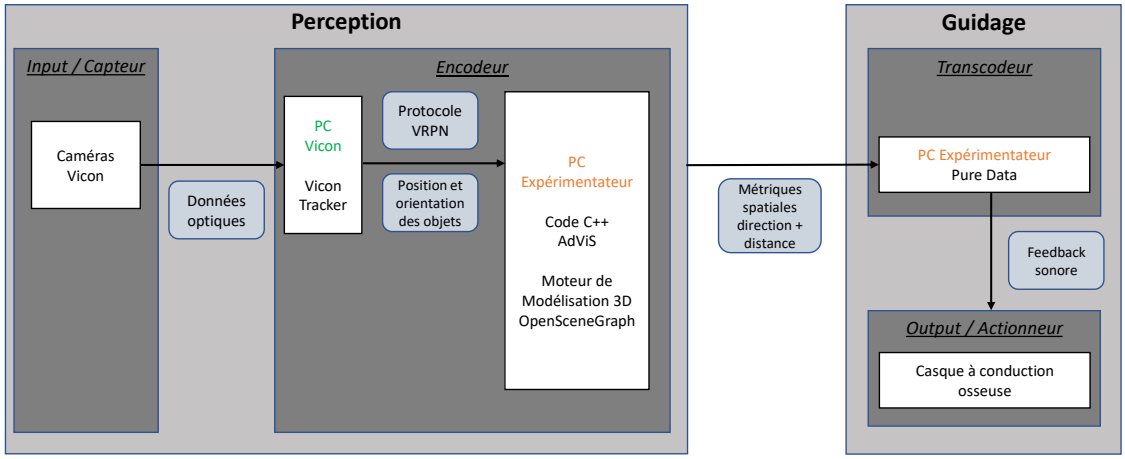


Figure 1.1 – Dispositif expérimental utilisé dans toutes les études présentées dans cette thèse.

1.4 Méthodologie : environnement virtuel et système de capture de mouvement

Pour répondre aux questions posées, cette thèse s'organise en trois séries expérimentales, qui partagent un protocole expérimental commun mis en œuvre dans un espace de capture de mouvement. Toutes les études présentées utilisent le même dispositif expérimental (Figure 1.1) où les participants, immergés dans un environnement virtuel, doivent atteindre une cible virtuelle. Le système de capture de mouvement suit en temps réel soit des points anatomiques du corps de l'utilisateur (tâche d'atteinte de cible avec la main), soit la position du pistolet factice (tâche de visée de cible au pistolet). Les coordonnées de ces points sont transmises via un protocole VRPN à un logiciel C++ central qui gère trois fonctions clés : il restitue l'environnement 3D via OpenScene Graph (OSG), calcule les métriques spatiales à convertir en sons, et pilote le protocole expérimental. Ces métriques sont converties en sons par Pure Data et transmises au participant via un casque à conduction osseuse connecté sans fil à un boîtier récepteur.

L'utilisation d'un environnement virtuel nous permet de contrôler les paramètres expérimentaux comme la forme, la taille et la position des cibles. Grâce au système de capture de mouvement, qui repère les objets portés par le participant, le corps de celui-ci peut être utilisé comme référentiel pour mesurer les écarts à la cible, plutôt que d'utiliser un référentiel externe comme une caméra portée en main ou sur la tête. Ce dispositif permet ainsi de localiser avec précision les points anatomiques ou le pistolet factice, de calculer finement les écarts à la cible et de déterminer avec précision le moment où celle-ci est atteinte.

Dans nos trois séries d'expériences, nous menons des comparaisons systématiques entre métaphores sonores, sur trois tâches d'atteinte de cibles différentes (que nous décrirons en section 1.6), afin d'identifier les plus efficaces et intuitives et optimiser le guidage vers une cible. Les différents modes de guidage sont à chaque fois évalués et comparés sur la base de métriques objectives (temps d'atteinte de cible) et subjectives (scores d'ergonomie donnés par les participants).

Les spécificités de chaque série expérimentale sont données dans la section suivante.

1.5 But des trois séries expérimentales

Les séries expérimentales A, B et C permettent de répondre respectivement aux sous-questions 1, 2 et 3 présentées dans la sous-section 1.3 :

1. **Série A : Optimisation et validation empirique de sonifications pour le guidage du corps et de la main vers des cibles 3D.** Dans cette série, nous optimisons les sonifications de guidage en testant des combinaisons de paramètres auditifs (fréquence, intensité, stéréo, guidage vocal, etc.) et des modes de présentation des métaphores sonores (séquentiel vs. parallèle) (Études A1, A2 et A3, Chapitre 3). Cette série d'études s'appuie sur les principes de discrimination auditive fine des paramètres sonores pour assurer la précision du guidage, et de correspondances cross-modales pour assurer l'intuitivité du guidage.
2. **Série B : Validation empirique de sonifications avec la loi de Fitts pour le guidage de la main vers des cibles 3D.** Dans cette série, nous vérifions que la loi de Fitts est applicable et donc qu'il est possible de l'utiliser pour comparer des sonifications. Ensuite, nous appliquons la loi de Fitts comme métrique standardisée pour évaluer l'efficacité des sonifications, en comparant leur *throughput* (TP) (Études B1 et B2, Chapitre 4). Cet axe répond au besoin de standardisation en reprenant la loi de Fitts, déjà validée pour les tâches visuelles mais inédite en sonification non spatialisée.
3. **Série C : Application et validation des sonifications optimisées en 3D pour le guidage de la visée 2D d'une cible au pistolet.** Dans cette série, nous adaptons les sonifications optimisées dans les deux premières séries au contexte sportif du laser-run, sur une tâche de tir au pistolet (Études C1 et C2, Chapitre 5). Cet axe poursuit l'idée de simplification du guidage pour réduire la charge cognitive, tout en l'adaptant à un contexte exigeant en rapidité et en précision (laser-run).

Les trois séries expérimentales permettre de répondre à ces deux limites de la littérature identifiées :

(1) les sonifications sont rarement optimisées par des comparaisons empiriques systématiques : dans toutes les études de ces trois séries, nous comparons systématiquement plusieurs modes de guidage sonore afin d'identifier les plus efficaces et intuitifs.

(2) les systèmes sont parfois trop complexes ce qui peut créer une surcharge cognitive et nuire aux performances : nous partons du codage le plus simple (information binaire de bonne direction) que nous enrichissons au fur et à mesure (ajout d'un codage de l'écart directionnel, séparation des dimensions) en testant à chaque fois l'intérêt de cet ajout d'information. Nous testons différentes métaphores sonores que nous avons choisies pour leur simplicité (e.g. codage binaire plutôt que continu), leur intuitivité (e.g. utilisation de la stéréo qui fait appel aux capacités

naturelles du système auditif à localiser un son, utilisation de sons verbaux), ou leur lien avec des associations cross-modales (e.g. une fréquence montante = il faut monter la main).

Dans la série B, nous nous intéressons à la dernière limite de la littérature que nous avons mentionnée :

(3) l'absence de standardisation de l'évaluation des dispositifs, qui rend difficile la comparaison entre études : nous proposons d'utiliser une métrique tirée de la loi de Fitts, le *throughput* (TP) pour évaluer un dispositif de guidage sonore. Le TP tient compte de paramètres expérimentaux qui peuvent varier entre les études, à savoir la taille des cibles et la distance entre le doigt et la cible au début de l'essai. Ainsi, le TP peut permettre la comparaison entre études, à condition de fournir l'équation complète de la loi obtenue dans chaque condition de guidage qu'on veut tester.

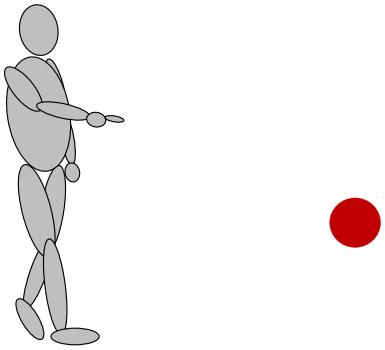
1.6 Trois tâches d'atteinte de cible pour trois séries expérimentales

A chacune de nos séries expérimentales est associé un type de tâche particulier (voir Figure 1.2).

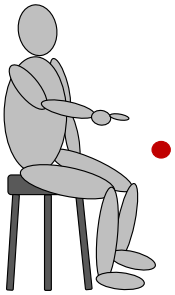
1.6.1 Série expérimentale A : Optimisation et validation empirique de sonifications pour le guidage du corps et de la main vers des cibles 3D

La série A (chapitre 3) vise à optimiser les métaphores sonores pour l'atteinte de cible 3D. Pour cela, nous avons créé une tâche d'atteinte de cible avec la main générique où les participants doivent marcher et atteindre de la main des cibles virtuelles qui apparaissent l'une après l'autre dans un espace virtuel (voir Figure 1.2(a)).

Nous voulons déterminer quelle quantité d'information il est nécessaire de transmettre à l'utilisateur pour à la fois maximiser la performance tout en minimisant la charge cognitive. Pour ça, nous procédons donc en comparant les modes de sonification par paire. Après chaque comparaison, nous gardons la sonification la plus efficace pour la prochaine comparaison avec une version que nous estimons améliorée. Dans l'étude A1 (section 3.2), nous partons de la sonification la plus simple, un signal binaire indiquant la bonne direction. Nous la comparons à une sonification ajoutant une information continue de déviation angulaire par rapport à la cible. Enfin, nous augmentons la précision en utilisant deux paramètres sonores pour dissocier les écarts verticaux et horizontaux, tout en réduisant la charge cognitive en utilisant un signal binaire pour l'un des écarts. Dans l'étude A2 (section 3.3), nous menons des tests sur une dimension pour trouver le meilleur codage de chaque dimension. Enfin dans l'étude A3 (section 3.4), nous comparons une sonification dissociant les trois dimensions en trois métaphores sonores, présentées simultanément (mode parallèle), à une sonification utilisant les mêmes métaphores sonores, mais présentées une par une (mode séquentiel). Les métaphores sonores utilisées pour les deux modes



(a) *Série expérimentale A.* Tâche d'atteinte de cible 3D an deux phases : marche puis atteinte avec la main. La taille de la cible est fixe. La position de la cible varie entre les essais.



(b) *Série expérimentale B.* Tâche d'atteinte de cible 3D simplifiée : atteinte avec la main. La taille de la cible et la position de la cible varient entre les essais.



(c) *Série expérimentale C.* Tâche d'atteinte de cible 2D : visée de cible au pistolet. La taille de la cible et sa position sont fixes.

Figure 1.2 – Illustration des trois tâches d'atteinte de cible guidée par le son correspondant aux trois séries expérimentales. Les cibles sont représentées en rouge.

de sonification sont celles testées dans l'étude A2 ayant obtenu les meilleures performances.

1.6.2 Série expérimentale B : Validation empirique de sonifications avec la loi de Fitts pour le guidage de la main vers des cibles 3D

La série B a deux objectifs : (1) proposer et tester une métrique standardisée, basée sur la loi de Fitts pour faciliter la comparaison avec d'autres systèmes de guidage de la littérature ; (2) répliquer les résultats de la série précédente et vérifier la robustesse des différences observées entre les modes de guidage. Dans les deux études de cette série expérimentale, nous utilisons une tâche d'atteinte de cible virtuelle en 3D similaire à celle du chapitre précédent, mais simplifiée. En effet ici les cibles sont à portées de main pour que le geste d'atteinte se rapproche de celui du pointage (voir Figure 1.2(b)). La phase d'atteinte avec la main étant la plus difficile pour les participants (comme la série A nous l'a fait remarquer), elle est la plus pertinente pour effectuer des comparaisons. Ici, la taille des cibles n'est plus fixe mais varie, afin de pouvoir calculer plusieurs niveaux de difficulté (Indice de Difficulté de Fitts). Les variations de taille de cible et de distance à la cible permettent d'utiliser la loi de Fitts et de calculer une métrique, le TP, qui permettra la comparaison avec d'autres études de la littérature.

Dans ce chapitre, nous menons deux comparaisons, qui reprennent celles utilisées dans le chapitre précédent afin de répliquer les résultats obtenus à la fois avec le temps d'atteinte de cible et le TP de Fitts. Ainsi, l'étude B1 (section 4.2) reprend la comparaison entre une sonification qui utilise une seule métrique pour coder une direction globale et une sonification qui dissocie les dimensions verticale et horizontale entre deux paramètres sonores. L'étude B2 (section 4.3) compare une sonification qui présente les écarts sur les trois dimensions l'un après l'autre (mode séquentiel) à une sonification qui les présente simultanément (mode parallèle).

1.6.3 Série expérimentale C : Application et validation empirique des sonifications améliorées en 3D pour le guidage de la visée 2D d'une cible au pistolet

La série C s'intéresse à l'application du guidage par métaphores sonores testé sur des tâches d'atteinte avec la main en 3D, à une tâche d'atteinte de cible en 2D, le tir au pistolet. Ainsi, toujours dans le même espace de capture de mouvement, les participants doivent viser et "tirer" avec un pistolet factice équipé d'une gâchette sur une cible virtuelle fixe (voir Figure 1.2). L'utilisation du système de capture de mouvement et d'une cible virtuelle nous permet de reproduire les caractéristiques géométriques du pas de tir utilisé en laser-run. Dans l'étude C1 (section 5.2) du chapitre 5, le protocole est similaire à celui des chapitres précédents, le participant compare deux modes de sonification et doit tirer un certains nombres de fois dans chaque condition. Dans l'étude C2 (section 5.3), nous nous rapprochons du scénario de tir du laser-run en introduisant une contrainte temporelle en plus d'une contrainte de précision. Les participants ont un nombre illimité de tir pour atteindre la cible un certain nombre de fois en un temps limité. Les modes de

sonification testés ici, par paires, sont directement inspirés de ceux des chapitres précédents, afin de vérifier si les découvertes faites sur les deux tâches 3D se retrouvent sur une tâche de visée 2D. De plus, nous les comparons au mode de guidage sonore utilisé par les athlètes déficients visuels dans une épreuve comparable, le tir à la carabine (épreuve de biathlon).

1.7 Contributions

Les contributions de cette thèse sont :

- Elle propose des sonifications optimisées, validées empiriquement, et adaptées aux capacités perceptives et cognitives des utilisateurs.
- Elle démontre l'utilité du *throughput* (TP), une métrique issue de la loi de Fitts, pour l'évaluation des DSS, ouvrant la voie à des comparaisons entre systèmes normalisées.
- Le projet ANR SAM-Guide qui a financé ces travaux de thèse se terminera par l'organisation d'une épreuve de laser-run pour déficient visuel. Le laser-run est une épreuve du pentatlon moderne qui combine course et tir au pistolet laser. C'est un sport récent, qui n'a pas encore été adapté aux athlètes déficients visuels. Les travaux menés dans cette thèse pourront donc avoir des applications sur la partie tir de cet objectif ludique.

Les principes explorés ici dans un contexte de jeu préparent des applications plus larges, car les tâches utilisées sont génériques. En effet, les métaphores sonores optimisées pourraient être appliquées au guidage de la main dans certaines situations concrètes, par exemple pour aider une personne malvoyante à localiser le bon bouton d'interphone, un produit dans un rayon de supermarché ou encore l'entrée d'une boutique lorsque le GPS l'a conduite à proximité. Combinées aux progrès techniques récents, notamment en vision par ordinateur sur smartphone, ces métaphores sonores pourraient être utilisées grâce à une simple application sur smartphone qui localiserait la cible visée grâce à l'image captée par la caméra du téléphone.

1.8 Plan du manuscrit

Le document est organisé comme suit : après un état de l'art approfondi sur les DSS et la perception auditive (chapitre 2), les chapitres 3, 4 et 5 détaillent nos trois séries expérimentales, tandis que le chapitre 6 propose une discussion générale des résultats et des perspectives pour de futures recherches.

Les listes des figures, des tableaux et des abréviations sont disponibles en annexes 7.3, 7.2 et 7.4 respectivement. La liste des publications liées à cette thèse est disponible en annexe 7.5.

Les trois chapitres expérimentaux comportent volontairement des répétitions sur la problématique, les questions de recherche et les limites de la littérature. Ces répétitions ont pour but de permettre la lecture de chaque chapitre de manière indépendante.

2. État de l'art

Sommaire

2.1 La substitution sensorielle 12
2.1.1 La substitution sensorielle vision-vers-toucher 12
2.1.2 La substitution sensorielle vision-vers-audition 13
2.1.3 Limites des dispositifs de substitution sensorielle 14
2.1.4 Synthèse 15
2.2 La perception auditive 15
2.2.1 Perception des paramètres sonores 15
2.2.2 Localisation d'une source sonore 16
2.2.3 Correspondances cross-modales 18
2.2.4 Synthèse 18
2.3 Dispositifs de substitution sensorielle appliqués au guidage vers une cible 19
2.3.1 Perception 19
2.3.2 Guidage 20
2.3.3 Évaluation 24
2.3.4 Synthèse 25
2.4 Sonification 26
2.4.1 Conditions pour une sonification de données 3D précise 26
2.4.2 Défis liés à la charge cognitives 27
2.4.3 Équilibre entre précision et simplicité 27
2.4.4 Rôle des conventions culturelles et des correspondances incarnées . . 27
2.4.5 Synthèse 28
2.5 La loi de Fitts 28
2.5.1 Première formulation de la loi de Fitts 28
2.5.2 Autres formulations de la loi de Fitts 30
2.5.3 La loi de Fitts adaptée à des tâches en 3D 31
2.5.4 La loi de Fitts adaptée à l'atteinte de cible sans la vue 31
2.5.5 Évaluer un système de guidage sonore en utilisant les paramètres de la loi de Fitts. 33
2.5.6 Synthèse 34

Ce chapitre propose un état de l'art sur les DSS et la perception auditive. Nous commençons par introduire les premiers pas de la substitution sensorielle et les limites des premiers DSS (section 2.1). Nous expliquons ensuite le fonctionnement du système auditif (section 2.2). Les sections suivantes passent en revue les DSS développés pour le guidage vers une cible par le son (section 2.3) et décrivent le principe de la sonification (section 2.4). Nous finissons par exposer le principe de la loi de Fitts et son potentiel comme méthode d'évaluation des DSS (section 2.5).

2.1 La substitution sensorielle

La substitution sensorielle a été conçue comme une technologie d'assistance permettant aux personnes souffrant de déficiences d'un sens d'accéder aux informations normalement accessibles par ce sens via une autre modalité sensorielle (Paul Bach-y-Rita, Collins et al. 1969). Lorsque la vision est déficiente, l'idée est d'utiliser un dispositif de substitution sensorielle (DSS) pour transformer un signal vidéo d'une caméra à l'aide d'un algorithme et le transmettre via un ou plusieurs systèmes sensoriels fonctionnels : auditif (Auvray et al. 2005 ; Hanneton, Auvray et Durette 2010 ; P. B. Meijer 1992) et/ou tactile (Paul Bach-y-Rita et Kercel 2003 ; Tyler, Danilov et Paul Bach-y-Rita 2003). Cette approche suppose qu'en raison de la neuroplasticité, après un entraînement, le cerveau est capable d'apprendre à percevoir et à naviguer dans le monde des stimuli physiques de manière plus efficace, et par conséquent, de compenser partiellement les fonctions du système sensoriel visuel (Paul Bach-y-Rita 1987, 2003). Lors du développement de ces méthodes de sonification d'images dans les DSS visuo-auditifs, une question fondamentale est de savoir quelle quantité d'information doit être convertie en sons, et si ces informations doivent être simplement transmises ou prétraitées de manière significative lors de la conversion (Brown, Simpson et Proulx 2014).

2.1.1 La substitution sensorielle vision-vers-toucher

Les premières études sur la substitution sensorielle ont été réalisées dans les années 1960. Paul Bach-y-Rita est connu comme le pionnier des systèmes de substitution visuo-tactile. Ses premiers dispositifs de substitution sensorielle, communément appelés systèmes de substitution tactile-vision ou *Tactile-Vision Substitution Systems* (TVSS) (Paul Bach-y-Rita 1967 ; Paul Bach-y-Rita, Collins et al. 1969 ; White et al. 1970), visaient à utiliser la plasticité cérébrale des personnes aveugles pour restaurer leurs capacités visuelles sur la base d'une entrée sensorielle tactile. L'un des premiers TVSS consistait en une caméra (capteur) montée sur un trépied et manipulée par le participant, qui pouvait ainsi balayer les objets devant lui. Au moyen d'un commutateur (transcodeur), une matrice de stimulateurs vibrants (effecteurs) montée sur une chaise projetait l'image capturée sur le dos des participants. Bach-y-Rita a travaillé sur la transmission de l'information à plusieurs parties du corps, dont l'abdomen, le dos, la cuisse, le front, la langue et le bout des doigts (Bach-y-Rita 1989 ; Paul Bach-y-Rita 1972, 1995, 2004 ; Collins et Paul Bach-y-Rita 1973).

2.1.2 La substitution sensorielle vision-vers-audition

Butorova (Butorova et Sergeev 2025) classe les dispositifs de substitution audio-visuel (ou AVSS, Audio-Vision Substitution System) en trois catégorie. Le dispositif the vOICe (P. B. Meijer 1992)) entre dans la première catégorie, qui utilise la *sonification colonne par colonne*. The vOICe est le premier AVSS. Ce dispositif convertit une image capturée par une caméra en un paysage sonore. Chaque paysage sonore est un balayage de l'image en colonne, de gauche à droite. Dans une colonne, la fréquence représente l'axe vertical : un pixel situé à une position plus élevée sera converti en une fréquence plus élevée. Son intensité sonore est fonction de sa luminosité : un pixel avec une luminosité plus élevée sera converti en intensité plus élevée. On retrouve dans cette catégorie le dispositif de Martinez (Martínez et al. 2011) qui, comme the vOICe, sonifie chaque pixel de l'image en codant la luminosité du pixel par l'intensité sonore, et la position verticale du pixel par la fréquence sonore. Cet algorithme est basé sur une approche symétrique, dans laquelle les oreilles gauche et droite reçoivent les signaux auditifs obtenus à partir de la moitié gauche et de la moitié droite de l'image respectivement. Dans la même catégorie, l'algorithme EyeMusic (Abboud et al. 2014) traite l'image colonne par colonne, de gauche à droite. Pour chaque pixel de la colonne, un instrument de musique indique la couleur, la fréquence varie en fonction de l'élévation du pixel, et le volume sonore en fonction de la luminosité. Enfin, avec l'algorithme Topo-Speech (Heimler et al. 2019 ; Netzer et al. 2021), la sonification de l'image colonne par colonne est complétée par des infirmations vocales sur l'identité des objets.

La deuxième catégorie d'AVSS encode l'ensemble de l'image en un signal auditif complexe (*sonification de l'image entière*). Le PVSA (Capelle et al. 1998) ne balaye pas le champ visuel mais utilise une transformation de fréquence acoustique pixel par pixel, où la position du pixel sur l'axe vertical est codée en fréquence, la position horizontale en intensité binaurale et en différence de phase, et la luminosité est codée en amplitude du signal auditif. De même, the Vibe (Auvray et al. 2005 ; Hanneton, Auvray et Durette 2010) ne balaie pas le champ visuel, mais le divise en plusieurs champs présentés simultanément, leur position étant codée par l'équilibre des fréquences des canaux audio gauche-droite.

Enfin, le troisième groupe d'algorithmes utilise la profondeur comme principale donnée visuelle (*sonification 3D*). MeloSee (Fristot et al. 2012 ; Stoll et al. 2015) reçoit une carte de profondeur en entrée et renvoie une mélodie correspondante en sortie. Les auteurs ont introduit des champs récepteurs pour sous-échantillonner les images d'entrée. Plus l'objet est proche, plus l'amplitude du signal auditif est élevée, et la fréquence sonore représente la verticalité. La position horizontale détermine le son stéréo appliqué à la source sonore binaurale. Les sons de tous les champs réceptifs sont joués simultanément. Le dispositif Synaestheatre (Hamilton-Fletcher, Mengucci et Medeiros 2016 ; Hamilton-Fletcher, Obrist et al. 2016 ; Richardson et al. 2019) transforme les informations spatiales et colorimétriques en 3D en sons esthétiques qui varient en temps réel. En particulier, une grille de 13 x 7 points de profondeur est sélectionnée par une caméra, avec des informations de couleur dans chacun des 91 emplacements spatiaux classés dans l'une des sept catégories de couleurs (noir, gris, blanc, rouge, vert, bleu et jaune). Chacun de ces

résultats potentiels est associé à un son qui est joué lorsqu'un objet est présent à cet endroit (une forme de spatialisation). La position horizontale de chaque pixel donne au son les différences de timing et d'intensité interaurales qui lui sont associées. La position verticale donne la hauteur du son et la couleur sa qualité timbrale. La distance par rapport à la caméra indique l'intensité sonore de chaque son, les objets situés au-delà de la distance maximale fixée étant silencieux.

2.1.3 Limites des dispositifs de substitution sensorielle

Les dispositifs de substitution sensorielle présentés ci-dessus partagent la même ambition de chercher à remplacer l'œil, et donc de fournir aux personnes aveugles et malvoyantes toutes les expériences cognitives et comportementales qui accompagnent la vision. Cependant, bien que certains utilisateurs soient capables de rapporter une phénoménologie visuelle détaillée après plusieurs mois d'utilisation intensive du dispositif, la difficulté d'apprentissage de ces systèmes semble être un frein à leur utilisation. Selon Bach-y-Rita, l'expérience subjective avec le TVSS est comparable à celle de la vision (Paul Bach-y-Rita 2004). Malgré cette affirmation, l'apprentissage du système est long et il est utilisé pour des tâches simples telles que la reconnaissance de lettres ou d'objets simples, ou la reconnaissance de la direction du mouvement d'un objet. Le système testé par des aveugles qui l'avaient appris pendant plus de 40 heures a montré qu'il fallait environ quinze minutes pour reconnaître un nouvel objet (White et al. 1970).

Dans leur présentation des limites de la substitution sensorielle, Lenay et al. 2000 proposent d'utiliser le terme de "suppléance perceptive" plutôt que substitution sensorielle. D'une part, la substitution n'est pas sensorielle, mais plutôt sensori-motrice. Les travaux de Bach-y-Rita montrent en effet que le sujet aveugle ne peut reconnaître des formes statiques sur la matrice tactile. Il faut qu'il puisse manipuler la caméra et apprendre comment les variations de ses sensations sont liées à ses actions. D'autre part, ce n'est pas réellement une substitution, un constat fait par les personnes malvoyantes qui étaient déçues ou déprimées quand elles commençaient à utiliser ce type de dispositif, car elles espéraient faire l'expérience du monde visible et découvrir une dimension de l'existence qui leur était inconnue. A la place, ce genre de dispositif permet de réaliser des tâches qui seraient impossibles ou difficiles sans la vue, mais ne permettent pas d'avoir une expérience réelle de la vision.

Lenay et al. 2000 définissent des conditions pour l'appropriation d'un dispositif de substitution sensorielle. Le système doit être autonome et léger, discret mais doit pouvoir jouer un rôle symbolique comme la canne blanche qui signale la fragilité de la personne pendant son trajet, et doit être peu coûteux. Le dispositif doit correspondre à une attente réelle des populations concernées. On doit également prendre en compte le protocole d'apprentissage, en gardant à l'esprit que le dispositif ne donne pas accès à une vision dégradée mais à un mode de perception nouveau.

Kilian et al. 2022 proposent des prérequis à prendre en compte lors de la conception d'un DSS. Parmi elles, la charge cognitive et la facilité d'apprentissage , qui, dans les systèmes présentés, semblent être négligées au profit de la résolution et de la quantité d'informations.

En effet, dans les AVSS présentés plus haut, le principe est de sonifier tous les pixels d'une image, avec une résolution variable. The vOICE avec originellement une résolution de 64 x 64 pixels (P. B. Meijer 1992), qui a été augmentée à 176 dans les versions plus récentes (Ward et P. Meijer 2010). Dans les autres systèmes, la résolution va de 3 x 3 (Heimler et al. 2019 ; Netzer et al. 2021) à 124 x 124 pixels (Capelle et al. 1998). Or, une étude où les participants devaient reconnaître des formes visuelles grâce à un DSS utilisant l'algorithme the vOICe a montré que la substitution sensorielle peut fonctionner avec peu d'information et qu'elle atteint un plafond à 9 x 9 pixels (Brown, Simpson et Proulx 2014). De plus, les sons correspondants aux contours des formes plutôt qu'aux images pleines étaient plus faciles à interpréter. Les résultats de cette étude pencheraient plutôt vers une simplification des informations transmises plutôt qu'une augmentation de la résolution. De même, dans (Kristjánsson et al. 2016), les auteurs soulèvent le problème des limites attentionnelles. Pour éviter la surcharge sensorielle, il est préférable de transmettre uniquement les informations essentielles, et donc de concevoir des DSS centrés sur des tâches spécifiques (ex. navigation, évitement d'obstacles).

2.1.4 Synthèse

Les dispositifs de substitution sensorielle (DSS) visant à compenser la cécité en convertissant des informations visuelles en sons ou en stimuli tactiles reposent sur un principe ambitieux : exploiter la neuroplasticité cérébrale pour remplacer la vision. Pourtant, malgré des avancées technologiques notables, après des mois d'entraînement, les utilisateurs ne parviennent qu'à accomplir des tâches simples comme reconnaître des formes basiques. L'idée d'une expérience sensorielle équivalente à la vue reste illusoire, et plutôt qu'une substitution sensorielle, certains préfèrent parler de "suppléance perceptive" (Lenay et al. 2000).

Un problème central réside dans la surcharge cognitive générée par ces dispositifs. Les systèmes comme the vOICe ou le PVSA, qui convertissent intégralement l'image en paysage sonore, exigent une attention soutenue et une capacité d'interprétation élevée. Les études de Brown (Brown, Simpson et Proulx 2014) et Kristjansson (Kristjánsson et al. 2016) soutiennent que moins d'informations peuvent paradoxalement améliorer la performance, suggérant que la quantité d'informations transmises doit être optimisée (et donc être adaptée aux capacités perceptives de l'utilisateur), et non maximisée. Tout laisse à penser que les concepteurs de DSS devraient se concentrer sur l'assistance sur une tâche précise plutôt que de chercher à remplacer la vue.

2.2 La perception auditive

2.2.1 Perception des paramètres sonores

Notre perception des sons repose sur une transformation des ondes acoustiques en signaux nerveux, opérée par les différentes structures de l'oreille. Chaque caractéristique du son - sa hau-

teur, son intensité et son timbre - est traitée de manière spécifique grâce à l'anatomie particulière de notre système auditif.

La cochlée, dans l'oreille interne, est le centre de traitement des fréquences sonores. Sa membrane est étroite et rigide à la base, et répond préférentiellement aux hautes fréquences, tandis que son apex, plus large et flexible, est spécialisée dans les basses fréquences. Cette disposition permet une discrimination fine des hauteurs sonores, avec une sensibilité optimale entre 1500 et 6000 Hz, particulièrement autour de 3000 - 4000 Hz (Scharine, Cave et Letowski 2009). Le seuil de discrimination est de seulement 1 à 3 Hz pour les tons purs entre 200 et 2000 Hz (Wier, Jesteadt et Green 1977). Notre système auditif nous permet de distinguer entre 640 et 4000 hauteurs différentes (Zwicker et Fastl 2013). Une étude a montré des capacités de discrimination des fréquences supérieures chez les personnes non-voyantes (Arnaud, Gracco et Ménard 2018 ; Gougoux et al. 2004 ; Rokem et Ahissar 2009 ; Wan et al. 2010).

L'intensité sonore est principalement codée par l'oreille moyenne, où les osselets (marteau, enclume et étrier) protègent l'oreille interne des sons trop intenses tout en amplifiant les sons faibles. Nous pouvons détecter des différences d'à peine 0,4 dB pour les sons complexes, et de 1 à 2 dB pour les sons purs (Jesteadt, Wier et Green 1977). Nous sommes capables de discerner environ 120 pas d'intensité (Zwicker et Fastl 2013), une sensibilité qui diminue avec l'augmentation de l'intensité sonore.

Le timbre est la caractéristique qui nous permet de distinguer deux instruments jouant la même note. Cette caractéristique est principalement le résultat du traitement des harmoniques et de l'enveloppe temporelle du son par notre système auditif. Les différentes phases d'un son (son attaque, son maintien et son extinction) sont analysées séparément et contribuent à notre perception globale du timbre. L'acuité (sharpness) est une composante du timbre qui est liée à la concentration d'énergie dans les hautes fréquences (plus les fréquences dominantes d'un son complexe sont élevées, plus le son est perçu comme "sharp"). Les autres composantes du timbre sont la brillance (*brightness*, liée au centroid spectral), la rugosité (*roughness*, liée aux modulations rapides) et l'ampleur (*fullness*, liée à la largeur spectrale). Jusqu'à 250 pas de sharpness peuvent être distingués (Pedrielli, Carletti et Casazza 2008 ; Zwicker et Fastl 2013). Une étude a montré que les personnes non-voyantes avaient de meilleures capacités de discrimination du timbre (Wan et al. 2010).

Ces seuils de discrimination montrent que les attribuent monoraux (non spatiaux) offrent une résolution bien supérieure à celle de la localisation spatiale qui, comme nous le verrons dans la section suivante, est limitée à quelques dizaines de positions dans l'espace (Blauert 1997 ; Ziemer et Schultheis 2019a).

2.2.2 Localisation d'une source sonore

Plusieurs mécanismes du fonctionnement du système auditif permettent à un auditeur de localiser une source sonore. Ces caractéristiques sont résumées par Carlini, Bordeau et Ambard

2024.

Il existe tout d'abord des indices binauraux :

- La différence temporelle interaurale (ITD) : la différence de temps d'arrivée du son entre les deux oreilles. Elle permet de localiser les sons de basse fréquence (< 1,5 kHz) (Brughera, Dunai et Hartmann 2013 ; Mills 1958).
- La différence de niveau interaurale (ILD) : la différence d'intensité perçue entre les oreilles, due à l'ombre acoustique de la tête. Elle est efficace plutôt pour les hautes fréquences (> 4 kHz) (Shaw 1974 ; Yost 2016).

Des indices monoraux entrent également en jeux. Les fonctions de transfert liées à la tête ou HRTF (Head Related Transfer Functions) permettent de caractériser les modifications qu'un son subit quand il se propage à partir d'une position source dans l'espace, notamment les modifications apportées par le corps de l'auditeur. Elles fournissent des informations sur l'élévation de la source sonore et la distinction avant-arrière. Ces fonctions varient selon l'anatomie individuelle (Majdak, Baumgartner et Jenny 2020 ; Wightman et Kistler 1989). Enfin, d'autres indices permettent de réduire les erreurs dans la localisation d'une source sonore : la réverbération aide à estimer la distance (Zahorik 2002) et réduit les confusions avant-arrière (Begault, Wenzel et Anderson 2001), et les mouvements de la tête améliorent la précision de la localisation en enrichissant les indices acoustiques (Brimijoin et Akeroyd 2014 ; Wightman et Kistler 1997). L'entraînement avec feedback visuel améliore la précision (Garcia et al. 2017).

Le système auditif humain présente des limites dans sa capacité à localiser les sons avec précision :

- **Localisation dans le plan horizontal (azimut)**. En position frontale (0°), l'erreur moyenne est de 1-2° (Makous et Middlebrooks 1990 ; Tabry, Zatorre et Voss 2013). Elle atteint environ 20° en périphérie (±90°) (Tabry, Zatorre et Voss 2013). Les confusions avant-arrière sont fréquentes pour des fréquences inférieures à 2,5 kHz (Stanley Smith Stevens et Newman 1936). Le seuil de discrimination (MAA - *Minimum Audible Angle*) est de 1-2° en position frontale (Mills 1958) et monte jusqu'à 3 à 7° pour des sons en position latérale ou arrière (Aggius-Vella, Kolarik et al. 2020).
- **Localisation sur le plan vertical (élévation)**. La localisation est moins précise qu'en azimut : l'erreur moyenne est de 4-5° en position frontale (Makous et Middlebrooks 1990) et pour des élévations extrêmes (±45°), l'erreur peut dépasser 30° (Tabry, Zatorre et Voss 2013).
- **Estimation de la distance**. Elle est meilleure autour de 1 mètre, avec une surestimation en dessous d'un mètre, et une sous-estimation au-delà (Zahorik 2002). L'erreur d'estimation de distance est autour de 15-20 cm pour des sources proches, et supérieure à 1 m pour des sources lointaines (Aggius-Vella, M. Gori et al. 2022).

Certaines études mettent en évidence des capacités de localisation auditive supérieures chez

les non-voyants pour l'azimut (Doucet et al. 2005 ; Lessard et al. 1998 ; Voss, Tabry et Zatorre 2015), l'élévation (Voss, Tabry et Zatorre 2015) et la distance (Kolarik, Cirstea et Pardhan 2013).

2.2.3 Correspondances cross-modales

Ce qu'on appelle les correspondances cross-modales sont des associations non-arbitraires systématiques entre les modalités sensorielles (par exemple, des sons graves vont être associés à des objets grands) (Spence 2011). Elles sont étudiées dès les années 1920, par exemple avec l'effet "bouba/kiki" où des formes arrondies sont associées à des sons doux (Ramachandran et Hubbard 2001, 2003 ; Sapir 1929).

Spence 2011 distingue trois catégories de correspondances cross-modales :

- **Statistiques** : basées sur des co-occurrences naturelles dans l'environnement (ex. : de gros objets produisent des sons plus graves) (Grassi 2005).
- **Structurelles** : résultent de l'organisation neuronale commune à plusieurs modalités sensorielles (ex. : intensité lumineuse et sonore partagent des codes neuronaux similaires (S. Stevens 1957)).
- **Sémantiques** : médiées par le langage et les métaphores linguistiques (ex. : le mot "haut" utilisé pour parler d'un son aigu et d'une position verticale (Martino et Marks 1999))

Une des correspondances cross-modales les plus étudiées est celle entre hauteur tonale et hauteur visuelle ou spatiale. En effet, les sons aigus sont plutôt associés à des objets situés en haut, et les sons graves à des objets en bas dans l'espace (Bernstein et Edelstein 1971 ; Evans et Treisman 2010 ; Melara et O'Brien 1987 ; Pratt 1930). Une étude (Hamilton-Fletcher, Pieniak et al. 2020) tend à démontrer l'existence de cette association chez les non-voyants. Pourtant, deux autres études suggèrent la nécessité d'une expérience visuelle pour que certaines correspondances cross-modales émergent (Deroy et al. 2016 ; Fryer, Freeman et Pring 2014).

Il existe également une correspondance entre tonalité sonore et taille visuelle : les sons aigus sont associés à des objets petits, et les sons graves à des objets grands (Gallace et Spence 2006 ; Parise et Spence 2009 ; Smith et Sera 1992). Enfin, il existe une correspondance entre direction spatiale et direction du mouvement sonore : un son qui monte en hauteur peut nous faire percevoir un mouvement visuel vers le haut, et inversement (Maeda, Kanai et Shimojo 2004).

Ces correspondances ont pour effet que des stimuli congruents sont traités plus rapidement (Gallace et Spence 2006) et intégrés plus fortement (Parise et Spence 2009).

2.2.4 Synthèse

Le système auditif humain présente des caractéristiques remarquables mais aussi des limitations importantes à prendre en compte lors de la conception d'un dispositif de guidage sonore.

D'un côté, nous sommes très performants pour discriminer finement des paramètres acoustiques comme la fréquence et l'intensité sonore. Cela signifie qu'une légère variation de hauteur ou de volume peut être perçue et interprétée avec une grande précision, ce qui permet de coder des informations de manière très dense sur ces dimensions. En revanche, notre capacité à localiser une source sonore dans l'espace reste relativement imprécise. Les indices de localisation - comme les différences de temps et d'intensité entre les deux oreilles - offrent une résolution spatiale limitée, surtout dans les environnements complexes ou en présence de bruit.

Dans ce contexte, il est plus judicieux, pour concevoir un système de guidage par le son, de ne pas miser sur une spatialisation sonore trop fine, mais plutôt d'utiliser des variations claires et distinctes de paramètres auditifs (fréquence, intensité, timbre), qui sont mieux perçues et interprétées par l'auditeur. Cette approche peut sembler moins intuitive au départ que la spatialisation (par exemple, entendre un son 30° à gauche et s'orienter devant soi à gauche), mais elle peut être rendue plus accessible et efficace grâce aux correspondances cross-modales qui font appel à des intuitions perceptives partagées (par exemple, une montée de fréquence peut être associée à une montée dans l'espace). Ainsi, plutôt que de chercher à reproduire fidèlement la vision par le son, il s'agit de concevoir un langage auditif fonctionnel, adapté aux capacités spécifiques de notre système auditif et surtout à la tâche. Cette stratégie, bien qu'exigeant un apprentissage, offre un meilleur compromis entre efficacité, accessibilité cognitive et performance perceptive.

2.3 Dispositifs de substitution sensorielle appliqués au guidage vers une cible

2.3.1 Perception

Un DSS guidant une personne vers une cible recueille des données spatiales par l'intermédiaire d'un capteur (généralement une caméra). L'écart par rapport à la cible est calculé à partir de ces données. Le cadre de référence peut être la tête (Katz et al. 2012 ; Liu, Stiles et Meister 2018 ; Thakoor et al. 2014) ou la main (Hild et Cheng 2014 ; Manduchi et Coughlan 2014 ; Shih et al. 2018) en fonction de la position de la caméra. Dans certaines études, le guidage passe d'un référentiel centré sur la tête pour la navigation vers l'emplacement de la cible à un référentiel centré sur la main pour un guidage précis de la main vers la cible (May et al. 2019 ; Zientara et al. 2017).

Le cadre de référence pour le choix des métriques spatiales à encoder ne peut ignorer la logique propre au corps des systèmes moteurs et cognitifs. Par exemple, fournir toute la matrice de pixels d'une scène en profondeur (P. B. Meijer 1992) est inutile pour atteindre un seul point dans l'espace, et le surplus de données est difficile à interpréter. Par ailleurs, le fait de ne fournir que la position absolue de la cible entraîne également des erreurs d'interprétation (Gao et al. 2022) même si la quantité de données est réduite à l'essentiel. Pour atteindre un objet cible ou un point dans l'espace, il faut que la personne soit informée de l'écart entre sa position et la position cible.

Reste à savoir quel référentiel utiliser pour mesurer cet écart.

La plupart du temps, les données spatiales à transmettre aux personnes privées de vue sont collectées par une ou plusieurs caméras. Lorsque le cadre de référence est une caméra, celle-ci est à la fois le capteur (qui collecte les données) et le pointeur (cadre de référence dans lequel sont calculés les écarts par rapport à la cible). Il s'agit d'un cadre de référence allocentrique : les données spatiales sont encodées par rapport à un objet, la caméra. En revanche, un cadre de référence égocentré permet d'encoder les données en référence à un système de coordonnées qui place le corps de l'utilisateur comme origine et définit les axes des repères par rapport à l'orientation du corps de l'utilisateur (Ruvolo 2021). Pour les personnes aveugles et malvoyantes, un cadre de référence égocentré est plus efficace (Ruvolo 2021) : en raison du déficit de feedback visuel, il est en effet difficile pour les personnes privées de vue de faire correspondre les coordonnées de la caméra à leur système de coordonnées égocentré.

Un cadre de référence égocentré semblerait donc plus approprié pour guider une personne malvoyante vers une cible. Il paraît pertinent de choisir le doigt comme cadre de référence, car la main est au centre d'une tâche d'atteinte d'objet.

2.3.2 Guidage

Les trois méthodes de guidage par le son. Trois types de retour sonore peuvent être utilisés pour guider une personne : le guidage vocal, également appelé instructions verbales, la sonification spatialisée et la sonification non spatialisée.

Le guidage vocal dans une tâche d'atteinte de cible peut correspondre à des commandes telles que «tourne à droite», «avance», « en haut à gauche», etc.

La sonification consiste à utiliser des sons non-verbaux pour transmettre des données perceptives (Bonebright et al. 2010).

La sonification spatialisée utilise la capacité du système auditif à localiser une source sonore en recréant virtuellement sa position. Pour ce faire, on utilise soit la stéréo, soit les Head Related Transfer Functions (HRTFs), décrits en section 2.2.2. Les HRTFs permettent de transformer un signal monophonique en signal stéréophonique contenant des indices comme l'ITD ou l'ILD. Ce sont ces indices, entre autres, qui permettent à l'être humain de localiser une source sonore. L'acquisition des HRTFs se fait par l'enregistrement de signaux acoustiques reçus par un micro dans chaque oreille lorsqu'un clic est émis à plusieurs positions dans l'espace (Li et Peissig 2020). Les HRTFs peuvent être individualisées (effectuées auprès de la personne qui entendra les sons spatialisés) ou non-individualisés (effectuées auprès d'un auditeur différent ou auprès d'un mannequin). Le résultat dépend de la taille de la tête, la forme du pavillon de l'oreille et la position des épaules.

La sonification non spatialisée utilise les propriétés physiques des sons telles que le volume, la fréquence, le tempo, etc. Par exemple, la distance par rapport à la cible peut être convertie en

tempo.

Avantages et inconvénients de ces trois méthodes de guidage Chacune de ces trois méthodes présente des avantages et des inconvénients.

Le guidage vocal est considéré comme naturel et intuitif, et ne nécessite pas de formation approfondie. Cependant, il peut être considéré comme lent, car il faut du temps pour entendre et traiter chaque instruction, et il y a un délai avant de donner l'instruction suivante. En revanche, la sonification non spatialisée peut faire varier continuellement les paramètres sonores tels que la fréquence, le tempo ou l'intensité en temps réel, reflétant instantanément les mouvements de l'utilisateur et permettant un guidage plus rapide.

La sonification spatialisée a l'avantage d'être naturelle et intuitive, mais elle n'est pas très précise car elle est limitée par la résolution spatiale du système auditif (Carlini, Bordeau et Ambard 2024 ; Middlebrooks 2015), qui, comme nous l'avons vu en section 2.2.2 est peu élevée. Lors de l'utilisation de HRTFs pour spatialiser le son, l'utilisation de HRTFs non-individualisées n'est pas recommandée, car elle augmente les erreurs de localisation, par exemple le risque de confusion entre une source sonore provenant de l'avant ou de l'arrière de la personne, d'en haut ou d'en bas, et réduit la perception d'une source sonore externe à la tête de l'auditeur (Best et al. 2020 ; Wenzel et al. 1993). Mais l'utilisation de HRTFs individualisées est un processus plus long, qui nécessite du matériel et des infrastructures spécifiques (Li et Peissig 2020). De plus, la localisation via HRTF est affectée par l'utilisation d'un casque à conduction osseuse (Voong et Oehler 2019). Un casque à conduction osseuse diffuse le son directement dans l'oreille interne par des vibrations appliquées sur l'os crânien, plutôt que de passer par le tympan. Ce type de casque est particulièrement utile pour les personnes ayant un handicap visuel, qui doivent particulièrement se reposer l'audition pour percevoir leur environnement, car il n'obstrue pas le conduit auditif.

La sonification non spatialisée est moins intuitive et naturelle que le guidage vocal ou la sonification spatialisée, car l'association entre les mesures spatiales et le son est arbitraire, et nécessite un processus d'apprentissage plus long pour maîtriser l'outil. Cependant, elle fournit des indications plus précises, grâce à des seuils de discrimination faibles pour la fréquence et l'intensité par exemple, comme nous l'avons vu en section 2.2.1. Cela signifie, par exemple, que lorsqu'il s'agit de transmettre un écart par rapport à la cible, l'encodage par des changements de fréquence permet une granularité beaucoup plus fine que le son spatialisé. Par conséquent, la sonification non spatialisée peut fournir aux utilisateurs des indications plus précises. Une étude a montré que la localisation des sons était meilleure lorsque l'on ajoutait une sonification non spatialisée à une sonification spatialisée (Song et al. 2022). En outre, il est possible d'exploiter le fait que certaines associations entre les paramètres spatiaux et sonores (correspondances cross-modales, décrites en section 2.2.3) sont plus naturelles que d'autres pour créer un système utilisant la sonification non spatialisée qui soit intuitif et simple à utiliser.

Utilisation des trois méthodes pour le guidage vers des cibles De nombreuses études combinent des instructions verbales et une sonification ou des vibrations. Dans (Manduchi et Coughlan 2014), les sujets doivent trouver et cadrer un objet sur un mur à l'aide d'une application de reconnaissance d'objets sur leur smartphone. Des commandes vocales leur indiquent s'ils doivent tourner à gauche, à droite, en haut ou en bas pour cadrer l'image sur la cible. Le rythme de bips indique ensuite la distance qui les sépare de la cible. Cette étude ne teste pas l'efficacité du feedback, mais l'impact du taux de rafraîchissement et du champ de vision de l'application. Dans (Troncoso Aldas et al. 2020), l'application pour smartphone AIGuide guide la personne à l'aide d'instructions verbales pour localiser et saisir un objet. Lorsque l'objet n'est pas bien cadré, quatre commandes vocales sont utilisées pour corriger la position : gauche, droite, haut, bas. L'application donne alors la position exacte de l'objet (l'objet est à x mètres, y degrés à gauche, z centimètres en dessous de la caméra). En outre, le tempo d'un son indique la distance qui sépare l'objet de la cible. Le feedback audio est comparé à un feedback tactile. Les résultats montrent des préférences variées pour les feedbacks audio et tactiles, bien qu'il n'y ait pas de différence significative entre les conditions en termes de performance (temps d'atteinte). Le système Third Eye de (Zientara et al. 2017) indique quatre directions par quatre commandes vocales : haut, bas, gauche, droite. Ces commandes permettent de guider la tête en direction de l'objet, puis les vibrations prennent le relais pour guider la main vers l'objet.

La sonification spatialisée et la sonification non spatialisée sont le plus souvent combinées pour guider une personne vers une cible. Le système CARA de (Liu, Stiles et Meister 2018) utilise une "voix" provenant de la position de l'objet. La fréquence sonore de la voix augmente à mesure que la distance entre le sujet et la cible diminue. Dans un environnement virtuel, les sujets de l'étude (Lokki et Grohn 2005) se déplaçaient physiquement et devaient atteindre une porte qui émettait un bruit rose depuis sa position. Dans une condition, le rythme des salves de bruit rose augmentait à mesure que la distance diminuait, et un son bref indiquait la hauteur de la source sonore par sa fréquence. Dans (May et al. 2019), les sujets devaient se déplacer et saisir un objet dans un environnement virtuel. Dans une première phase de navigation où la tête était prise comme cadre de référence, le tempo d'un son augmentait au fur et à mesure que le sujet s'approchait de la cible. Dans une deuxième phase où la main était guidée vers la cible lorsque la tête entrait dans un rayon de 0,75 m autour de la cible, un son court de fréquence basse, moyenne ou haute indiquait la hauteur de la cible (Area Cue). Des bips indiquaient la proximité de la cible, leur fréquence et leur rythme augmentant avec la proximité (Proximity Feedback). Les sons étaient spatialisés. Dans (J. C. Lock et al. 2020), les participants devaient pointer une cible en ajustant la position d'une caméra. La spatialisation du son indiquait la position horizontale de la cible, tandis que l'élévation était donnée par la variation continue de la fréquence du son en fonction de l'angle d'élévation de la caméra. Dans (Hu et al. 2022), les auteurs ont comparé l'audio spatialisé à trois autres types de feedback : des instructions verbales (avant, arrière, droite, gauche, la fréquence sonore des instructions variant en fonction de la distance), des vibrations (un moteur sur chaque biceps donnait la direction droite-gauche et un moteur sur chaque poignet donnait la direction avant-arrière, l'intensité variait en fonction de la distance), et une sonification non

spatialisée (celle de Meijer (P. B. Meijer 1992), the vOICe). Sur un écran apparaissaient le curseur de la souris (un cercle) et la cible (un autre cercle, plus grand que le curseur). Les participants devaient atteindre la cible avec le curseur. Dans (Gao et al. 2022), les auteurs explorent comment la sonification non spatialisée peut être combinée au son spatialisé, dans une tâche de pointage. Ils testent différents paramètres sonores combinés au son spatialisé. Enfin, dans (Guarese et al. 2024), les participants devaient trouver un objet sur une étagère. La position horizontale de la cible était toujours codée par un son spatialisé binaire, dans l'oreille droite si la cible était à droite de la main du participant, et dans l'oreille gauche si la cible était à gauche. Le tempo de ce son et son intensité diminuaient au fur et à mesure que la main s'éloignait de la cible. Les auteurs ont comparé un codage absolu de la hauteur de la cible (une fréquence qui correspond à la hauteur de la cible, indépendamment de la position de la main) à des codages dynamiques dans lesquels la fréquence du son varie en fonction de la hauteur de la main par rapport à la hauteur de la cible. Les participants ont abandonné plus rapidement avec le codage absolu et ont été moins précis dans la localisation de la cible qu'avec le codage dynamique.

Les études comparant différents types de retour d'information indiquent une meilleure performance dans la recherche d'une cible en utilisant une combinaison de sonification non spatialisée et de sonification spatialisée, par rapport à la sonification spatialisée seule (Gao et al. 2022 ; Lokki et Grohn 2005 ; May et al. 2019). Dans J. C. Lock et al. 2020, différentes gammes de fréquences étaient testées, et la gamme la plus élevée était la plus performante pour les participants malvoyants, alors que la gamme moyenne était la plus performante pour le groupe de participants voyants avec les yeux bandés. Dans (Liu, Stiles et Meister 2018), le système CARA a surpassé the vOICe dans des tâches de localisation et de navigation. Dans (Hu et al. 2022), les auteurs ont constaté une meilleure performance pour le son spatialisé que pour la sonification non spatialisée (the vOICe), mais les participants n'ont pas réussi à terminer la tâche avec the vOICe, car ils n'ont pas réussi à différencier le curseur de la cible. Cela confirme la nécessité d'adapter le retour sonore aux besoins de la tâche. Le participant n'a pas besoin de connaître sa position absolue et la position absolue de la cible (comme c'est le cas avec the vOICe), mais de connaître la distance entre les deux et la direction dans laquelle il doit se déplacer pour réduire cette distance.

Dans la plupart des études utilisant la sonification, les informations concernant plusieurs dimensions (la position horizontale de la cible, sa position verticale et la distance à la cible) sont données simultanément (Guarese et al. 2024 ; Hu et al. 2022 ; Liu, Stiles et Meister 2018 ; J. C. Lock et al. 2020). Enfin, les études utilisant le guidage vocal utilisent très majoritairement des instructions verbales séquentielles, avec des commandes telles que « droite », « gauche », « haut », « bas » (Ahmetovic et al. 2020 ; Hild et Cheng 2014 ; Troncoso Aldas et al. 2020 ; Zientara et al. 2017), et très peu utilisent des commandes telles que « en haut à droite », « en bas à gauche », etc. (Thakoor et al. 2014) ou indiquent la position exacte de la cible sur toutes les dimensions à la fois (Troncoso Aldas et al. 2020).

2.3.3 Évaluation

Type de tâche Les études précédentes testant des dispositifs de guidage explorent des scénarios variés pour tester l'efficacité des systèmes. Certains évaluent des tâches de localisation et/ou de saisie d'objets réels (Ahmetovic et al. 2020 ; Hu et al. 2022 ; Manduchi et Coughlan 2014 ; Thakoor et al. 2014 ; Troncoso Aldas et al. 2020) (comme des produits alimentaires), ou de navigation dans un environnement réel (Liu, Stiles et Meister 2018) mettant l'accent sur des situations réelles de la vie de tous les jours. A l'inverses, d'autres utilisent des environnements virtuels pour contrôler finement les conditions expérimentales (J. C. Lock et al. 2020 ; Lokki et Grohn 2005 ; May et al. 2019). Enfin, des travaux (Guarese et al. 2024 ; Zientara et al. 2017) intègrent la réalité augmentée, combinant des éléments physiques et numériques.

Les tâches elles-mêmes couvrent un spectre large :

- **Navigation** : Liu et al. (Liu, Stiles et Meister 2018) étudient la capacité des systèmes à guider les utilisateurs sur de longues distances dans des bâtiments inconnus, tandis que Zientara et al. (Zientara et al. 2017) se concentrent sur la navigation en magasin. Dans (Lokki et Grohn 2005), les participants doivent naviguer dans un environnement virtuel jusqu'à atteindre une porte.
- **Saisie de cible** : certaines études (Guarese et al. 2024 ; Hu et al. 2022 ; May et al. 2019 ; Thakoor et al. 2014 ; Troncoso Aldas et al. 2020) testent leur système de guidage sur l'atteinte de cible avec la main.
- **Cadrage ou pointage** : les systèmes de guidage peuvent être appliqués au cadrage d'un objet au centre de l'image de la caméra (Ahmetovic et al. 2020 ; J. C. Lock et al. 2020 ; Manduchi et Coughlan 2014)

Protocole d'entraînement Dans la plupart des études présentées, les participants disposent d'une phase d'entraînement pour se familiariser avec le système. Par exemple, Troncoso Aldas et al. 2020 proposent un tutoriel intégré à l'application AIGuide, complété par une session Zoom pour expliquer les retours sensoriels. De même, Zientara et al. 2017 organisent des sessions préliminaires pour s'exercer à la navigation et à la saisie d'objets. Lokki et Grohn 2005 laissent les participants s'exercer jusqu'à ce qu'ils se sentent à l'aise.

Il est important que les participants soient actifs durant la phase d'apprentissage pour faire le lien entre leurs actions et le feedback qui en résulte. L'importance d'un apprentissage actif a déjà été montré à l'époque du dispositif de substitution visuo-tactile de Bach-y-Rita (Paul Bach-y-Rita 1967). En effet, il a montré que les participants n'arrivaient pas à identifier des formes avec le dispositif s'ils ne pouvaient pas manipuler la caméra. La manipulation active de la caméra permet au participant d'apprendre comment les variations de ses sensations sont liées à ses actions. Il en est de même pour le dispositif the vOICe de Meijer (P. B. Meijer 1992). Une étude explorant l'apprentissage des constances de longueur et d'orientation en 3D chez des utilisateurs de the vOICe a mis en évidence l'importance des mouvements auto-initiés (comme l'inclinaison de la

tête) pour améliorer l'apprentissage.

Mesures d'évaluation Les études recourent à des mesures objectives et subjectives pour évaluer les systèmes.

L'indicateur de performance le plus utilisé est le temps mis pour atteindre la cible, la pointer ou naviguer du point A au point B, selon la tâche utilisée dans les études (Ahmetovic et al. 2020 ; Guarese et al. 2024 ; Hu et al. 2022 ; Liu, Stiles et Meister 2018 ; J. C. Lock et al. 2020 ; Lokki et Grohn 2005 ; Manduchi et Coughlan 2014 ; May et al. 2019 ; Thakoor et al. 2014 ; Troncoso Aldas et al. 2020 ; Zientara et al. 2017). Il est utilisé dans toutes les études citées.

Beaucoup de tests incluent également une mesure de précision comme l'erreur angulaire (J. C. Lock et al. 2020), la distance à la cible (Guarese et al. 2024) ou le taux de réussite (Ahmetovic et al. 2020). Il est également possible d'estimer l'efficacité des trajectoires en mesurant la distance parcourue dans une condition de feedback, comparée à une distance optimale ou à la distance parcourue dans une autre condition de feedback (Liu, Stiles et Meister 2018 ; Lokki et Grohn 2005).

Des mesures subjectives complètent généralement ces données. La charge cognitive peut être évaluée via des questionnaires comme le NASA-TLX (Guarese et al. 2024 ; May et al. 2019), et l'utilisabilité par le System Usability Scale (Ahmetovic et al. 2020 ; Thakoor et al. 2014). Le ressenti subjectif peut aussi être évalué par des questions ouvertes ou des échelles de satisfaction conçues spécifiquement pour l'étude (Hu et al. 2022 ; Manduchi et Coughlan 2014 ; Troncoso Aldas et al. 2020 ; Zientara et al. 2017).

Malgré l'utilisation du temps d'atteinte comme mesure d'évaluation dans toutes les études de ce type, la comparaison des dispositifs entre les études reste compliquée, du fait de la grande variabilité des conditions expérimentales (environnement virtuel ou réel, taille de la cible, position des cibles, etc.). Certaines études utilisent un indice de performance calculé grâce à la loi de Fitts, lié également au temps mis pour atteindre la cible (Hu et al. 2022 ; J. C. Lock et al. 2020).

2.3.4 Synthèse

Les dispositifs de substitution sensorielle (DSS) pour le guidage vers une cible reposent souvent sur des caméras portées sur la tête ou en main pour calculer les écarts spatiaux, mais un référentiel égocentré s'avère plus adapté aux personnes malvoyantes, car il évite les confusions entre les coordonnées de la caméra et leur perception corporelle.

Beaucoup de systèmes combinent différentes méthodes de guidage pour de meilleurs résultats. La plupart des systèmes transmettent les informations sur les trois dimensions spatiales pour une grande précision, au risque de surcharger l'utilisateur. Les évaluations de ces DSS, menées en environnements réels ou virtuels, se basent principalement sur le temps d'atteinte de la cible, une mesure pertinente mais difficile à comparer entre études en raison de la diversité des protocoles

expérimentaux.

Pour transmettre les informations de guidage, trois méthodes sonores sont employées : le guidage vocal, intuitif mais lent ; la sonification spatialisée, naturelle mais peu précise en raison des limites du système auditif ; et la sonification non spatialisée, qui, bien que moins intuitive, offre une meilleure précision grâce à la discrimination fine des fréquences et des intensités.

La sonification non spatialisée peut être rendue plus naturelle en exploitant des associations cross-modales, comme lier une fréquence élevée à une position haute ou un tempo rapide à la proximité. Elle peut donc être à la fois très précise et intuitive, ce qui la rend très pertinente pour le guidage vers une cible.

2.4 Sonification

2.4.1 Conditions pour une sonification de données 3D précise

Les travaux fondateurs sur la sonification dans l'espace tridimensionnel ont mis en évidence plusieurs conditions nécessaires à une restitution précise des données spatiales. À partir de trois études (Arfib et al. 2002 ; Barrass 2005 ; S. Ferguson et al. 2006), Ziemer et Schultheis (Ziemer et Schultheis 2018) identifient trois exigences critiques concernant les affichages auditifs pour la navigation :

- être interprétables de manière interactive ;
- présenter des paramètres auditifs orthogonaux d'un point de vue perceptif ;
- permettre une intégration cohérente des axes.

Ils ajoutent ensuite que les dimensions codées doivent être codées de manière linéaire, continue et être dotées d'une résolution perceptive élevée, faiblement interférentes entre elles, disposer d'une origine audible sans son de référence, et être présentées dans un flux audio unique.

L'orthogonalité perceptive signifie que lorsque deux quantités sont sonifiées simultanément, elles peuvent être toutes les deux interprétées. En outre, si une quantité change, le changement de son peut être attribué à la quantité correspondante et interprété sans ambiguïté. L'orthogonalité ne se réfère pas à des paramètres audios indépendants, mais à des impressions d'écoute indépendantes. Certains attributs du son qui sont physiquement orthogonaux, tels que la fréquence et l'amplitude, interfèrent de manière perceptive.

Tout en reconnaissant l'importance du respect de ces critères pour une sonification précise des données 3D, nous nous interrogeons sur la pertinence de maximiser la précision au détriment de la charge cognitive. Cette question est soulevée dans plusieurs travaux récents sur la sonification, qui mettent en avant la nécessité de limiter la quantité d'informations transmises et de trouver un équilibre entre précision et accessibilité.

2.4.2 Défis liés à la charge cognitives

Bien que ces conditions favorisent une sonification précise, elles soulèvent la question de la surcharge cognitive. La sonification de Ziemer et Schultheis 2018, par exemple, combine cinq dimensions psychoacoustiques continues dans un flux unique, ce qui exige une forte capacité d'interprétation. Pour répondre à ce défi, Simmons et al. 2024 distinguent deux approches : l'approche traditionnelle, reposant sur des paramètres auditifs classiques (fréquence, intensité, etc.), et l'approche incarnée, qui s'appuie sur des connaissances contextuelles socialement acquises pour faciliter l'interprétation.

Comment le soulignent Roddy et Bridges 2020 dans leur modèle ESLM (Embodied Sonification Listening Model), la compréhension des sonifications repose sur des connaissances incarnées, c'est-à-dire des expériences sensorielles, motrices et culturelles. Une sonification trop complexe peut saturer ces capacités, rendant l'interprétation difficile voire impossible pour des novices ou dans des contextes exigeants (comme la navigation en temps réel). Wirfs-Brock et al. 2021 illustrent cette idée en montrant comment des techniques narratives peuvent simplifier l'interprétation de données complexes en s'appuyant sur des métaphores musicales familières (arpèges de piano, percussions). De la même manière, Kantan, Spaich et Dahl 2021 soulignent que dans la rééducation motrice, des métaphores musicales intuitives améliorent l'efficacité tout en réduisant la charge cognitive.

2.4.3 Équilibre entre précision et simplicité

Concevoir une sonification efficace revient à trouver un équilibre délicat entre richesse informative et simplicité perceptive. Dubus et Bresin 2013 rappellent que les mappings les plus utilisés reposent sur une logique écologique (e.g. fréquence pour valeurs numériques), mais que leur efficacité reste peu évaluée empiriquement. John H Flowers 2005 alerte contre les sonifications trop complexes, qui peuvent saturer la perception.

Des solutions innovantes existent, comme l'utilisation de la musique personnelle des utilisateurs comme support de sonification (J. Ferguson et S. Brewster 2019), permettant de réduire la charge cognitive grâce à la congruence perceptive, concept confirmé dans les travaux sur les modèles mentaux (J. Ferguson et S. A. Brewster 2018).

2.4.4 Rôle des conventions culturelles et des correspondances incarnées

Plusieurs auteurs (Gresham-Lancaster 2012 ; Vickers 2016) insistent sur la nécessité d'intégrer des conventions culturelles et musicales pour rendre la sonification des données 3D plus intuitive. Par exemple, un auditeur familiarisé avec la musique occidentale interprétera une montée chromatique comme une tension croissante (Roddy et Bridges 2020). Cependant, comme le note Wilson 2002 dans son analyse de la cognition incarnée, ces associations dépendent fortement du contexte et des expériences individuelles, ce qui complique la création de mappings

universellement efficaces.

Enfin, Preziosi et Coane 2017 et Walker et Kramer 2005 soulignent l'importance du symbolisme sonore et des tests empiriques pour valider les mappings. Une sonification réussie doit non seulement respecter les limites de la perception auditive, mais aussi s'appuyer sur des associations incarnées. On peut faire le lien ici avec la notion de correspondances cross-modales définie plus tôt. Certains paramètres sonores sont associés à certains paramètres visuels ou spatiaux, comme la hauteur sonore qui est associée à la hauteur spatiale ou visuelle. Cette association pourrait être une métaphore linguistique, du fait qu'on parle à la fois de hauteur de son et de hauteur d'un objet dans l'espace. L'intuitivité de la sonification dépendra donc de l'association systématique ou non du paramètre auditif qui sera utilisé pour transmettre l'information spatiale, association systématique qui dépend elle-même des conventions culturelles.

2.4.5 Synthèse

En résumé, une sonification de l'espace 3D efficace doit satisfaire les exigences perceptives (linéarité, orthogonalité, continuité, intégration), tout en évitant la surcharge cognitive. Les travaux récents insistent sur la nécessité de s'appuyer sur des métaphores familières, des connaissances incarnées et des conventions culturelles pour renforcer l'intuitivité. Comme le rappelle Vickers et al. 2014, une sonification réussie repose sur un équilibre : transmettre suffisamment d'informations pour être utile, sans dépasser les capacités perceptives et cognitives de l'auditeur, ce qui suppose une validation empirique systématique adaptée à chaque tâche.

2.5 La loi de Fitts

2.5.1 Première formulation de la loi de Fitts

La loi de Fitts pourrait fournir une métrique pour évaluer nos dispositifs de guidage et les comparer à d'autres dispositifs dans littérature. Cette loi prédit le temps nécessaire à une personne pour pointer une cible (la toucher avec le doigt ou un stylet) en fonction de la taille de la cible et de la distance entre le pointeur et la cible. La figure 2.1 présente la loi de Fitts et ses paramètres.

Pour l'établir, Fitts a cherché à identifier le facteur limitant la vitesse des mouvements contrôlés du corps (Fitts 1954). Il a émis deux hypothèses :

- La vitesse des mouvements contrôlés est limitée par la force musculaire.
- La vitesse des mouvements contrôlés est limitée par la capacité de traitement de l'information du système nerveux humain.

Pour répondre à sa question de recherche, Fitts a créé une tâche de tapotement réciproque, dans laquelle les participants devaient toucher alternativement deux cibles (des rectangles) avec

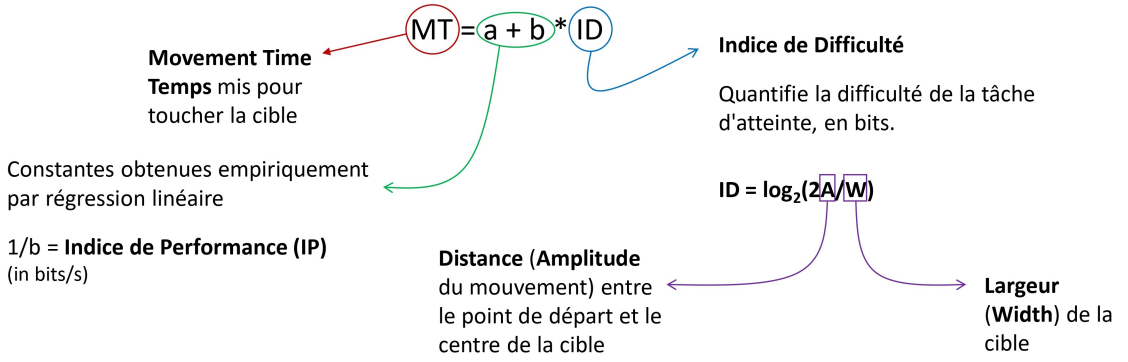


Figure 2.1 – Loi de Fitts et ses paramètres.

un stylet. Deux facteurs variaient d'un essai à l'autre : la largeur des rectangles et la distance entre les deux rectangles.

Pour tester la première hypothèse, Fitts a fait varier le poids du stylet. Pour tester la seconde hypothèse, Fitts a dû quantifier la quantité d'informations contenues dans la tâche. Il a utilisé une analogie entre les amplitudes mécaniques et électriques et la connaissance de la théorie de l'information. Cette analogie est légitime car la théorie de l'information est également valable pour les ondes mécaniques. Il a donc créé un indice de difficulté (ID) :

$$ID = log_2(\frac{2A}{W}) \tag{2.1}$$

L'unité de mesure de l'ID est le bit, qui est la quantité d'information binaire nécessaire pour représenter la difficulté de la tâche d'atteinte. La lettre A signifie "Amplitude" et se réfère à la distance entre la position initiale du pointeur et le centre de la cible. La lettre W signifie "Width" et représente la largeur de la cible. Les résultats de la tâche de tapotement réciproque ont montré une forte corrélation entre l'ID et le temps mis pour atteindre la cible, alors que ce dernier était indépendant du poids du stylet. La vitesse des mouvements contrôlés n'est donc pas limitée par la force musculaire, mais correspond au concept de théorie de l'information. Les mouvements de pointage sont plus longs lorsque la cible est petite et éloignée. La formulation complète de la loi de Fitts est la suivante :

$$MT = a + b * log_2(\frac{2A}{W}) \tag{2.2}$$

Avec MT le temps de mouvement (Movement Time) (le temps qu'il faut pour atteindre la cible), et les constantes a- et b- dérivées empiriquement par régression linéaire. La constante b- est appelée partie informationnelle et a pour unité s/bits. Cette constante indique le temps nécessaire pour traiter un bit d'information. La constante a- a pour unité la seconde et est définie comme le temps de réaction par Fitts et Peterson (Fitts et Peterson 1964) bien qu'il n'y ait pas de consensus sur cette définition, comme nous le verrons par la suite.

Cette loi a d'abord été appliquée aux mouvements continus (les participants touchaient successivement deux cibles plusieurs fois de suite) (Fitts 1954). Plus tard, Fitts a montré qu'elle pouvait être généralisée aux mouvements discrets (les participants devaient rapidement toucher une cible puis revenaient au point de départ) (Fitts et Peterson 1964). La loi s'applique également lorsqu'une contrainte de vitesse est introduite pour démarrer le mouvement de réponse (Fitts et B. K. Radford 1966).

Cette loi est associée à une mesure de performance, d'abord connue sous le nom d'indice de performance ou IP (Fitts 1954), puis généralisée sous le nom de *throughput* ou TP dans la littérature sur les interfaces homme-machine. Cette mesure fournit des informations sur le nombre de bits transférés par seconde. Le TP a d'abord été défini comme (Fitts 1954) :

$$TP = \frac{ID}{MT} \tag{2.3}$$

Cependant, il existe une autre version du TP, et l'utilisation de cette version fait l'objet d'un débat dans la littérature. Nous y reviendrons plus tard.

2.5.2 Autres formulations de la loi de Fitts

À la suite de Fitts, d'autres chercheurs ont tenté d'obtenir un pouvoir prédictif supérieur à la formulation originale de la loi de Fitts en modifiant cette dernière. C'est le cas, par exemple, de MacKenzie (MacKenzie 1989), dont la version modifiée est également connue sous le nom de version de Shannon, car elle découle d'une tentative d'établir une meilleure analogie avec le 17e théorème de Shannon :

$$ID = log_2(\frac{A}{W} + 1) \tag{2.4}$$

Cette version a été largement adoptée par la communauté HCI et est la plus utilisée aujourd'hui. Cependant, plusieurs auteurs rejettent les arguments avancés par MacKenzie. Selon Drewes (Drewes 2010, 2023), l'analogie avec le 17ème théorème de Shannon n'est pas justifiée. Hoffmann 2013 estime également que cette analogie provient d'une mauvaise interprétation du 17ème théorème de Shannon. Il ajoute que cette version ne fournit pas toujours un meilleur ajustement aux données, et que même si c'était le cas, ce n'est pas une raison pour utiliser un modèle théoriquement invalide. En outre, il est vrai que cette formulation donne toujours une valeur positive de l'ID, un des arguments de MacKenzie, mais c'est déjà le cas dans toutes les conditions où le pointeur se trouve en dehors de la cible au début de l'essai. De plus, il n'est peut-être pas important d'avoir une valeur positive de l'ID, car Guiard et H. Olafsdottir 2011 ont montré que l'ID est mesuré sur une échelle d'intervalles, sans valeur zéro absolue.

Enfin, si l'élimination du facteur 2 de l'équation originale de Fitts n'affecte pas la valeur de la constante b-, elle affecte la constante a-, qui n'a donc plus la signification du temps de réaction

(Drewes 2010). C'est pourquoi certains qualifient cette constante a- de paramètre non formel, ce qui contribue à la confusion.

Hoffmann 2013 conclut que seule la version originale de Fitts (ainsi que la version modifiée de Welford (Welford 1970) que nous ne détaillerons pas ici), est théoriquement valide sur la base de la théorie de l'information. Selon Drewes 2023), le temps (MT) prédit par ces formules n'est valable que pour un temps moyen sur de nombreux gestes de pointage. Or, les performances de pointage présentent une forte variance interindividuelle et intra-individuelle. Par conséquent, la question de savoir laquelle de ces formules est correcte est essentiellement théorique, puisque dans la pratique, leurs résultats ne diffèrent que de quelques millisecondes. Néanmoins, il conseille à la communauté HCI d'ignorer et d'oublier les formules alternatives et toutes les recherches qui en découlent.

2.5.3 La loi de Fitts adaptée à des tâches en 3D

Si la loi de Fitts a d'abord été développée pour les mouvements de translation sur une dimension, elle convient également aux mouvements en 2D et est couramment utilisée pour évaluer les performances de systèmes tels que les souris d'ordinateur.

Il y a une quinzaine d'années, des chercheurs ont adopté la loi de Fitts pour analyser les performances des mouvements de l'utilisateur dans des environnements virtuels en 3D (Teather et al. 2009). Plusieurs problèmes se posent alors : en raison des degrés de liberté supplémentaires par rapport à une tâche 2D, une tâche de sélection en 3D est plus difficile. En outre, la tâche originale de Fitts consistait à toucher des cibles sur une table (Fitts 1954), de sorte que l'absence de retour haptique dans un environnement virtuel peut s'avérer problématique. Enfin, le système peut être sujet à des instabilités, qui peuvent être dues à des tremblements de la main, par exemple, ou à des bruits dans le système de suivi. Compte tenu de tous ces problèmes, la loi de Fitts n'est peut-être pas directement applicable à la modélisation des mouvements en 3D.

Plusieurs études ont donc proposé de nouvelles versions de la loi de Fitts, adaptées aux mouvements 3D (Barrera Machuca et Stuerzlinger 2019 ; Cha et Myung 2013 ; Clark, Bhagat et Riggs 2020 ; Murata et Iwase 2001). Cependant, notre objectif est d'utiliser une mesure standardisée pour évaluer et comparer les systèmes de guidage sonore. La multiplication des formules va à l'encontre de cet objectif. De plus, aucun de ces nouveaux modèles n'a mis l'accent sur les performances de l'utilisateur et aucun n'a défini une formulation appropriée du TP. Pour ces deux raisons, et pour celles évoquées dans la section précédente, nous retiendrons dans cette étude la première version de la loi de Fitts, définie par l'équation 2.2 et détaillée en figure 2.1.

2.5.4 La loi de Fitts adaptée à l'atteinte de cible sans la vue

Les premières tâches de Fitts étaient des tâches de pointage de cibles guidées par la vue, où l'on manipulait l'ID en faisant varier la taille des cibles et la distance à la cible. Depuis, un certain

nombre d'auteurs ont étudié l'effet de la sonification ajoutée aux informations visuelles sur le même type de tâche. Dans une étude portant sur l'atteinte d'une cible 1D sur un ordinateur, les participants ont reçu soit un retour sonore (son d'orgue), soit un retour tactile (vibrations de la souris) lorsque le curseur de la souris était sur la cible, soit aucun retour autre que visuel. De même, Batmaz et Stuerzlinger (Batmaz et Stuerzlinger 2021) ont testé l'effet d'un retour sonore indiquant au participant qu'il n'était pas sur la cible (son aigu ou grave continu), dans une tâche d'atteinte de cible en 3D dans un environnement virtuel. Dans ces deux études, les auteurs utilisent la loi de Fitts pour comparer les conditions de sonification, en comparant le coefficient de corrélation entre le temps nécessaire pour atteindre la cible (MT) et l'indice de difficulté (ID) obtenu pour chaque condition. Dans une autre étude (Edström et Hallberg 2014), les participants devaient atteindre des cibles avec leur main dans un environnement virtuel en 3D. La distance à la cible était codée par la fréquence, l'intensité ou le tempo, selon la condition, et les cibles étaient visibles. Les auteurs n'ont pas utilisé la loi de Fitts pour analyser leurs données, car ils n'ont pas fait varier la taille des cibles, mais ils la reconnaissent dans la visualisation des données, car les temps d'atteinte des cibles augmentent avec la distance à la cible.

Les tâches de Fitts peuvent également être utilisées dans le développement de dispositifs de substitution sensorielle pour les personnes aveugles et malvoyantes. La loi de Fitts est alors utilisée pour évaluer les performances dans des tâches d'atteinte de cibles sans vision, guidées par le son ou les vibrations. Dans une étude de Charoenchaimonkon et al. (Charoenchaimonkon, Janecek et al. 2010), trois modalités de retour sensoriel sont comparées pour atteindre des cibles 1D à l'aide d'un stylet : visuelle, auditive et tactile. Dans les modalités auditives et tactiles, un son ou une vibration est émis lorsque le stylet est sur la cible. Les auteurs comparent les performances avec les différents types de feedback en fonction de la difficulté de la tâche, calculée par ID, et constatent que le nombre de cibles atteintes avec le feedback tactile est plus élevé qu'avec le feedback audio pour les ID faibles, mais pas pour les ID plus élevés. Dans une autre expérience (J. C. Lock et al. 2020), des participants devaient pointer des cibles à l'aide de l'appareil photo d'un smartphone. Un son spatialisé était émis, la fréquence variant en fonction de l'angle d'élévation. Les auteurs comparent trois gammes de fréquences pour encoder cet angle. Les conditions sont comparées sur la base de l'erreur angulaire, et le *throughput* calculé avec TP = ID/MT. Les résultats montrent qu'une gamme de fréquences plus élevée produit une erreur angulaire plus faible, mais au détriment de la performance temporelle. Dans une autre étude (Marentakis et S. A. Brewster 2006), les participants devaient atteindre une cible indiquée par un son spatialisé (bruit blanc). Les auteurs ont comparé différentes conditions de mobilité et de retour d'information en calculant le TP. Enfin, Hu et al. 2022 ont comparé quatre types de retour d'information pour une tâche d'atteinte de cible en 3D dans le monde réel : un son spatialisé, des instructions verbales, des vibrations et une sonification. Ils ont comparé le TP pour chaque retour d'information et ont trouvé un meilleur TP pour la condition avec le son spatialisé que pour les autres types de retour d'information.

La loi de Fitts peut donc être utilisée pour décrire les performances d'un système de guidage sonore vers des cibles en 3D, car elle explique bien le temps nécessaire pour atteindre la cible

dans ces expériences.

2.5.5 Évaluer un système de guidage sonore en utilisant les paramètres de la loi de Fitts.

Pour vérifier que la loi de Fitts s'applique aux données, il semblerait que la pratique courante dans la communauté HCI soit de calculer la corrélation entre ID et MT, et de conclure que les données correspondent bien au modèle lorsque cette corrélation est proche de 1. Cependant, comme le souligne Drewes (Drewes 2010), une corrélation non nulle indique seulement un lien linéaire entre deux variables, mais la significativité de ce lien nécessite un test statistique supplémentaire. En outre, de nombreux chercheurs calculent la moyenne du temps d'atteinte pour différentes conditions de A et W qui donnent le même ID. Cette pratique rend la corrélation insignifiante, car elle suppose la validité de la loi de Fitts avant qu'elle n'ait été testée, puisque différentes combinaisons de A et W produisant la même ID devraient avoir des temps d'atteinte identiques (J. Gori et al. 2018).

Une fois que nous avons testé la validité du modèle par rapport aux données, nous pouvons utiliser la loi de Fitts pour évaluer les performances du système. Comme nous l'avons vu, la loi de Fitts est associée à une mesure de performance connue sous le nom de *throughput* (TP), et calculée avec l'équation 2.3 pour la première fois par Fitts en 1954 (Fitts 1954). Cette mesure fournit des informations sur la vitesse, en indiquant le nombre de bits transférés par seconde. Toutefois, Fitts et Radford (Fitts et B. K. Radford 1966) ont par la suite utilisé la pente de l'équation de régression comme indice de performance. La première version du TP est donc concurrencée par une autre version :

$$TP = \frac{1}{b} \tag{2.5}$$

La 1ère version est considérée comme meilleure par certains auteurs (MacKenzie 1992 ; Soukoreff et MacKenzie 2004) parce qu'elle prend en compte la valeur de l'ordonnée à l'origine a-, qui ne serait pas prise en compte si l'on utilisait la réciproque de la pente. D'autres (H. H. Olafsdottir et al. 2012 ; Zhai 2004) soutiennent au contraire que la première équation réduit une mesure multivariée à une mesure unique. En outre, H. H. Olafsdottir et al. 2012 prouvent dans une étude que la valeur du TP calculée par MacKenzie n'est pas invariante comme il l'a prétendu, mais varie à la fois en fonction de la consigne donnée au participant (si elle met l'accent sur la vitesse ou la précision) et de l'ordre dans lequel le TP est calculé (si un TP est d'abord calculé pour chaque ID et ensuite moyenné, ou si ID et MT sont d'abord moyennés globalement et ensuite le TP calculé). Il conseille de revenir à la définition "mathématiquement correcte" du TP, c'est-à-dire la réciproque de la pente, l'équation 2.5.

Enfin, Zhai 2004 souligne que la valeur de TP calculée par l'équation 2.3 dépendrait de la plage d'ID utilisée, ce qui va à l'encontre de l'objectif même de l'utilisation de la loi de Fitts dans

l'évaluation des dispositifs, car elle ne permet pas la généralisation à d'autres conditions expérimentales. Zhai conseille donc d'utiliser la réciproque de la pente (1/b) et l'ordonnée à l'origine (a-), car ces paramètres sont théoriquement intrinsèques au système étudié et peuvent donc être généralisés au-delà des paramètres cibles utilisés dans l'expérience. Soukoreff et MacKenzie 2004 contredisent Zhai en avançant que la réciproque de la pente n'est pas plus indépendante de la gamme des valeurs d'ID.

Les deux équations concurrentes du TP devraient en théorie donner la même valeur si l'ordonnée à l'origine était nulle, ce qui devrait être le cas en théorie, puisque pour un ID égal à 0, MT devrait logiquement être égal à 0. En pratique, ce n'est jamais le cas. La valeur a- peut s'expliquer par le temps de réaction, un retard inévitable dans le système psychomoteur ou un temps supplémentaire pour traiter le feedback (Fitts et B. K. Radford 1966), ou une tâche de mouvement secondaire inexpliquée, comme le temps qu'il faut pour appuyer sur un bouton (MacKenzie 1992). Il n'y a pas de consensus sur le sujet dans la littérature.

Ici, nous nous rangeons à l'avis de Olafsdottir (H. H. Olafsdottir et al. 2012) et Zhai (Zhai 2004), et utilisons la réciproque de la pente (équation 2.5) dans nos calculs de TP (voir chapitre 4 où nous utilisons la loi de Fitts pour l'évaluation des DSS), car il semble plus sûr de ne pas confondre les deux paramètres a- et b- en une seule métrique. Cependant, pour obtenir une image complète des résultats, nous fournirons les deux paramètres a- et b-, ainsi que les temps d'atteinte pour chaque condition de taille et de distance.

2.5.6 Synthèse

La loi de Fitts offre un cadre théorique prometteur pour évaluer l'efficacité des systèmes de guidage sonore en 3D, notamment via le *throughput* (TP) comme métrique clé. En effet, dans toutes les études évaluant un dispositif de guidage par le son, la principale métrique utilisée est le temps d'atteinte. Cependant, la comparaison entre dispositifs par le temps d'atteinte est peu pertinente à cause de la diversité des protocoles expérimentaux. Le TP de la loi de Fitts pourrait donc théoriquement être utilisé comme métrique standardisée pour comparer ce type de dispositifs. Bien que validée pour des tâches visuelles 2D et partiellement explorée avec des retours sonores spatiaux, son application à la sonification non spatialisée est insuffisamment documentée. Plusieurs variantes de la loi ont été proposées, dont celle de Mackenzie, largement utilisée en IHM, mais qui suscite des débats quant à sa pertinence. De même, le calcul du TP fait l'objet de discussions. Ces choix méthodologiques conditionnent la comparabilité des résultats entre études. En prenant en compte ces débats, nous avons décidé de tester l'applicabilité de la loi de Fitts pour évaluer notre dispositif de guidage et les différentes métaphores sonores utilisées afin de permettre la comparaison aux autres dispositifs de la littérature (chapitre 4).

3. Série expérimentale A : Optimisation et validation empirique de sonifications pour le guidage du corps et de la main vers des cibles 3D

Sommaire

3.1 Introduction 36

3.2 Étude A1 : Comparaison de sonifications explicites vs parcimonieuses pour l'atteinte d'une cible avec la main sur une tâche en 3D 38

3.2.1 Introduction 38

3.2.2 Matériel et Méthode 39

3.2.3 Résultats 45

3.2.4 Discussion 47

3.2.5 Conclusion 49

3.2.6 Résumé de l'étude A1 49

3.3 Étude A2 : Comparaison de métaphores sonores pour l'atteinte d'une cible avec la main sur une dimension. 51

3.3.1 Introduction 51

3.3.2 Matériel et Méthodes 54

3.3.3 Résultats 58

3.3.4 Discussion 63

3.3.5 Conclusion 65

3.3.6 Résumé de l'étude A2 65

3.4 Étude A3 : Comparaison d'un mode de sonification séquentiel à un mode de sonification parallèle et à un guidage vocal sur une tâche d'atteinte de cible 3D avec la main 67

3.4.1 Introduction 67

3.4.2 Matériel et Méthodes 69

3.4.3 Résultats 76

3.4.4 Discussion 81

3.4.5 Conclusion 83

3.4.6 Résumé de l'étude A3 84

3.1 Introduction

Cette thèse vise à répondre à la question suivante : comment concevoir et évaluer des métaphores sonores pour guider une personne vers une cible sans la vue ? Dans ce chapitre nous nous penchons sur la conception des métaphores sonores, et cherchons donc à répondre à la première des sous-questions que nous avons posées : Quels paramètres sonores sont les plus efficaces et intuitifs pour encoder les trois dimensions spatiales, sans dépasser les capacités d'interprétation des utilisateurs ?

Comme présenté dans l'état de l'art (section 2), les premiers travaux sur la substitution sensorielle visaient à remplacer la vision par des paysages sonores. Ces approches, trop riches en informations, surchargeaient les utilisateurs. Par la suite, les recherches se sont davantage concentrées sur des tâches spécifiques, telles que l'atteinte de cible. Le guidage sonore doit alors prendre en compte les limites du système auditif (comme la faible précision de la localisation spatiale) mais aussi ses atouts, comme la capacité de discrimination des fréquences et des intensités. Il doit également considérer les contraintes du système cognitif. La sonification non spatialisée peut paraître plus arbitraire et moins intuitive que le guidage vocal ou le son spatialisé. Cependant, grâce aux correspondances cross-modales et aux bonnes capacités discriminatives de l'audition, elle peut se révéler à la fois précise et efficace. Le défi consiste donc à trouver un équilibre entre précision et économie cognitive. Or, nous avons relevé un manque dans la littérature : nombre d'études proposent des systèmes de guidage sonore sans véritable comparaison empirique entre approches.

Ce chapitre s'attaque donc à deux des trois limites des DSS identifiés dans la littérature :

— Le manque de validation empirique systématique des sonifications : nous comparons systématiquement plusieurs modes de guidage sonore pour identifier les plus efficaces et intuitifs.
— La surcharge cognitive induite par des systèmes trop complexes : nous partons du codage le plus simple puis nous l'enrichissons au fur et à mesure, en testant à chaque étape l'intérêt de cet ajout d'information. Les métaphores sonores sont choisies pour leur simplicité (codage binaire plutôt que continu), leur intuitivité (e.g. utilisation de la stéréo et de sons verbaux), et leur lien avec des associations cross-modales connues (e.g. une fréquence montante = il faut monter la main.

Dans ce chapitre, nous confrontons plusieurs modes de sonification d'une tâche d'atteinte de cible avec la main en environnement virtuel 3D. La tâche peut se décomposer en deux phases : marcher vers la cible puis la toucher avec le doigt. Les participants doivent ainsi toucher successivement plusieurs cibles virtuelles du doigt, le plus rapidement possible. L'évaluation se fait en comparaison les sonifications par paires.

— Étude A1 (section 3.2) : Comparaison de sonifications explicites vs parcimonieuses pour l'atteinte d'une cible avec la main sur une tâche en 3D. Nous commençons par une sonifi-

cation simple avec un signal binaire indiquant simplement à l'utilisateur s'il pointe la cible ou non. Nous ajoutons ensuite une information continue sur l'écart angulaire entre la direction pointée et la cible, afin d'accroître la précision et de réduire le temps de recherche. Enfin, nous distinguons les écarts verticaux et horizontaux en deux paramètres sonores distincts. Cette approche enrichit l'information transmise tout en limitant la charge cognitive, puisqu'elle évite au participant d'inférer lui-même les composantes verticales et horizontales, et parce qu'elle combine paramètres binaires et continus.

— Étude A2 (section 3.3) : Comparaison de métaphores sonores pour l'atteinte manuelle d'une cible sur une dimension. Nous comparons différentes métaphores sonores pour coder séparément les écarts verticaux et horizontaux. Les résultats obtenus alimentent l'étude A3.

— Étude A3 (section 3.4) : Comparaison d'un mode de sonification séquentiel à un mode de sonification parallèle et à un guidage vocal sur une tâche d'atteinte de cible 3D avec la main. Nous confrontons une sonification parallèle (présentation simultanée des métaphores sonores de chaque dimension spatiale) et une sonification séquentielle (présentation successive des dimensions). Cette dernière modalité devrait réduire la charge cognitive en limitant la quantité d'informations à traiter en même temps, tout en préservant la précision. Dans cette même étude, nous comparons également la sonification séquentielle à un guidage vocal assuré par un humain.

3.2 Étude A1 : Comparaison de sonifications explicites vs parcimonieuses pour l'atteinte d'une cible avec la main sur une tâche en 3D

3.2.1 Introduction

Dans cette étude, nous testons un DSS visant à guider des personnes privées de vue pour atteindre une cible avec la main, en convertissant des écarts spatiaux entre la personne et la cible en sons. Nous cherchons à créer un dispositif qui soit le plus efficace possible (qui permette d'atteindre la cible facilement et rapidement) mais également confortable, qui ne crée pas de surcharge cognitive.

Pour obtenir un guidage 3D efficace, mais aussi pour réduire la charge cognitive, il faut faire des choix entre guidage explicite et parcimonieux :

- Encoder les trois dimensions spatiales de manière explicite (ce qui nécessite l'utilisation de nombreux paramètres sonores) OU réduire le nombre de dimensions encodées pour minimiser la quantité d'informations à traiter. Dans ce dernier cas, l'utilisateur doit inférer chaque dimension.
- Utiliser un seul flux auditif pour les trois dimensions pour avoir toutes les informations disponibles instantanément (pas besoin de changer son attention d'un flux à l'autre) OU utiliser plusieurs flux auditifs séparés pour réduire la quantité d'information dans chaque flux. Dans ce dernier cas, l'utilisateur a besoin de changer son attention d'un flux à l'autre.
- Utiliser des échelles continues (plus précises) OU discrètes (plus simples).

En tenant compte de ces différents choix et de la littérature sur le sujet (voir section 2.4), nous avons créé trois types de sonification des écarts à la cible, que nous avons comparés par paires. Nous partons de la sonification la plus simple en augmentant au fur et à mesure la quantité d'informations transmises, tout en essayant de limiter la charge cognitive en utilisant des codages intuitifs et en combinant échelle continue et discrète.

Comparaison 1 : Angle-to-Pitch (AP) vs Binary (BI). BI ne donne que des informations binaires : un son (bruit blanc) lorsqu'on pointe dans la bonne direction vs aucun son. Des études ont montré que de simples indices sonores binaires peuvent suffire à créer un système de guidage (Marston et al. 2007).

AP donne un écart angulaire entre la direction pointée et celle de la cible, exprimée par la fréquence du son, dans des échelles continues. Un seul paramètre sonore est utilisé pour exprimer

une seule métrique spatiale globale en trois dimensions. Avec AP, chacune des trois dimensions peut être déduite par la boucle action-perception.

AP possède plusieurs qualités mises en évidence par Ziemer et al. (Ziemer, Nuchprayoon et Schultheis 2019 ; Ziemer et Schultheis 2018, 2019a,b) : les dimensions sont intégrées en un seul flux auditif, la résolution perceptive est haute (grâce à l'utilisation de la fréquence) et les échelles sont continues. En revanche, les dimensions ne sont pas orthogonales : un seul paramètre sonore est utilisé pour encoder la direction sur toutes les dimensions spatiales. Cela pourrait nécessiter plus d'efforts pour interpréter la sonification et faire augmenter la charge cognitive. Pour éviter cela, l'écart directionnel peut être séparé en deux paramètres sonores différents pour coder l'écart vertical et l'écart horizontal. Mais intégrer plusieurs paramètres sonores en un seul flux peut mener à des erreurs d'interprétation et une surcharge cognitive.

Comparaison 2 : AP VS Dissociated Vertical-Horizontal (DVH). Avec DVH, on dissocie deux métriques spatiales en deux paramètres sonores donnés simultanément en deux flux audios superposés. L'écart angulaire sur la dimension horizontale est codé sur une échelle continue, par la fréquence du son. L'écart sur la dimension verticale est codé de manière binaire par la présence d'un bruit blanc de bonne hauteur. Ainsi les deux dimensions sont orthogonales, l'utilisateur peut les interpréter séparément, ce qui devrait simplifier la sonification et rendre le guidage plus précis. L'utilisation d'informations binaires devrait faciliter davantage la sonification et réduire la quantité de données à intégrer et donc la charge cognitive.

Pour comparer ces trois types de sonification, nous avons effectué des tests dans un environnement 3-D. Les participants, les yeux fermés, se déplaçaient pour chercher et atteindre des sphères virtuelles avec leur index. Chaque sonification a été évaluée par le temps mis pour atteindre la cible et par un questionnaire.

3.2.2 Matériel et Méthode

Participants

Nous avons réalisé deux comparaisons. Les participants étaient des étudiants de psychologie (leur participation leur permetteait d'obtenir un point bonus dans une matière), des étudiants d'autres filières et d'autres volontaires.

Comparaison 1 : 14 participants normo-voyants et 1 participant déficient visuel ont testé les métaphores BI et AP.

Comparaison 2 : 8 participants normo-voyants et 1 participant déficient visuel ont testé les métaphores AP et DVH.

Tous les participants ont donné leur consentement écrit avant participation à l'étude.

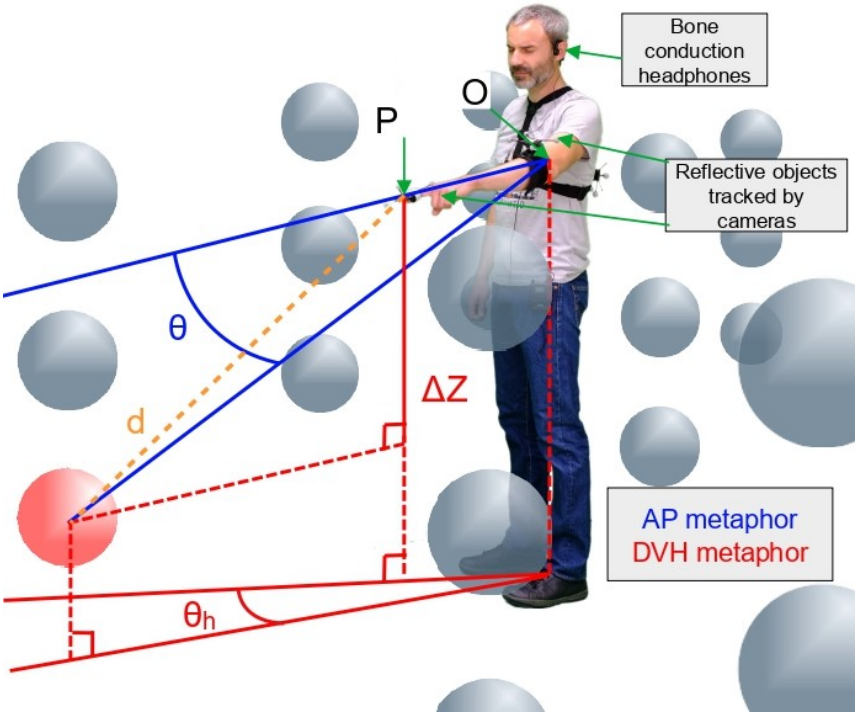


Figure 3.1 – Représentation de la tâche d'atteinte de cibles 3D.
La figure montre les objets avec réflecteurs utilisés pour capturer la position des membres du participant, la position des sphères virtuelles (la sphère rouge étant la cible à atteindre) et les métriques spatiales utilisées pour les modes de sonification BI (Binary), AP (Angle-to-Pitch) et DVH (Dissociated Vertical and Horizontal). La sphère cible est située à l'une des 27 positions possibles. Les points O (coude) et P (extrémité de l'index) déterminent la direction indiquée par le participant. Avec BI, un bruit blanc est déclenché lorsque la ligne OP coupe la cible (θ nul). Avec AP, la hauteur du son varie en fonction de θ. Avec DVH la hauteur du son varie en fonction de θ_h (le projeté de l'angle θ dans le plan horizontal) et un bruit blanc est déclenché lorsque P est à la même hauteur que la cible.

Matériel

Les tests prenaient la forme d'un jeu virtuel consistant à atteindre des sphères dans un espace 3D avec son index (voir Fig. 3.1). Le système de capture de mouvement utilisé était tel que celui décrit en section 1.4 (voir figure 1.1). Le participant portait deux objets équipés de réflecteurs captés par le système, qui permettaient de localiser les points anatomiques servant de référentiel dans le calcul des écarts à la cible : l'index et le coude.

La cible était une sphère de 30 cm de diamètre qui pouvait être positionnée à 27 endroits différents sur une grille 3*3*3 (voir Figure 3.2). Un bloc de test consistait en une série de 6 cibles consécutives à atteindre, positionnées de manière semi-aléatoire pour s'assurer que le participant parcourait toujours une distance minimale entre les cibles (environ 1,88 m). La séquence garantissait que chacune des trois coordonnées x ; y ; z des sphères variaient entre chaque essai.

Pendant l'expérience, le participant était équipé de :

- Objets avec réflecteurs, localisés par le système de capture de mouvement. Les objets étaient positionnés sur des parties du corps correspondant aux points O et P : une bande autour

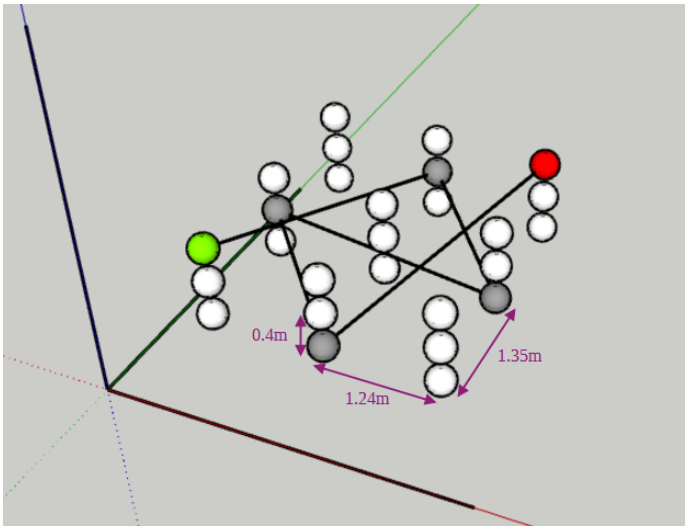


Figure 3.2 – Scène virtuelle représentant une séquence possible (un bloc) de 6 cibles parmi les 27 positions de sphères. La sphère verte représente la première cible, la sphère rouge la dernière cible, les sphères grises les cibles intermédiaires. Il y a une distance de 1,24 m entre deux colonnes de sphères en x, de 1,35 m en y, et trois hauteurs possibles en z : 70 cm, 110 cm et 150 cm.

du doigt plaçait le point P sur le bout de l'index de la main dominante, et un brassard du même côté plaçait le point O sur le coude (voir Figure 3.1). Le vecteur OP définissait la direction pointée par le participant.

- Un casque à conduction osseuse (Aftershokz Sportz3) connecté à un boîtier récepteur (Mipro MI909R) pour recevoir le retour sonore.
- Un clavier miniature (iclever Rii 2.4 GHz L * l * h = 151 * 59 * 12.5 mm) pour démarrer chaque essai.

Métriques spatiales

Plusieurs caractéristiques étaient extraites de la scène pour être transcodées en sons (Figure 3.1).

Pour les modes BI et AP, les métriques spatiales utilisées étaient les suivantes :

- L'angle θ qui correspond à l'angle formé par les demi-droites OP et OT (écart angulaire).
- La distance d = [PT] qui correspond à la distance entre le doigt et la cible (écart en distance).

Pour le mode DVH, nous utilisions deux métriques spatiales supplémentaires :

- L'angle θ_h qui correspond à la projection de l'angle θ sur le plan horizontal (écart angulaire azimutal).
- La différence de hauteur ΔZ = |Pz - Tz|, qui correspond à la valeur absolue de l'écart de hauteur entre la cible et le doigt (écart en distance).

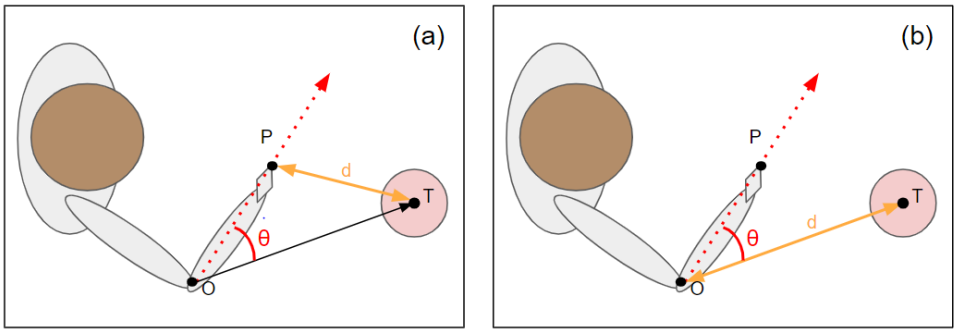


Figure 3.3 – Représentation d'un pointage naturel (a) vs. des coordonnées polaires classiques (b)
Pointage naturel : l'angle θ a pour origine O alors que la distance d a pour origine P. Coordonnées polaires classiques : l'angle θ et la distance d ont la même origine O.

Tableau 3.1 – Métaphores sonores utilisées pour la sonification BI

	Horizontale	**Verticale**	**Distance**
Bruit Blanc	OP intersecte la cible 80 dB		P entre dans la cible 85 dB
Buzzer			P dépasse la cible

Ces métriques nous permettaient de dissocier les écarts sur les dimensions verticale et horizontale pour le mode DVH.

O et P définissent la direction pointée par le participant et la distance entre son doigt et la cible. Pour la métaphore AP, nous avons utilisé un système de coordonnées polaires. Cependant, dans un système de coordonnées polaires classique, l'angle entre la direction pointée et la direction de la cible a une même origine O. C'est le cas dans la plupart des études qui utilisent une caméra pour estimer à la fois l'angle et la distance à la cible (Hild et Cheng 2014 ; Manduchi et Coughlan 2014 ; Shih et al. 2018 ; Troncoso Aldas et al. 2020). Ici, nous dissocions O et P. O est utilisé pour estimer l'orientation de l'angle θ, alors que P est utilisé pour estimer la distance à la cible d (Figure 3.3). Nous aurions pu utiliser O pour déterminer la distance à la cible, mais dans une tâche d'atteinte de cible avec la main, la distance est intuitivement considérée comme nulle quand la main touche la cible.

Sonifications

Les métaphores sonores utilisées pour chaque mode de sonification étaient les suivantes :

BI (voir tableau 3.1) : quand le participant pointe en direction de la cible (si la demi-droite OP intersecte la cible), un bruit blanc est généré. Si le participant ne pointe pas en direction de la cible, aucun son n'est généré.

Pour les modes de sonification AP et DVH, les métriques spatiales extraites étaient encodées en générant un son pur dont la fréquence sonore variait. Quand l'angle θ pour AP ou θ_h pour

Tableau 3.2 – Métaphores sonores utilisées pour la sonification AP

	Horizontale	**Verticale**	**Distance**
Fréquence	θmin = 5°, θmax = 90° *fmax* = 440 Hz, *fmin* = 110 Hz		
Bruit Blanc	OP intersecte la cible 80 dB		P entre dans la cible 85 dB
Buzzer			P dépasse la cible

DVH dépassait 90°, le son pur n'était plus généré, ce qui permettait au participant de savoir que la cible était derrière lui.

AP (voir tableau 3.2) : la fréquence du son pur varie sur un continuum entre *fmin* = 100 Hz pour θmax = 90° à *fmax* = 440 Hz pour θmin = 5° de manière à ce que :

1. aucun son ne soit déclenché quand la cible est derrière le participant ;
2. f soit dans le spectre audible tout en évitant les fréquences trop élevées qui peuvent être agressives sur un stimulus continu.

L'intensité était constante. Un bruit blanc était superposé au son pur quand la demi-droite OP coupait la cible (de la même manière que pour BI). La conversion angle-fréquence suivait la loi psychophysique de Stevens (S. Stevens 1957), une fonction de loi de puissance :

$$log(f) = Af * log(X) + Bf \tag{3.1}$$

Avec :

$$Af = \frac{log(fmax) - log(fmin)}{log(Xmin) - log(Xmax)} \tag{3.2}$$

Et :

$$Bf = log(fmin) - Af * log(Xmax) \tag{3.3}$$

Dans cette étude, X correspond à θ.

DVH (voir tableau 3.3) : cette sonification dissociait dimension verticale et horizontale. La fréquence du son pur variait en suivant la même loi de puissance que pour AP, mais en fonction de l'angle θ_h. L'écart directionnel horizontal était donc codé par la fréquence, alors que la dimension verticale était codée par l'activation d'un bruit blanc quand le doigt du participant était à la même hauteur que la cible (i.e. quand ΔZ < 15 cm (le rayon de la cible)). L'intensité était constante.

Pour les trois sonifications, un buzzer (signal carré) est déclenché lorsque le doigt du partici-

Tableau 3.3 – Métaphores sonores utilisées pour la sonification DVH

	Horizontale	**Verticale**	**Distance**
Fréquence	θmin = 5°, θmax = 90° *fmax* = 440 Hz, *fmin* = 110 Hz		
Bruit Blanc		$\|\Delta Z\|$ <15 cm 80 dB	P entre dans la cible 85 dB
Buzzer			P dépasse la cible

pant dépasse la cible ([OP] > [OT]). Lorsque le pointeur P pénètre dans la cible (d < 15 cm), seul un bruit blanc se substitue à la sonification, dont l'intensité est plus forte que le bruit blanc de bonne direction (condition AP) ou de bonne hauteur (condition DVH).

Protocole

L'expérience se déroulait dans l'espace de capture de mouvement décrit ci-dessus. Les participants devaient trouver plusieurs cibles présentées successivement. Chaque participant effectuait l'une des deux comparaisons suivantes : BI vs AP ou AP vs DHV. L'expérience commençait par une phase d'entraînement, qui consistait en un bloc de six essais par mode de sonification. Au cours de cette phase, le participant recevait des explications de l'expérimentateur sur les difficultés qui pouvaient survenir et sur la manière de les surmonter. La session se poursuivait par la réalisation de quatre blocs expérimentaux alternant les deux conditions de sonification. L'ordre de passage des deux sonifications était contrebalancé entre les participants. Chaque bloc expérimental était précédé d'un bloc de rafraîchissement de la sonification suivante. Chaque participant faisait autant de blocs de rafraîchissement que souhaité et indiquait lorsqu'il était prêt à commencer le bloc expérimental.

Chaque bloc se déroulait comme suit : le participant se plaçait au centre de la pièce et fermait les yeux. Dès qu'il était prêt, le participant initiait lui-même le premier essai en appuyant sur la touche N du mini-clavier. Un son de démarrage confirmait le début de l'essai. Le retour sonore était déclenché et changeait en fonction des mouvements du participant, selon le mode de sonification. Lorsque P restait dans la cible pendant 400 ms, un son indiquait la victoire et la fin de l'essai, et le retour sonore était désactivé. L'essai était interrompu si le participant n'atteignait pas la cible dans la limite des 180 secondes, et un son de cloche indiquant la défaite était déclenché. Après chaque essai, le participant restait sur place et commençait l'essai suivant en appuyant à nouveau sur la touche N. Il continuait ainsi jusqu'à ce que les six essais du bloc expérimental soient terminés.

À la fin de l'expérience, les participants remplissaient un questionnaire. Ils devaient donner leur avis général sur l'expérience, les difficultés rencontrées et évaluer sur une échelle de Likert chacune des deux conditions sur trois aspects : la facilité, le confort et l'efficacité.

Analyses

Pour comparer les sonifications, nous avons mesuré la performance objective de nos participants par le temps mis pour atteindre la cible. Comme notre variable réponse (temps pour atteindre la cible) est une durée dont la distribution est asymétrique, le modèle de Cox est apparu comme le plus approprié (Letué et al. 2018). Il nous permet d'analyser des mesures répétées sans faire la moyenne des données pour chaque participant, il tient donc compte de la variabilité intra- et inter-participants. Pour réaliser nos analyses, nous avons utilisé la fonction coxph du package Survival du logiciel R.

Les participants ont attribué un score d'ergonomie à chaque condition de sonification selon trois critères : l'efficacité, le confort et la facilité. Un modèle type des questionnaires utilisés tout au long de ce travail de thèse est inclus en annexe 7.1. Le score prend comme valeur un nombre entier. Nous avons choisi de nous intéresser à la distribution de probabilité des valeurs prises, et donc de considérer le score comme une variable catégorielle ordinale. Nous avons effectué une régression ordinale, en utilisant la fonction clm() du package ordinal du logiciel R.

3.2.3 Résultats

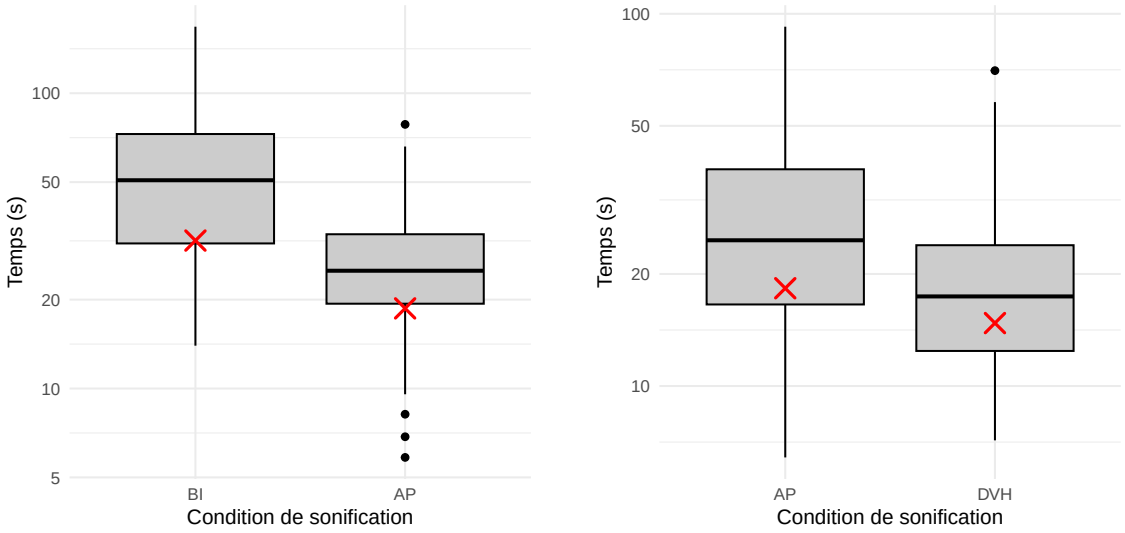
Comparaison 1 : Binary (BI) vs. Angle-to-Pitch (AP)

Temps d'atteinte de cible : Comme le montre la figure 3.4(a), les participants étaient en moyenne 29 secondes plus rapides avec la sonification AP (M = 27,70 s ; SD = 12,27 s) qu'avec BI (M = 56,68 s ; SD = 33,78 s) (z = -10,41 ; p < 0,001). Le participant déficient visuel a également obtenu de meilleurs temps pour AP (M = 19,02 s ; SD = 9,24 s) que pour BI (M = 35,55 s ; SD = 19,29 s). Ses temps médians sont représentés par des croix rouges sur la figure 3.4(a).

Scores d'ergonomie Selon les scores d'ergonomie attribués à chaque condition de sonification par les participants (voir Fig. 3.5), AP est considéré comme plus confortable que BI (z = -56. 62, p < 0,001), plus efficace (z = 2,74, p < 0,01) et plus facile à utiliser (z = 3,02, p < 0,01).

Comparaison 2 : Angle-to-Pitch (AP) vs. Dissociated Vertical and Horizontal (DVH)

Temps d'atteinte de cible : Comme le montre la figure 3.4(b), les participants étaient en moyenne 10 secondes plus rapides dans la condition DVH (M = 19,70s ; SD = 10,77s) que dans la condition AP (M = 29,85s ; SD = 17,70s) (z = 4,80 ; p < 0,01). Le participant malvoyant a obtenu des résultats similaires, avec de meilleurs temps pour la sonification DVH (M = 17,95 s : SD = 10,14 s) que pour la sonification AP (M = 21,95 s : SD = 12,60 s). Ses temps médians sont représentés par des croix rouges sur la figure 3.4(b).



(a) BI vs AP **(b)** AP vs DVH

Figure 3.4 – Temps d'atteinte de cibles (échelle log10) pour chaque condition de sonification pour les comparaisons BI vs AP et AP vs DVH.
BI (Binary) donne la bonne direction par un bruit blanc binaire. AP rajoute à BI l'écart de direction global exprimé en fréquence. DVH dissocie les écarts sur les dimensions horizontale (fréquence) et verticale (bruit blanc binaire). Chaque boîte représente la médiane (ligne horizontale), l'intervalle interquartile (boîte) et les valeurs extrêmes aberrantes (moustaches). Les croix rouges représentent les temps médians des participants malvoyants.

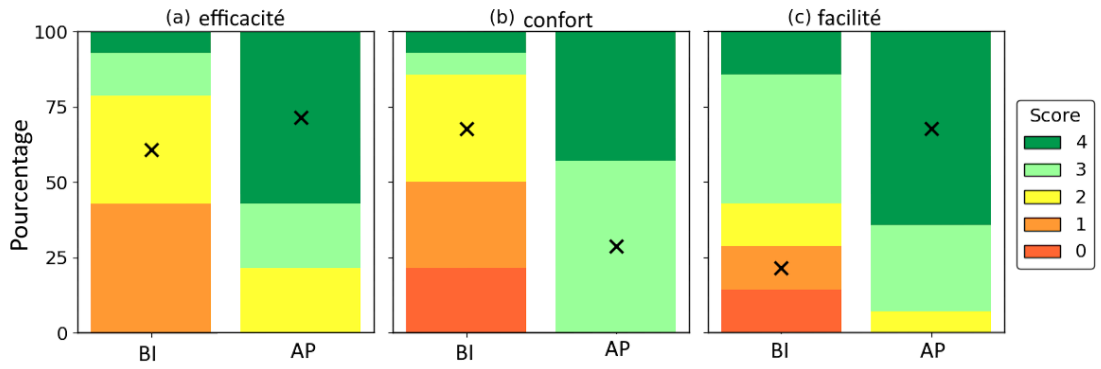


Figure 3.5 – Pourcentage de participants ayant donné chaque score dans les deux conditions de sonification BI et AP.
BI (Binary) donne la bonne direction par un bruit blanc binaire. AP rajoute à BI l'écart de direction global exprimé en fréquence.Les trois critères sont pris en compte (efficacité, confort, facilité). Le gradient de couleur représente les niveaux de notation, le rouge correspondant à 0 et le vert foncé à 4, de sorte que les zones plus vertes correspondent à des jugements plus positifs. Les croix noires indiquent le score attribué par le participant malvoyant.

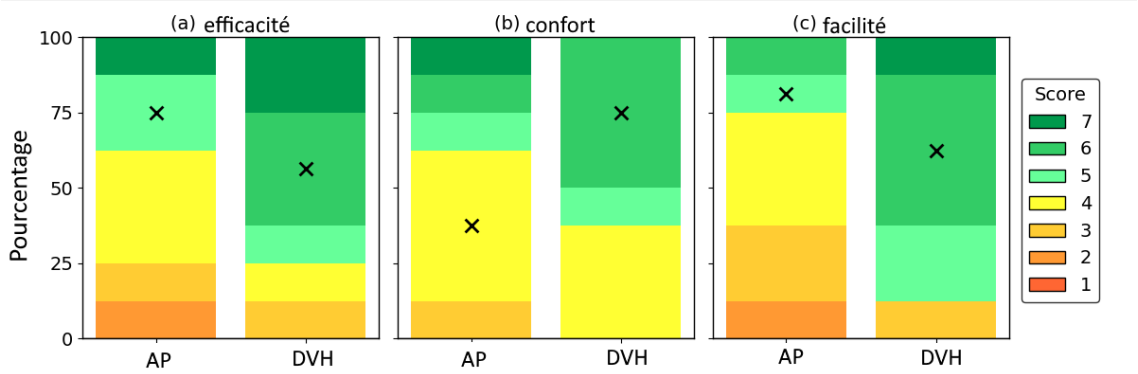


Figure 3.6 – Pourcentage de participants ayant donné chaque score dans les deux conditions de sonification AP et DVH.
AP rajoute au bruit blanc de bonne direction l'écart de direction global exprimé en fréquence. DVH dissocie les écarts sur les dimensions horizontale (fréquence) et verticale (bruit blanc binaire).Les trois critères sont pris en compte (efficacité, confort, facilité). Le gradient de couleur représente les niveaux de notation, le rouge correspondant à 1 et le vert foncé à 7, de sorte que les zones plus vertes correspondent à des jugements plus positifs. Les croix noires indiquent le score attribué par le participant malvoyant.

Scores d'ergonomie La tâche a été jugée plus facile avec DVH qu'avec AP (voir Fig. 3.6) (z = -2,29 ; p = 0,02).

Les scores donnés par les participants malvoyants sont similaires à ceux donnés par les participants voyants, avec de meilleurs scores pour AP que pour BI et de meilleurs scores pour DVH que pour AP.

3.2.4 Discussion

Notre objectif était de créer le dispositif de guidage sonore le plus efficace et le plus ergonomique possible pour aider les personnes aveugles et malvoyantes à atteindre leur cible. La plupart du temps, la distance entre l'utilisateur et la cible est calculée à partir du capteur, souvent une caméra portable ou montée sur la tête. Ici, le cadre de référence est le corps du participant.

Nous avons créé trois modes de sonification en tenant compte des conseils de la littérature pour concevoir un guidage sonore précis, mais en tenant compte également de la charge cognitive imposée à l'utilisateur. Nous avons testé ces sonifications par paires, en partant de la plus simple (signal binaire) puis en enrichissant au fur et à mesure l'information transmise, dans le cadre d'une tâche d'atteinte d'une cible virtuelle avec la main. La performance des participants était mesurée par le temps nécessaire pour atteindre la cible, et leur évaluation de chaque sonification était recueillie à l'aide d'un questionnaire.

Nous avons d'abord comparé BI, qui donne le minimum d'information pour réussir la tâche, une information binaire de bonne direction, à AP, qui donne en plus de cette information binaire un écart directionnel par rapport à la cible, exprimé par la fréquence du son sur une échelle continue. Les résultats ont montré la supériorité de AP sur BI en termes de rapidité mais également de facilité, de confort et d'efficacité. Transmettre uniquement une information binaire de bonne

direction n'est pas suffisant. Il y a un avantage à donner les déviations à la cible à la manière d'un jeu chaud-froid. Avec BI, le participant doit balayer l'espace à la recherche d'informations sonores et doit faire beaucoup de mouvements pour balayer méthodiquement la zone et tomber sur la cible. Alors qu'AP fournit rapidement et sans effort des informations sur une métrique spatiale tridimensionnelle qui sépare le participant de la cible. La première comparaison AP vs BI montre donc que des informations supplémentaires sont bénéfiques par rapport à la sonification la plus parcimonieuse possible.

AP possède plusieurs qualités identifiées par Ziemer et al. (Ziemer, Nuchprayoon et Schultheis 2019 ; Ziemer et Schultheis 2018, 2019a,b), mais ses dimensions ne sont pas orthogonales. En effet, un seul paramètre sonore est utilisé pour coder la direction sur les axes vertical et horizontal. Dans la condition AP, nous avons observé que les sujets « erraient » autour de la cible lorsqu'ils en étaient proches. Ce problème pourrait être dû au manque d'orthogonalité des dimensions. À distance, les variations d'angle tridimensionnelles proviennent essentiellement de mouvements sur un plan horizontal, tandis que les composantes verticales peuvent être négligées ; à proximité, les écarts d'azimut et d'élévation contribuent au même continuum sonore. Une confusion entre l'axe de gravité et le plan horizontal n'est pas cohérente avec la représentation incarnée de l'espace (Graf et Klam 2006 ; Lamy 2006). Les boucles d'action-perception permettent de déduire les composantes le long de l'un ou l'autre axe avec AP, mais au prix d'une charge cognitive supplémentaire. Ce coût a été allégé grâce au mode de sonification DVH.

En effet, dans un deuxième temps, nous avons comparé AP à DVH, une sonification dont les écarts sur l'axe horizontal sont codés sur une échelle continue par la fréquence du son, tandis que les écarts sur l'axe vertical sont codés de manière binaire, par le déclenchement d'un bruit blanc de bonne hauteur. Les résultats ont montré l'avantage de DVH : les participants ont été plus rapides et ont trouvé cette métaphore plus facile que AP. Avec DVH, les deux dimensions sont orthogonales : le participant peut les interpréter séparément. Elles sont séparées en deux flux, qui ne se chevauchent que lorsque la bonne hauteur est atteinte. Le bruit blanc indiquant la bonne hauteur est un flux supplémentaire contenant une information binaire, il n'est en théorie pas nécessaire de détourner son attention de l'autre flux pour l'interpréter. L'utilisation d'informations binaires simplifie la sonification, réduit la quantité d'informations à intégrer et diminue la charge cognitive. La métaphore DVH soulage le participant de la charge cognitive nécessaire à la séparation des dimensions. Cette charge est encore réduite en codant la position verticale de la cible sous forme binaire (Gao et al. 2022), ce qui s'avère suffisant pour guider quelqu'un vers une cible, comme le suggéraient déjà des travaux antérieurs (Marston et al. 2007). Le mode DVH représente ainsi un compromis efficace entre un guidage parcimonieux (informations binaires) et explicite (dissociation des dimensions).

Les résultats de cette seconde comparaison sont cohérents avec ceux de Lokki et al. (Lokki et Grohn 2005), une étude dans laquelle les temps d'atteinte de cible étaient plus courts avec un son indiquant l'élévation de la cible (ajouté au son spatialisé). Cependant, cette expérience a été réalisée en déplaçant un avatar dans un monde virtuel, ce qui rend impossible la comparaison des temps d'atteinte de cible avec notre étude. Au contraire, dans une autre étude (May et al. 2019), les

participants se déplaçaient physiquement dans une pièce de 3*3 m tout en étant immergés dans un environnement virtuel grâce à un casque de réalité virtuelle. La position de la cible sur l'axe horizontal était codée par le son spatialisé et sur l'axe vertical par la hauteur du son. La latence moyenne pour atteindre la cible était de 25 s. Nous avons obtenu des latences comparables, avec 19,70 s pour DVH et 29,85 s pour AP.

3.2.5 Conclusion

Des participants privés de vue atteignent une cible plus rapidement et préfèrent être guidés par des déviations directionnelles par rapport à la cible plutôt que par l'indication de la bonne direction seule. Ces avantages sont renforcés lorsque les dimensions verticale et horizontale sont transmises par deux flux sonores distincts. Dans cette étude, les participants malvoyants se comportent de manière similaire aux participants normo-voyants ayant les yeux bandés.

L'avantage d'une sonification dissociant les écarts verticaux et horizontaux présentés simultanément soulève une nouvelle question : dissocier ces écarts non seulement de manière spatiale mais aussi temporelle pourrait-il apporter un bénéfice ? En effet, les descriptions des stratégies des participants suggèrent qu'ils réduisent l'écart à la cible sur une dimension à la fois. Ils n'auraient donc pas besoin de toutes les informations simultanément. Cette idée s'inscrit dans les conclusions de l'état de l'art et de notre problématique : un dispositif efficace limite la quantité d'information, ce qui économise des ressources cognitives tout en améliorant les performances.

Cette hypothèse sera explorée dans l'étude A3 de ce chapitre (section 3.3). La prochaine étude comparera différentes métaphores sonores codant séparément l'écart horizontal et l'écart vertical, sur deux tâches : l'atteinte d'une hauteur cible et l'atteinte d'une direction cible.

3.2.6 Résumé de l'étude A1

Les dispositifs de substitution sensorielle (DSS) permettent aux personnes privées de vue de percevoir des informations normalement acquises par la vision via une autre modalité sensorielle. Cette étude évalue un DSS conçu pour guider l'utilisateur vers une cible en convertissant des informations spatiales en informations sonores. L'objectif est de concilier précision et économie cognitive, en trouvant un équilibre entre guidage parcimonieux et explicite. Pour cela, nous procédons par comparaison de sonifications par paire, en partant du plus simple puis en enrichissant progressivement la quantité d'information. Nous limitons la charge cognitive en utilisant des métaphores intuitives et en fournissant certains écarts de manière discrète plutôt que continue.

Trois types de sonification ont été comparés : Binary (BI), indiquant uniquement la direction de la cible par un bruit blanc ; Angle-to-Pitch (AP), codant l'écart angulaire 3-D entre la main et

la cible en fréquence continue, en plus du bruit blanc de bonne direction ; et Dissociated Vertical-Horizontal (DVH), dissociant les dimensions horizontale et verticale, en utilisant des variations de fréquence continues pour coder la direction horizontale, et un bruit blanc pour la hauteur.

Les participants, les yeux fermés, devaient atteindre des cibles virtuelles dans un environnement 3-D. Les performances ont été mesurées par le temps nécessaire pour atteindre la cible et évaluées via un questionnaire portant sur efficacité, confort et facilité d'utilisation.

Les résultats montrent que AP est plus rapide et mieux perçu que BI, tandis que DVH améliore encore les performances et la facilité d'usage par rapport à AP. Cette sonification permet de traiter séparément les dimensions verticale et horizontale, réduisant la charge cognitive tout en conservant une précision élevée. DVH constitue ainsi un compromis efficace entre parcimonie et explicitation des informations, démontrant qu'une sonification optimisée peut guider efficacement vers une cible sans surcharge cognitive.

Les résultats de cette étude ont été publiés dans (Fons, Huet, Pellerin, Gerber et al. 2023).

3.3 Étude A2 : Comparaison de métaphores sonores pour l'atteinte d'une cible avec la main sur une dimension.

3.3.1 Introduction

Dans l'expérience précédente, nous avons comparé trois sonifications guidant une personne privée de vue vers une cible. Les résultats ont montré que les performances étaient meilleures lorsque la sonification dissocie explicitement les dimensions verticale et horizontale, en utilisant deux paramètres sonores différents pour coder les deux dimensions (DVH).

Cette dissociation s'aligne sur une stratégie communément adoptée par les utilisateurs : chercher la cible le long d'une dimension spatiale à la fois. Les dimensions étant orthogonales, elles peuvent être interprétées séparément. Afin de créer un dispositif plus efficace qui s'adapte mieux aux limites cognitives et à la stratégie intuitive des utilisateurs, ainsi qu'à la tâche, nous cherchons maintenant à comparer la sonification DVH à une sonification qui dissocie les dimensions spatiales dans le temps, c'est-à-dire qui les présente étape par étape, une dimension à la fois. Comme nous l'avons souligné précédemment, diminuer la quantité d'informations pourrait réduire la charge cognitive et ainsi améliorer les performances.

Dans cette nouvelle sonification, les informations sur les écarts par rapport à la cible dans chaque dimension seront présentées de manière séquentielle, et non pas toutes à la fois. Cette méthode permet d'adapter la quantité d'informations transmises aux besoins de l'utilisateur, afin de réduire la charge cognitive. Elle s'aligne également sur la manière dont les systèmes de guidage vocal, couramment utilisés dans les dispositifs pour malvoyants, fournissent des instructions. En effet, ces systèmes donnent généralement des instructions une dimension à la fois (par exemple, à gauche, à droite, vers l'avant, vers l'arrière). En appliquant ce principe aux métaphores sonores, nous pouvons combiner la clarté intuitive des instructions verbales avec l'efficacité de la sonification.

Pour tester l'efficacité de cette sonification séquentielle - appelée DVHseq - nous la comparerons à notre sonification DVH précédente, qui dissocie les dimensions en deux paramètres sonores et les présente en parallèle (DVHpar). Les deux sonifications utiliseront les mêmes conversions de métriques spatiales en son (mêmes métaphores sonores sur chacune des dimensions), la seule différence étant que les informations sont présentées de manière séquentielle ou parallèle. En mode séquentiel, la dimension dont l'écart est le plus important est donnée en premier, puis on bascule sur une autre dimension quand l'écart est réduit sous un certain seuil. Cependant, notre sonification DVH précédente n'est pas directement compatible avec un mode séquentiel. En effet, en mode séquentiel, pour le codage de la dimension verticale, il n'y a pas de son tant que le doigt n'est pas à la bonne hauteur. Cela pourrait créer une confusion, car le silence pourrait

indiquer soit que le doigt est dans la mauvaise position, soit que le système ne fonctionne pas.

Nous devons donc sélectionner des paramètres sonores compatibles à la fois en mode séquentiel et parallèle. Pour ce faire, nous avons mené deux tests portant séparément sur les dimensions verticale et horizontale. Comme nous le verrons, chaque métaphore a été choisie pour certaines caractéristiques qui devraient en théorie la rendre intuitive, précise, ou les deux. Pour la sonification séquentielle, il est essentiel que les participants puissent facilement passer d'une dimension à l'autre. Par conséquent, les paramètres sonores choisis pour chaque dimension doivent être intuitifs et suffisamment simples pour être rapidement interprétés. Pour évaluer cela, nous avons comparé quatre métaphores sonores qui codent la différence de hauteur entre le doigt et la cible pour la dimension verticale :

- *Continuous* : Cette métaphore convertit continuellement la distance entre le doigt et la cible sur le plan vertical en hauteur. Plus la distance diminue, plus le son devient aigu. Il s'agit de l'un des paramètres sonores les plus couramment utilisés dans les études sur la sonification non spatialisée pour le guidage sonore(Dubus et Bresin 2013). La gamme de fréquences utilisée (110 Hz à 440 Hz) a été choisie pour éviter les sons trop aigus, qui peuvent être désagréables (Kurakata, Mizunami et Matsushita 2013).
- *Discrete* : Un son aigu de fréquence fixe est généré lorsque la cible est au-dessus de la main, et un son grave est généré lorsque la cible est en dessous de la main. Cette métaphore repose sur la tendance naturelle des êtres humains à associer les sons aigus aux positions spatiales les plus élevées et les sons graves aux positions les plus basses (Evans et Treisman 2010 ; Hamilton-Fletcher, Pieniak et al. 2020 ; Pedley et Harper 1959 ; Pratt 1930 ; Rusconi et al. 2006).
- *Shepard* : Cette métaphore génère une illusion auditive de Shepard-Risset, où le son semble monter ou descendre indéfiniment en tonalité. Plusieurs sons espacés d'une octave sont joués simultanément. Au fur et à mesure, les sons les plus graves diminuent progressivement de volume, tandis que les sons les plus aigus augmentent de volume. Cela crée l'effet d'un son dont la fréquence monte continuellement (ou descend continuellement, si on fait le processus inverse). Dans notre métaphore, un son montant est utilisé lorsque la cible se trouve au-dessus de la main, et un son descendant lorsque la cible se trouve en dessous. Ceci est basé sur l'association naturelle entre les changements de fréquence et la direction du mouvement de la main (Maeda, Kanai et Shimojo 2004).
- *Formant* : Dans cette métaphore, la voyelle [ɑ] est générée lorsque la cible est en dessous de la main, et [o] lorsque la cible est au-dessus de la main. Cette métaphore tire parti de l'association naturelle entre ces voyelles et les positions verticales correspondantes en français : la voyelle [ɑ] se rapproche du mot "bas" prononcé [bɑ], et [o] se prononce comme le mot "haut" Cette métaphore ressemble aux instructions verbales, qui sont utilisées dans de nombreux systèmes de guidage (Ahmetovic et al. 2020 ; Hild et Cheng 2014 ; Manduchi et Coughlan 2014 ; Thakoor et al. 2014 ; Troncoso Aldas et al. 2020 ; Zientara et al. 2017) et sont naturelles et intuitives, tout en étant plus rapides à interpréter que des mots entiers.

Au total, nous disposons de quatre métaphores sonores de complexité et de précision variables. La métaphore *Continuous* a l'avantage d'indiquer avec précision la distance par rapport à la cible, grâce aux variations continues de fréquence et à la haute résolution perceptive du son. Les autres métaphores offrent des informations binaires sur la direction de la cible, le codage binaire étant potentiellement plus facile à interpréter que les sons continus. Parmi celles-ci, la métaphore *Shepard* est la plus complexe, mais elle pourrait être très intuitive si les participants peuvent associer la direction du changement de fréquence sonore à la direction du mouvement de leur main vers la cible. Les métaphores *Discrete* et *Formant*, bien que plus simples, devraient néanmoins être intuitives en raison des associations spatiales naturelles et culturelles des sons utilisés.

Pour la dimension horizontale, nous avons comparé trois métaphores sonores codant la différence angulaire entre la direction pointée par la main et la direction de la cible :

- *Continuous* : À l'instar de la dimension verticale, cette métaphore code de manière continue l'écart angulaire en hauteur du son. La fréquence augmente ou diminue en fonction de la distance angulaire entre la main et la cible.
- *Stereo* : Un son constant est généré dans l'oreille droite lorsque la cible est à droite, et dans l'oreille gauche lorsque la cible est à gauche. Cette métaphore est une forme de sonification spatialisée, souvent considérée comme plus intuitive et nécessitant moins de temps d'apprentissage que la sonification non spatialisée. Cependant, comme la sonification spatialisée reproduit des indices acoustiques pour simuler la position de la cible, elle est confrontée aux mêmes limites que la perception spatiale auditive. On sait que la perception d'une source sonore dans le plan horizontal est relativement imprécise (Middlebrooks 2015) (voir section 2.2.2 pour revue sur la localisation d'une source sonore). Cette métaphore présente donc un indice binaire simplifié : un son dans une oreille ou dans l'autre.
- *Continuous + Stereo* : Cette métaphore combine les deux précédentes. La hauteur du son varie continuellement en fonction de la distance angulaire, et le son est présenté soit à l'oreille droite (si la cible est à droite), soit à l'oreille gauche (si la cible est à gauche).

La métaphore *Stereo* présente l'avantage d'indiquer directement la direction de la cible, contrairement à la métaphore *Continuous*. En outre, s'agissant d'un signal binaire, elle peut être plus facile à interpréter que la métaphore *Continuous*. Cependant, la métaphore *Continuous* fournit des informations plus précises en codant la distance par rapport à la cible par la hauteur. La combinaison *Continu + Stereo* a pour but de combiner les avantages des deux métaphores, bien qu'il soit possible que la combinaison crée un excès d'informations, ce qui la rendrait plus compliquée à interpréter que les métaphores individuelles.

Pour les tâches verticales et horizontales, les différentes métaphores présentent des avantages et des inconvénients distincts, allant de métaphores plus précises à des métaphores plus simples et plus intuitives. Toutes ces métaphores ont été créées en tenant compte des correspondances cross-modales (voir section 2.2.3) afin qu'elles soient les plus intuitives possibles. Pour déterminer quelle métaphore est la plus efficace et la plus confortable pour les participants, nous avons utilisé

un paradigme dans lequel les métaphores étaient présentées de manière aléatoire d'un essai à l'autre pour les deux tâches. Cette conception nous permet d'évaluer quelles caractéristiques - simplicité, précision, signaux binaires ou continus, direction ou distance - sont les plus bénéfiques dans le contexte d'une sonification séquentielle.

3.3.2 Matériel et Méthodes

Participants

Deux groupes distincts de participants ont testé les métaphores pour les dimensions verticale et horizontale.

Pour la dimension verticale, 14 participants normo-voyants (âge moyen : 20,3 ± 1,8 ans) ont pris part à l'expérience. Tous les participants étaient des étudiants en psychologie et ont reçu un point bonus pour un cours de leur choix en échange de leur participation.

Pour la dimension horizontale, 11 participants normo-voyants (âge moyen : 23,3 ± 4,5 ans) ont été recrutés. Sept d'entre eux étaient des étudiants en psychologie, qui ont bénéficié de la même incitation sous forme de points bonus. Les autres participants étaient des étudiants d'autres disciplines, recrutés au laboratoire.

Tous les participants ont donné leur consentement éclairé à l'écrit avant de participer à l'étude.

Matériel

Système de capture de mouvement et équipement du participant Les tests ont été réalisés sous la forme d'un jeu virtuel, dans lequel les participants devaient atteindre une hauteur (dimension verticale) ou une direction (dimension horizontale) spécifique avec leur index.

Pour la tâche horizontale, les participants portaient un harnais sur le dos et un pointeur sur le doigt, tandis que pour la tâche verticale, seul le pointeur sur le doigt était utilisé. Ces objets captés par le système de capture de mouvement étaient utilisées pour déterminer la position de deux points anatomiques : le milieu du torse (point O) et l'extrémité du doigt (point P). Le dispositif était le même que celui de la précédente étude (voir section 3.2.2 et Figure 1.1). Le participant portait un casque à conduction osseuse connecté à un boîtier récepteur porté à la ceinture, comme dans toutes les autres études.

Un bouton était fixé sur un pied de microphone à une hauteur de 1,25 m, servant à la fois de bouton de démarrage et de référence pour le calcul de la distance par rapport à la cible.

Cibles Pour la tâche verticale, chaque cible était représentée par un plan horizontal. La largeur du plan et sa distance par rapport au point de départ variaient, ce qui donnait 72 conditions possibles. Il y avait en effet :

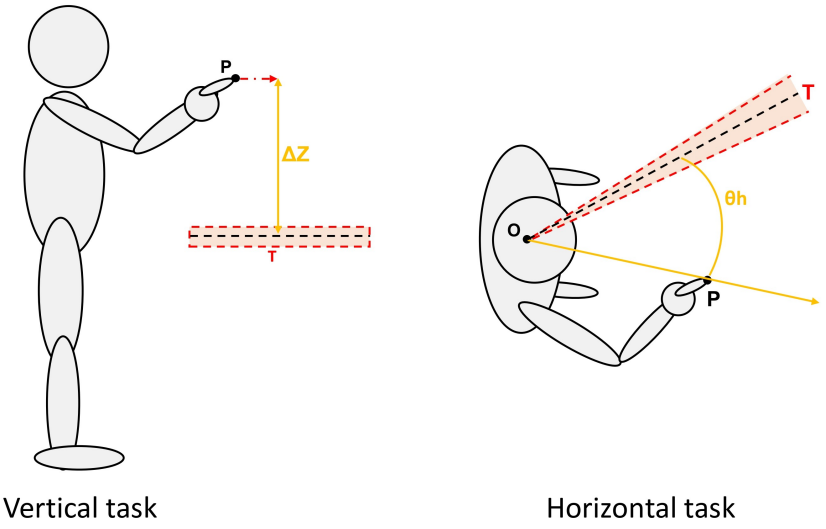


Figure 3.7 – Illustration des métriques spatiales utilisées pour les tâches d'atteinte de cible verticale et horizontale.
θ_h est la projection de l'angle formé par les demi-droites OP et OT sur un plan horizontal. ΔZ est la projection de la distance [PT] sur l'axe vertical Z.

- 6 hauteurs de cible : 0,80 cm, 0,95 cm, 1,10 cm, 1,40 cm, 1,55 cm, 1,70 cm.
- 3 épaisseurs de cible : 3,75 cm, 7,5 cm, 15 cm.
- 4 métaphores : *Continuous*, *Discrete*, *Shepard*, *Formant*.

Chaque participant effectuait 3 blocs de test, chacun contenant 24 cibles (72 essais au total, un pour chaque condition).

Pour la tâche horizontale, chaque cible représentait une direction à atteindre, avec une largeur et un angle qui variaient entre les essais, ce qui donne 54 conditions :

- 6 angles : -90°, -55°, -30°, 30°, 55°, 80°.
- 3 largeurs de cible : 3,75°, 7,5°, 15°.
- 3 métaphores : *Continuous*, *Stereo*, *Continuous + Stereo*.

Chaque participant effectuait trois blocs de test, chacun contenant 18 cibles (54 essais au total, un par condition).

Pour les deux tâches, chaque bloc contenait un nombre égal d'essais pour chaque taille de cible, hauteur ou angle, et métaphore. Les essais étaient présentés dans un ordre semi-aléatoire afin de garantir que la taille, la hauteur et la métaphore varient au cours de deux essais successifs.

Métriques spatiales

Les métriques spatiales pour les deux tâches unidimensionnelles sont représentées dans la figure 3.7.

Pour la dimension verticale, une seule métrique spatiale a été utilisée pour générer les métaphores sonores :

$$\Delta Z = Pz - Tz \tag{3.4}$$

Cela correspond à la projection de la distance entre la cible T et le doigt P sur l'axe vertical Z. Une valeur positive de ΔZ indique que la cible est au-dessus du doigt, tandis qu'une valeur négative indique que la cible est en dessous du doigt.

Pour la dimension horizontale, une seule métrique spatiale a également été utilisée pour générer les métaphores sonores : θ_h. Elle correspond à la distance angulaire entre la direction de la cible $\overrightarrow{OT}$ et la direction torse-index $\overrightarrow{OP}$. Une valeur positive de θ_h signifie que la cible est à gauche de la direction pointée, une valeur négative signifie que la cible est à droite.

Métaphores sonores

Pour la dimension verticale, quatre métaphores sonores ont été créées :

- *Continuous* : on génère une sinusoïde dont la fréquence varie en fonction de ΔZ sur un continuum allant de *fmin* = 110 Hz pour $|\Delta Z|$ = 2 m, à *fmax* = 440 Hz pour $|\Delta Z|$ = 0,01 m. De cette manière, f se situe dans le spectre audible tout en évitant les fréquences trop élevées, qui sont désagréables.
- *Discrete* : si $\Delta Z < 0$, un son aigu à une fréquence constante de 440 Hz est généré. Si $\Delta Z > 0$, un son grave d'une fréquence constante de 110 Hz est généré.
- *Shepard* : une illusion de Shepard-Risset est générée. Si $\Delta Z < 0$, l'illusion est montante : la fréquence du son donne l'impression de monter continuellement. Si $\Delta Z > 0$, l'illusion est descendante : la fréquence du son donne l'impression de descendre continuellement.
- Formant : si $\Delta Z < 0$, alors la voyelle [o] est générée en continu. Si $\Delta Z > 0$, alors la voyelle [ɑ] est générée en continu.

Pour la dimension horizontale, trois métaphores sonores ont été créées :

- *Continuous* : un son pur est généré, dont la fréquence varie en fonction de θ_h, sur un continuum allant de *fmin* = 110 Hz pour θ_h = 180°, à *fmax* = 440 Hz pour θ_h = 0,01°.
- *Stereo* : un son pur d'une fréquence constante de 220 Hz est envoyé dans l'oreille droite lorsque $\theta_h < 0$, et dans l'oreille gauche lorsque $\theta_h > 0$.
- *Continuous + Stereo* : combinaison des deux métaphores précédentes. Un son pur est généré et varie de la manière décrite ci-dessus, soit dans l'oreille droite, soit dans l'oreille gauche, comme pour la métaphore *Stereo*.

Pour les métaphores Continuous et Continuous + Stereo, la conversion de θ_h ou ΔZ (X dans les équations suivantes) en fréquence suit la loi psychophysique de Stevens (S. Stevens 1957)

dont nous avons cité les équations 3.1, 3.2 et 3.3 plus tôt.

Protocole

L'expérience se déroulait dans l'espace de capture de mouvement décrit précédemment. Les participants devaient trouver plusieurs cibles présentées successivement, guidés par les métaphores décrites dans les sections précédentes.

Procédure L'expérience commençait par une phase d'entraînement, au cours de laquelle les participants réalisaient un bloc d'entraînement pour chaque métaphore sonore (quatre dans l'expérience verticale, trois dans l'expérience horizontale). L'ordre des blocs d'entraînement était contrebalancé. Chaque bloc d'entraînement comprenait 18 cibles (3 tailles de cibles * 6 hauteurs ou directions de cibles). Pendant cette phase, l'expérimentateur pouvait communiquer avec le participant pour clarifier les métaphores sonores si nécessaire.

Après avoir terminé l'entraînement, les participants passaient aux trois blocs expérimentaux. Seules les données de ces blocs expérimentaux ont été utilisées dans les analyses. La durée totale de l'expérience était d'environ 30 minutes.

Déroulement d'un bloc expérimental :

1. Le participant, portant l'équipement décrit plus haut, se tient devant le bouton de départ.
2. Le participant appuie sur le bouton, un son indique le début de l'essai.
3. Le feedback sonore est déclenché. Le son varie en fonction de la position du doigt du participant par rapport à la cible, selon la métaphore déclenchée.
4. L'essai se termine quand le doigt du participant entre dans la cible et y reste pendant 400 ms (afin d'être sûr que le participant ait vraiment trouvé la cible et ne soit pas rentré dedans accidentellement), ou quand la limite de temps est atteinte (10 secondes).
5. Si le participant a atteint la cible, un son de victoire est déclenché, sinon un autre son indique la défaite.
6. Le participant déclenche l'essai suivant en appuyant sur le bouton S.

Questionnaire : A la fin de l'expérience, les participants remplissaient un questionnaire. On leur demandait :

- Leur opinion générale sur l'expérience.
- Les difficultés qu'ils ont rencontrées.
- Si des similarités entre certaines métaphores ont apporté de la confusion.
- De donner un score d'ergonomie à chaque métaphore sur une échelle de Likert en 7 points, sur trois critères : efficacité, confort, facilité à utiliser le dispositif.

Analyses

Les analyses ont été réalisées avec RStudio (RStudio 2023.12.1+402) et Jamovi (Jamovi 2.3.29). Pour chaque dimension, nous avons évalué et comparé quantitativement les performances des différentes métaphores de plusieurs manières.

Comme notre variable de réponse, le temps d'atteinte de cible, est une durée avec une distribution asymétrique, nous avons utilisé une analyse de Cox pour comparer les temps d'atteinte de cible entre les conditions. Pour réaliser cette analyse, nous avons utilisé la fonction coxph du package Survival du logiciel RStudio.

Enfin, les métaphores ont été comparées qualitativement, à l'aide des scores d'ergonomie donnés par les participants sur trois critères, sur une échelle de 1 à 7. Nous avons effectué une régression ordinale pour comparer les scores en fonction de la métaphore, en utilisant la fonction clm() du package ordinal du logiciel RStudio.

Des analyses supplémentaires de la trajectoire du doigt pendant l'essai ont été effectuées et seront détaillées dans les sections suivantes.

Nous avons également utilisé la loi de Fitts pour comparer nos métaphores. Nous reviendrons sur ces analyses dans le prochain chapitre (voir section 4.2), qui sera dédié à l'utilisation de la loi de Fitts pour évaluer et comparer des métaphores sonores guidant le geste d'atteinte de cible.

3.3.3 Résultats

Temps d'atteinte de cible

Tâche verticale Le temps moyen pour atteindre la cible était de 3,68 ± 1,45 s pour la métaphore *Formant*, 4,35 ± 2,17 s pour la métaphore *Continuous*, 4,18 ± 2,13 s pour la condition *Discrete*, et 4,46 ± 2,33 s pour la condition *Shepard*. Les temps d'atteinte sont représentés sur la figure 3.8.

Les temps d'atteinte de cible ont été comparés à l'aide d'un modèle de Cox. Les résultats montrent que les temps d'atteinte de cible sont plus courts pour la condition *Formant* que pour la condition *Continuous* (z = 4,12, p <. 001), pour la condition *Formant* que pour la condition *Discrete* (z = 3,28, p < .01) et pour la condition *Formant* que pour la condition Shepard (z = -4,85, p > .001). Les autres comparaisons n'étaient pas significatives.

Tâche horizontale Les temps d'atteinte de cible moyens étaient de 2,76 ± 0,84 s pour la métaphore *Continuous + Stereo*, 3,72 ± 1,74 s pour la métaphore *Continuous*, et 3,69 ± 1,56 s pour la métaphore *Stereo*. Les temps d'atteinte sont représentés sur la figure 3.9.

Les temps d'atteinte des cibles ont été comparés à l'aide d'un modèle de Cox. Les résultats montrent que les temps d'atteinte de cible sont plus courts pour la métaphore *Continuous + Stereo* que pour la métaphore *Continuous* (z = 7,22, p < .001), et plus courts pour la métaphore *Continuous*

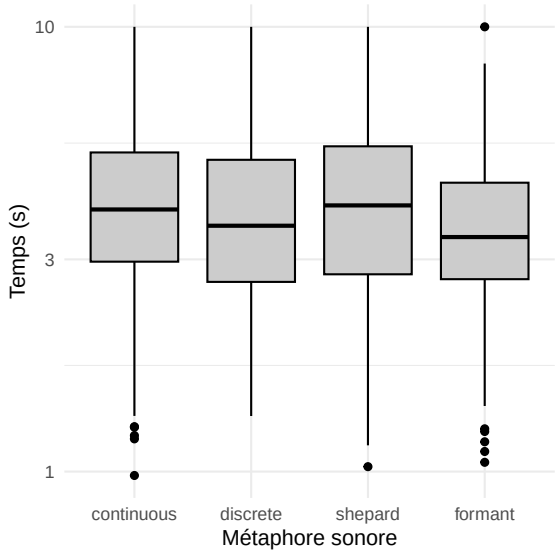


Figure 3.8 – Temps d'atteinte de cible (échelle log10) pour les quatre métaphores testées dans la tâche verticale.
Chaque boîte représente la médiane (ligne horizontale), l'intervalle interquartile (boîte) et les valeurs extrêmes aberrantes (moustaches).

+ Stereo que pour la métaphore *Stereo* (z = 7,22, p < .001). La comparaison entre *Continuous* et *Stereo* n'était pas significative.

Scores

Tâche verticale Nous avons effectué une régression ordinale pour comparer les scores d'ergonomie par métaphore, ce qui n'a révélé aucune différence significative entre les métaphores pour l'un ou l'autre des critères (voir Figure 3.10).

Tâche horizontale Nous avons effectué une régression ordinale pour comparer les scores d'ergonomie par métaphore, ce qui a révélé un meilleur score pour *Continuous + Stereo* que pour *Continuous* et *Stereo*, pour le critère d'efficacité (z = 2,74, p = 0,02). Il n'y a pas de différence significative entre *Continuous* et *Stereo* pour ce critère, et aucune différence entre les trois métaphores pour les critères de facilité et de confort (voir Figure 3.11).

Direction relative du premier mouvement

Tâche verticale En plus des analyses principales, nous avons voulu évaluer si les métaphores aidaient les participants à diriger leur main immédiatement vers la cible. Un test binomial a été effectué pour déterminer si la proportion des premiers mouvements vers ou à l'opposé de la direction de la cible différait de 50 % pour chaque métaphore. Trois métaphores (*Shepard*, *Formant* et *Discrete*) devraient permettre aux participants d'identifier immédiatement la direction de la cible. Nous nous attendions donc à un plus grand nombre de premiers mouvements vers la cible

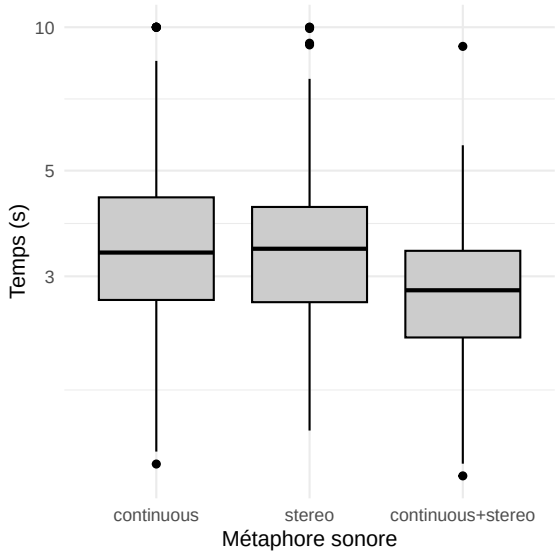
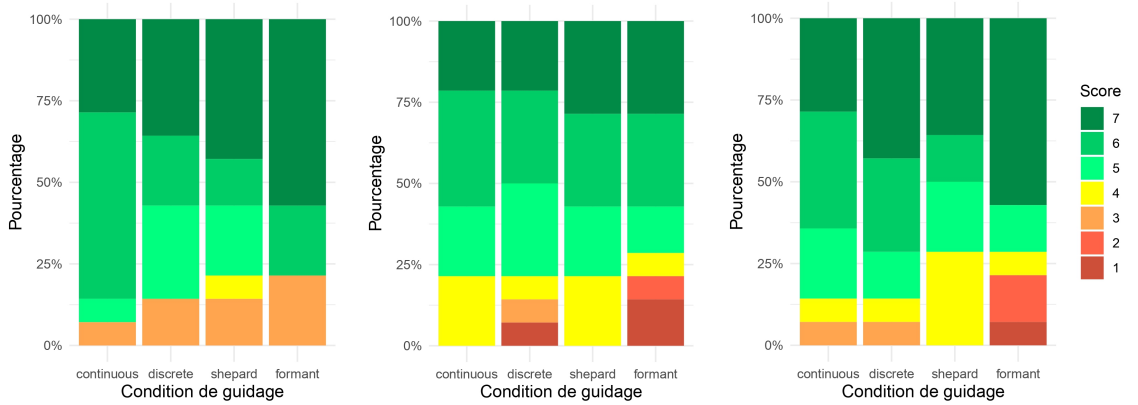
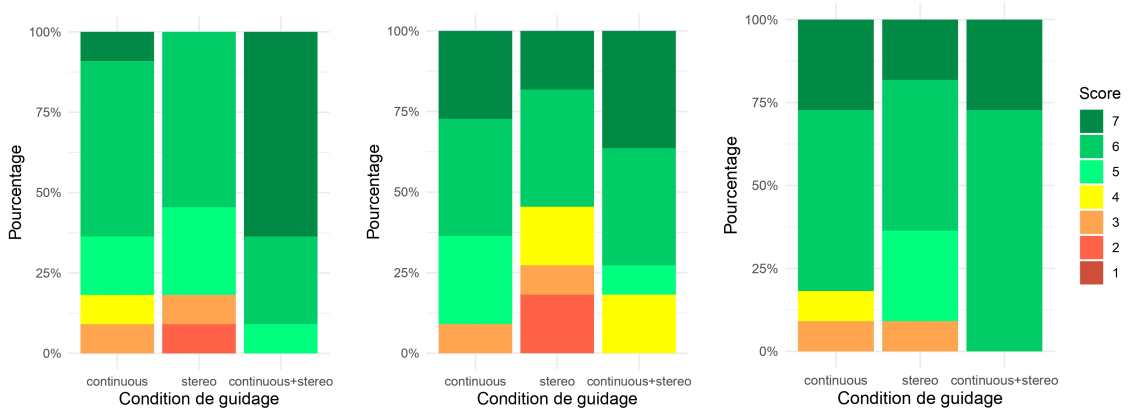


Figure 3.9 – Temps d'atteinte de cible (échelle log10) pour les trois métaphores testées dans la tâche horizontale.
Chaque boîte représente la médiane (ligne horizontale), l'intervalle interquartile (boîte) et les valeurs extrêmes aberrantes (moustaches).



(a) Efficacité **(b)** Confort **(c)** Facilité

Figure 3.10 – Pourcentage de participants ayant donné chaque score pour les quatre métaphores de la tâche verticale. Les trois critères sont pris en compte (efficacité, confort, facilité). Le gradient de couleur représente les niveaux de notation, le rouge correspondant à 1 et le vert foncé à 7, de sorte que les zones plus vertes correspondent à des jugements plus positifs.



(a) Efficacité **(b)** Confort **(c)** Facilité

Figure 3.11 – Pourcentage de participants ayant donné chaque score pour les trois métaphores de la tâche horizontale. Les trois critères sont pris en compte (efficacité, confort, facilité). Le gradient de couleur représente les niveaux de notation, le rouge correspondant à 1 et le vert foncé à 7, de sorte que les zones plus vertes correspondent à des jugements plus positifs.

pour ces métaphores. Un test unilatéral a été utilisé pour ces trois métaphores, et un test bilatéral pour la métaphore *Continuous*. La proportion de premiers mouvements vers la cible était significativement plus élevée que 50 % pour la métaphore *Discrete* (p < .001), mais pas pour les autres métaphores.

Tâche horizontale Un test binomial a été effectué pour évaluer si la proportion des premiers mouvements vers ou à l'opposé de la cible différait de 50 % pour chaque métaphore. Les métaphores *Stereo* et *Continuous + Stereo* permettent normalement aux participants de déterminer immédiatement la direction de la cible, nous nous attendions donc à un plus grand nombre de premiers mouvements vers la cible pour ces métaphores. Un test unilatéral a été effectué pour *Stereo* et *Continuous + Stereo*, et un test bilatéral pour *Continuous*. Les tests binomiaux ont donné $p = 0{,}04$, $p < .001$, et $p < .001$ pour *Continuous*, *Continuous + Stereo*, et *Stereo*, respectivement. La proportion de premiers mouvements vers la cible était significativement plus élevée que 50 % pour les trois métaphores.

Direction absolue du premier mouvement

Tâche verticale Nous avons cherché à vérifier l'existence d'un biais dans la direction du premier mouvement (si les participants avaient tendance à se déplacer vers le haut ou vers le bas, quelle que soit la position de la cible). Un test binomial bilatéral a été effectué pour déterminer si la proportion des premiers mouvements vers le haut ou vers le bas différait de 50 % pour chaque métaphore. Les résultats ont montré que la proportion de premiers mouvements vers le bas était significativement supérieure à 50 % pour toutes les métaphores : *Continuous* ($p = 0{,}001$), *Discrete* ($p = 0{,}006$), *Formant* ($p = 0{,}002$) et *Shepard* ($p = 0{,}001$).

Tâche horizontale Un test binomial bilatéral a été effectué pour déterminer si la proportion des premiers mouvements vers la gauche ou vers la droite différait de 50 % pour chaque métaphore. Les tests binomiaux ont montré que la proportion des premiers mouvements vers la gauche n'était pas significativement différente de 50 %, pour aucune des métaphores.

Changements de direction durant l'essai

Tâche verticale Une ANOVA a été réalisée pour comparer le nombre de changements de direction entre les métaphores. Les résultats ont révélé une différence significative dans le nombre de changements de direction en fonction de la métaphore ($F(3, 1002) = 5{,}02$, $p = 0{,}002$). Les analyses post hoc ont montré qu'il y avait moins de changements de direction avec la métaphore *Continuous* qu'avec la métaphore *Discrete* ($t(1003) = -3{,}84$, $p < .001$). Les autres comparaisons n'étaient pas significatives.

Tâche horizontale Une ANOVA a été réalisée pour comparer le nombre de changements de direction entre les métaphores. Les résultats ont montré une différence significative dans le nombre de changements de direction pendant le test en fonction de la métaphore ($F(2, 591) = 18$, $p < .001$). Les analyses post hoc ont révélé que le nombre de changements de direction était plus faible pour la métaphore *Continuous + Stereo* que pour la métaphore *Continuous* ($t(591) = 4{,}87$, $p < .001$), et plus faible pour la métaphore *Continuous + Stereo* que pour la métaphore *Stereo* ($t(591) = -5{,}48$, $p < .001$). La comparaison entre *Continuous* et *Stereo* n'était pas significative.

Nombre de dépassements de la cible pendant l'essai

Tâche verticale Une ANOVA a été réalisée pour comparer le nombre de dépassements entre les métaphores. Les résultats ont montré une différence significative dans les dépassements en fonction de la métaphore ($F(3, 1000) = 6{,}23$, $p < .001$). Les analyses post hoc ont révélé que les dépassements étaient moins nombreux avec la métaphore Continuous qu'avec les métaphores Shepard ($t(1003) = 3{,}54$, $p = 0{,}002$), *Formant* ($t(1003) = -3{,}24$, $p = 0{,}007$), et *Discrete* ($t(1003) = -3{,}73$, $p = 0{,}001$). Les autres comparaisons n'étaient pas significatives.

Tâche horizontale Une ANOVA a été réalisée pour comparer le nombre de dépassements entre les métaphores. Les résultats ont montré une différence significative dans le nombre de dépassements en fonction de la métaphore ($F(2, 591) = 20{,}6$, $p < .001$). Les analyses post hoc ont révélé que le nombre de dépassements était plus faible pour la métaphore *Continuous + Stereo* que pour la métaphore *Stereo* ($t(591) = -5{,}89$, $p < .001$), et plus faible pour la métaphore *Continuous* que pour la métaphore *Stereo* ($t(591) = -5{,}17$, $p < .001$). La comparaison entre *Continuous + Stereo* et *Continuous* n'était pas significative.

3.3.4 Discussion

Tâche verticale

Nous avons comparé quatre métaphores pour l'encodage de l'écart vertical à la cible dans une tâche d'atteinte de cible 1D.

Les résultats ont montré que la métaphore *Formant* permettait d'obtenir de meilleures performances en termes de temps d'atteinte de cible. Elle encode l'écart à la cible sous une forme binaire et proche de sons verbaux, ce qui la rend facile à interpréter et intuitive. Cependant, les commentaires des participants n'ont pas montré de différences significatives dans l'appréciation des métaphores. Les avis étaient partagés, peut-être en raison de la confusion entre les métaphores pendant la phase de test. Nous avons présenté les métaphores de manière aléatoire afin d'identifier la métaphore la plus intuitive, qui serait la plus facile à interpréter malgré les distractions. Les métaphores ont été conçues pour être distinctes, mais une confusion entre *Discrete* et *Formant* a pu se produire. Bien que les voyelles [o] et [ɑ] aient été générées avec la même fréquence fondamentale, certains participants ont perçu [o] (cible plus élevée) comme étant plus grave que [ɑ] (cible plus basse), la confondant avec la métaphore *Discrete*. En effet, la perception de la hauteur des voyelles n'est pas seulement liée à leur fréquence fondamentale, mais aussi à la position des formants. Un formant est une concentration d'énergie acoustique dans certaines bandes de fréquences spécifiques qui résulte des résonances dans le conduit vocal humain lors de la production de la voyelle. Ils jouent un rôle essentiel dans la reconnaissance des voyelles. Les premier et deuxième formants F1 et F2 de la voyelle [o] sont plus bas que les formants F1 et F2 de la voyelle [ɑ].

Pour voir si une confusion a pu se produire, nous avons analysé les temps d'atteinte des cibles pendant la phase d'entraînement. Les résultats ont montré que les temps d'atteinte de cible étaient plus courts pour *Discrete* que pour les métaphores *Continuous*, *Shepard* et *Formant*. Une différence de temps significative entre les phases d'entraînement et de test n'a été constatée que pour la métaphore *Discrete*, ce qui suggère une confusion avec *Formant*. Cependant, l'analyse des trajectoires a révélé que les participants étaient plus susceptibles de commencer dans la bonne direction avec *Discrete*, contrairement à *Formant*.

La métaphore *Continuous* a permis aux participants de dépasser moins souvent la cible, car elle fournissait une distance par rapport à la cible, ce qui permettait de s'arrêter avec précision. En revanche, les métaphores qui ne fournissent que la direction vers la cible permettent théoriquement de diriger le mouvement directement vers la cible (ce qui n'a pas fonctionné ici en raison de la confusion entre les métaphores), mais pas d'adapter la vitesse. Par conséquent, les participants doivent passer la cible pour entendre un changement de son et connaître sa position. Malgré cela, les performances de la métaphore *Continuous* en termes d'atteinte de la cible sont inférieures à celles de la métaphore Formant (et de la métaphore *Discrete* si l'on tient compte de la phase d'entraînement). La précision de la métaphore *Continuous* n'a pas surpassé la simplicité des trois autres métaphores binaires.

Les participants ont éprouvé des difficultés avec la métaphore *Shepard*, bien qu'elle donne automatiquement la bonne direction. Nombre d'entre eux ont eu du mal avec les fréquences montantes et descendantes, commençant souvent dans la mauvaise direction, et il leur a fallu plus de temps pour corriger leur erreur.

Un biais a également été observé dans la direction du premier mouvement, la plupart des participants se déplaçant vers le bas, peut-être pour des raisons de confort ou de fatigue du bras après avoir appuyé sur le bouton de démarrage. Ce biais peut expliquer pourquoi les participants ne se sont pas dirigés directement vers la cible avec les métaphores *Formant* et *Shepard*.

En conclusion, les sons binaires simples et intuitifs sont préférables aux sons qui varient continuellement, qui sont plus précis mais plus difficiles à interpréter, ou aux sons binaires complexes comme l'illusion de Shepard-Risset. Un contrôle minutieux de la similarité des métaphores est également crucial pour éviter les erreurs d'interprétation.

Tâche horizontale

Nous avons comparé trois métaphores sonores dans une tâche d'atteinte de cible horizontale 1D. L'une des métaphores utilisait un changement de fréquence continu pour représenter la différence d'angle entre la direction pointée et la cible (*Continuous*). Une autre métaphore indiquait la direction relative de la cible de manière binaire en utilisant un son stéréo (*Stereo*), et la troisième combinait ces deux métaphores (*Continuous + Stereo*).

Les résultats ont montré une meilleure performance avec la métaphore *Continuous + Stereo*, y compris des temps d'atteinte de cible plus rapides et de meilleures évaluations de la part des participants. Cela s'est reflété dans l'analyse de la trajectoire du doigt, en accord avec les propriétés de chaque métaphore.

Pour la plupart des essais, les métaphores indiquant la direction de la cible (*Stereo* et *Continuous + Stereo*) ont permis de se diriger directement dans la direction de la cible. Les métaphores indiquant la distance (*Continuous* et *Continuous + Stereo*) réduisaient le nombre de dépassements, ce qui diminuait le nombre de cas où les participants passaient à côté de la cible sans la valider. La métaphore *Continuous + Stereo* combinait ces avantages.

La métaphore *Stereo* est une forme simplifiée de son spatialisé, offrant des indications de direction intuitives et naturelles. Bien que le son spatialisé soit généralement moins précis, la forme binaire utilisée ici permet de détecter plus facilement la position de la cible en faisant passer le son d'une oreille à l'autre.

La métaphore *Continuous*, bien que plus précise, nécessite un effort cognitif plus important pour être interprétée. Elle indique la distance par rapport à la cible, ce qui permet aux participants d'ajuster leur vitesse de déplacement et d'éviter de dépasser la cible. Cependant, elle nécessite un mouvement initial dans une direction aléatoire pour déterminer la position de la cible.

La combinaison de ces métaphores n'a pas submergé les utilisateurs d'informations. Au contraire,

elle a créé une redondance utile, améliorant à la fois l'efficacité et l'expérience de l'utilisateur.

3.3.5 Conclusion

Pour guider la main d'une personne privée de vue le long de l'axe vertical, une métaphore utilisant deux timbres distincts (voyelles) pour indiquer la direction (« haut » ou « bas ») permet d'atteindre la cible plus facilement que d'autres métaphores, comme l'utilisation de deux fréquences distinctes, l'illusion de Shepard montante/descendante, ou une variation continue de fréquence en fonction de la distance à la cible. Cependant, l'analyse des trajectoires montre que la métaphore binaire utilisant deux fréquences favorise la capacité des participants à se diriger spontanément vers la bonne direction, contrairement aux autres métaphores. En parallèle, la métaphore à variation continue de fréquence permet de réduire les dépassements de cible et les changements de direction grâce à un retour permanent sur la distance à la cible.

Pour le guidage sur le plan horizontal, une métaphore combinant codage continu de l'écart directionnel en fréquence sonore et stéréo (son à droite pour une cible à droite, à gauche pour une cible à gauche) s'avère plus efficace que des métaphores utilisant uniquement la stéréo ou uniquement la fréquence. Cette combinaison permet d'atteindre rapidement la direction correcte, de se diriger instantanément vers la cible, de limiter les changements de direction et de réduire les dépassements.

Ces résultats illustrent plusieurs points importants : les sons simples (voyelles binaires ou fréquences binaires) peuvent parfois surpasser des sons complexes (illusion de Shepard), et la combinaison de métaphores complémentaires permet de combiner leurs avantages respectifs (distance continue + direction stéréo) sans générer de surcharge cognitive.

Contrairement à de nombreuses études dans la littérature, nous avons mené des comparaisons empiriques systématiques pour identifier la métaphore sonore la plus efficace et intuitive, et nous avons observé que les différences de trajectoire entre métaphores expliquent à la fois les préférences des participants et les variations de performance mesurées par le temps d'atteinte.

Les enseignements tirés de cette étude serviront à construire deux modes de sonification pour le guidage 3D : un mode séquentiel, transmettant les trois dimensions spatiales avec trois métaphores sonores présentées l'une après l'autre, et un mode parallèle, les présentant simultanément. Ces modes seront étudiés dans l'étude A3 (section 3.4).

3.3.6 Résumé de l'étude A2

Les résultats de l'étude précédente montrent que coder séparément les dimensions verticale et horizontale avec deux paramètres sonores distincts permet un guidage plus efficace et plus

confortable de la main d'une personne privée de vue. L'objectif de cette étude était donc de valider empiriquement l'efficacité de plusieurs métaphores sonores en les testant séparément sur chacune de ces deux dimensions avant de combiner les plus efficaces en une sonification de guidage 3D complète.

Dans cette étude, nous avons testé différentes métaphores sonores sur deux tâches en une dimension.

Pour la tâche verticale (atteindre une hauteur cible), quatre métaphores ont été testées : *Continuous* (variation continue de fréquence selon la distance), *Discrete* (deux fréquences binaires), *Shepard* (illusion de Shepard montante/descendante), et *Formant* (deux voyelles indiquant haut ou bas). Pour la tâche horizontale (atteindre une direction cible), trois métaphores ont été comparées : *Continuous* (écart angulaire codé en fréquence continue), *Stereo* (son dans l'oreille correspondant à la direction), et *Continuous + Stereo* (combinaison des deux).

Les résultats, basés sur les temps d'atteinte, les évaluations subjectives et l'analyse des trajectoires, montrent que *Formant* et *Continuous + Stereo* sont les plus efficaces et ergonomiques. Contrairement à beaucoup d'études dans la littérature, nous avons mené des comparaisons empiriques systématiques. Ces comparaisons ont révélé des différences de trajectoires au cours des essais entre les métaphores. Ces différences peuvent expliquer les préférences des participants et les variations de performance mesurée par le temps d'atteinte de cible. Les métaphores que nous avons identifié comme étant plus efficaces et ergonomiques seront utilisées pour concevoir deux modes de sonification 3D : séquentiel et parallèle (section 3.4).

Les résultats de cette étude ont été publiés dans (Fons, Huet, Pellerin et Graff 2025a).

3.4 Étude A3 : Comparaison d'un mode de sonification séquentiel à un mode de sonification parallèle et à un guidage vocal sur une tâche d'atteinte de cible 3D avec la main

3.4.1 Introduction

Nous poursuivons ici nos travaux sur l'atteinte avec la main d'un objet dans un environnement 3D, guidée par le son. Loin de vouloir remplacer complètement la vue, nous cherchons à développer un système efficace pour assister une personne privée de la vue dans cette tâche spécifique mais générique. Nous voulons créer un système suffisamment précis pour effectuer la tâche rapidement, mais sans créer de surcharge cognitive. En analysant la littérature sur ce type de dispositif de guidage sonore, nous avons soulevé plusieurs limites : le manque de validation empirique des sonifications, et la surcharge cognitive induite par des systèmes trop complexes. Cette étude, dans la continuité des deux précédentes, s'attaque à ces deux limites en proposant des comparaisons systématiques de modes de guidages sonores sur une tâche générique d'atteinte de cible avec la main. Pour diminuer la charge cognitive, les sonifications que nous proposons ont l'avantage de limiter la quantité d'information transmise, de prendre en compte les limites et atouts du système auditif et les limites cognitives de l'humain. En plus de prendre en compte les correspondances cross-modales identifiées dans la littérature, les nouvelles sonifications que nous proposons dans cette étude s'inspirent du son spatialisé (stéréo) et du guidage par la voix (séquentialisation) pour augmenter l'intuitivité du guidage, tout en conservant la précision d'une sonification non spatialisée.

Dans l'étude A1 précédente (section 3.2), nous avons démontré l'avantage de dissocier spatialement les dimensions verticales et horizontales en deux flux sonores présentés simultanément, pour guider des personnes malvoyantes vers des cibles virtuelles (Fons, Huet, Pellerin, Gerber et al. 2023). Cette séparation permet aux participants de se concentrer sur une dimension spatiale à la fois, ce qui leur permet d'atteindre la cible plus rapidement et de manière plus intuitive. Cela suggère que les sonifications présentant les dimensions séparément dans le temps peuvent être plus simples et mieux adaptées à la stratégie la plus intuitive.

Nous avons ensuite comparé diverses métaphores pour guider le mouvement dans les dimensions verticale et horizontale (étude A2 3.3). Les résultats ont montré que l'utilisation de voyelles (métaphore *Formant*) pour le codage de la position verticale améliorait la vitesse, tandis qu'un codage continu de la déviation de la direction en stéréo (métaphore *Continuous + Stereo*) était plus rapide que le codage par la fréquence seule ou la stéréo seule pour la dimension horizontale.

Dans cette étude A3, nous montrons l'avantage d'utiliser un mode de guidage séquentiel. Cette stratégie est comparée à un mode de sonification parallèle. La stratégie parallèle combine toutes

les métaphores sonores codant chaque dimension spatiale en un seul signal ; la stratégie séquentielle décompose ce signal unique en une série de signaux unidimensionnels présentés les uns après les autres. En mettant à disposition toutes les informations nécessaires dans chaque signal, la stratégie parallèle est théoriquement plus rapide, mais elle semble plus exigeante en termes de charge cognitive. Cette stratégie séquentielle est inspirée à la fois des résultats des études précédentes, de nos observations informelles qu'un guide humain semble utiliser spontanément des commandes séquentielles, et des conseils de la littérature d'optimiser la quantité d'informations transmises plutôt que de la maximiser (voir section 2.4). Nous comparons donc également notre sonification séquentielle à un guidage vocal.

Pour démontrer l'avantage de la sonification séquentielle sur la sonification parallèle, les sonifications ne doivent différer que par leur présentation temporelle et avoir les mêmes métaphores sonores (conversions de métriques spatiales en sons) pour chaque dimension. Le choix de ces métaphores a été fait en suivant les conseils donnés dans la littérature pour créer la sonification la plus précise possible, et en suivant les résultats de l'étude précédente (Étude A2, section 3.3) qui a comparé plusieurs métaphores pour transcoder la position verticale et la position horizontale d'une cible (Fons, Huet, Pellerin et Graff 2025a).

Les métaphores retenues sont :

- Azimut : l'écart angulaire entre la direction pointée projeté dans le plan horizontal et la direction de la cible est converti en fréquence sur une échelle continue. Un effet stéréophonique vient renforcer la métaphore : les sons sont diffusés dans l'oreille droite lorsque la cible est à droite, dans l'oreille gauche lorsque la cible est à gauche, dans les deux oreilles quand la cible est correctement pointée.
- Profondeur : la distance entre le doigt et la cible est continuellement convertie en tempo. Le tempo est naturellement associé à la distance, étant utilisé pour signaler la proximité d'un obstacle, par exemple sur le radar de recul d'une voiture. Un tempo élevé est conceptuellement associé à la proximité et à un sentiment d'urgence (May et al. 2019). Il fournit des indications plus précises que l'intensité(Parseihian et al. 2016).
- Hauteur : la voyelle [ɑ] (pour "bas") est générée lorsque la cible est en dessous de la main du participant. La voyelle [o] (pour "haut") est produite lorsque la cible est au-dessus de la main du participant.

En mode séquentiel, ces métaphores sont présentées l'une après l'autre. La direction azimutale est a priori donnée en premier, quand la personne n'est pas tournée dans la bonne direction au début de l'essai, puis la distance, une fois que le participant est dans la bonne direction, puis la hauteur de la cible, une fois que le participant a corrigé à la fois la direction et la distance. Bien entendu, si le participant s'écarte de la bonne direction, cette information sera à nouveau donnée. Il en va de même pour la distance.

Nous avons testé nos différents types de guidage sur une tâche d'atteinte de cible en 3D avec la main. Les participants se déplaçaient dans un espace de capture de mouvement, les yeux fermés,

pour atteindre une cible virtuelle invisible avec leur main, de la même manière que dans l'étude A1.

Dans une deuxième comparaisons, nous testons la même sonification séquentielle contre un guidage vocal. Dans la condition de guidage vocal, un guide donnait au participant des instructions sur la manière d'atteindre la cible. Les conditions de guidage ont été évaluées en fonction du temps nécessaire pour atteindre la cible et à l'aide d'un questionnaire sur les qualités ergonomiques du système.

Ces deux comparaisons nous ont permis de tester quatre hypothèses :

- Une sonification séquentielle permet d'atteindre les cibles plus rapidement qu'une sonification parallèle.
- Les utilisateurs préfèrent la sonification séquentielle à la sonification parallèle.
- Un guidage par sonification séquentielle permet d'atteindre les cibles plus rapidement qu'un guidage vocal.
- Les guides utilisent spontanément des commandes vocales séquentielles, c'est-à-dire qu'ils guident les participants sur une seule dimension spatiale à la fois.

3.4.2 Matériel et Méthodes

Participants

Chaque participant a comparé deux modes de guidage.

20 participants normo-voyants (22 ± 6 ans) ont comparé le mode de sonification séquentiel au mode parallèle.

Vingt participants normo-voyants ont comparé le mode de sonification séquentiel au guidage par la voix (22 ± 7 ans). Les participants ont été associés en dix dyades, dont l'une était le guide pour la condition de guidage vocal, et l'autre le participant devant atteindre les cibles dans les deux conditions de guidage. Par souci de simplicité, nous les appellerons respectivement guide et participant. Leur rôle a été déterminé au hasard, par tirage au sort.

Tous les participants et les guides étaient des étudiants en psychologie et ont reçu un point bonus dans un cours de psychologie expérimentale en échange de leur participation. Tous les participants et les guides ont donné leur consentement éclairé par écrit avant de participer à l'étude.

Matériel

Le système de capture de mouvement était identique à celui utilisé précédemment (voir Figure 1.1). Comme précédemment, les participants étaient équipés d'un harnais et d'une bande sur le doigt captés par le système de capture de mouvement pour détecter la position de leur torse

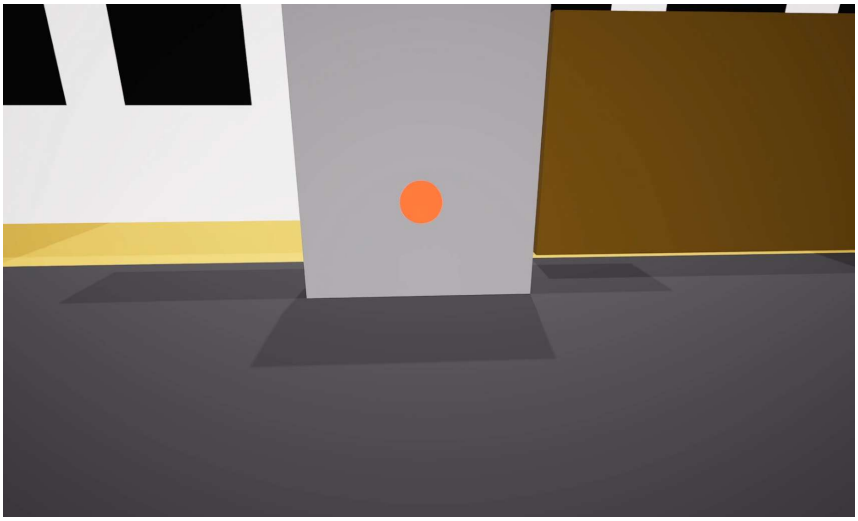


Figure 3.12 – Écran de visualisation de l'environnement virtuel contenant la cible.
Le guide visualise la scène du point de vue du doigt du participant. La sphère orange représente la cible.

(point O) et du bout de leur index (point P). Ils portaient également un casque à conduction osseuse pour entendre la sonification ou les commandes vocales du guide selon la condition, et un clavier miniature dans leur main dominante pour lancer les essais.

Le guide (opérant dans la condition de guidage vocal) était équipé d'un microphone (Boya BY-M1) pour communiquer avec le participant, ce qui lui permettait de donner des instructions verbales. Le participant n'avait aucun moyen de répondre ou de communiquer en retour avec le guide.

Les cibles utilisées ici étaient des sphères de rayon r = 7,5 cm qui pouvaient être positionnée à 27 endroits différents sur une grille 3*3*3 (voir Figure 3.2).

Pour la condition de guidage vocal, le protocole expérimental a été mis en œuvre à l'aide d'un programme Python, sur un second ordinateur, qui gérait la génération des cibles, la validation et l'enregistrement des microphones. Un programme Unreal Engine permettait au guide de visualiser l'environnement virtuel (voir Fig. 3.12), tandis qu'un programme C++ enregistrait les données relatives au torse et à la position des doigts. La communication entre ces composants était facilitée par le protocole Open Sound Control (OSC), qui transmettait des informations sur les cibles, notamment leur position, leurs dimensions et leur état d'acquisition (atteinte ou pas atteinte).

Le guide visualisait l'environnement virtuel contenant la cible, comme si une caméra était fixée sur le doigt du participant, en veillant à ce que la rotation du doigt n'affecte pas l'orientation de la caméra. Le guide opérait dos à la zone de capture de mouvement, sans autre communication que le microphone connecté aux écouteurs du participant. L'angle de vue de la caméra, avec un champ de vision (FOV) de 82°, était déterminé par l'écart angulaire entre le torse et le doigt.

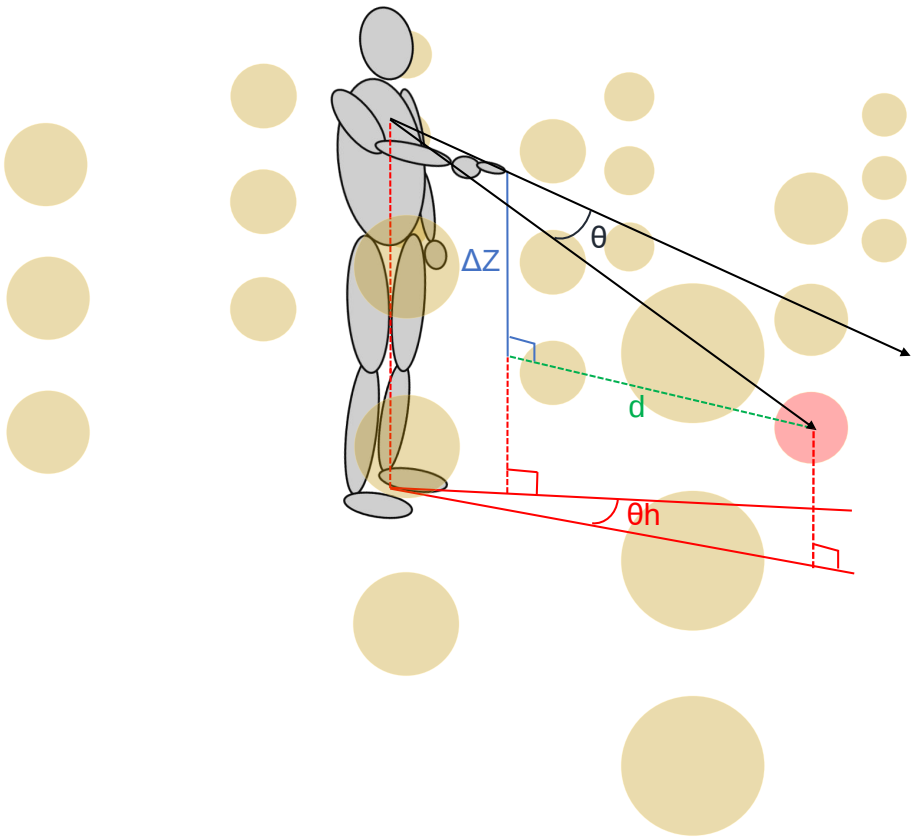


Figure 3.13 – Représentation des positions des cibles (sphères) et des mesures spatiales converties en sons.
Ici, la cible à atteindre (en rouge) est positionnée à la hauteur la plus basse et au milieu d'un petit côté extérieur de la grille. θ_h est la projection de l'angle formé par les droites OP et OT sur le plan horizontal parallèle au sol. d est la projection de la distance [PT] sur le plan horizontal parallèle au sol. ΔZ est la projection de la distance [PT] sur l'axe vertical Z.

Métriques spatiales

Les caractéristiques spatiales extraites de la scène (voir Fig 3.13) étaient déduites du point P, l'extrémité de l'index, du point O, le milieu intérieur du torse, et du point T, le centre de la cible sphérique. Le vecteur $\overrightarrow{OP}$ définit la direction de pointage du participant, à comparer à la direction de la cible $\overrightarrow{OT}$; la distance PT définit la proximité de la cible. Les trois mesures spatiales calculées sont les suivantes :

- L'angle azimutal θ_h : la projection de l'angle formé par les droites OP et OT sur le plan horizontal parallèle au sol (écart angulaire).
- La profondeur (distance) d : la projection de la distance [PT] sur le plan horizontal parallèle au sol (écart en distance).
- La différence de hauteur ΔZ : la projection de la distance [PT] sur l'axe vertical Z (écart en distance).

Tableau 3.4 – Métaphores utilisées pour les deux modes de sonification séquentiel et parallèle

	Horizontale	Verticale	Distance
Fréquence	$\theta_h min$ = 0,01°, $\theta_h max$ = 90° *fmin* = 110 Hz, *fmax* = 330 Hz		
Stéréo	Cible à droite : oreille droite Cible à gauche : oreille gauche Bonne direction : deux oreilles		
Timbre		ΔZ <0 =>[o] ΔZ >0 =>[ɑ]	
Tempo			dmin = 0,01 m, dmax = 6 m Tmax = 600 BPM, Tmin = 200 BPM
Bruit Blanc			P entre dans la cible
Buzzer			P dépasse la cible

Sonification

Ces métriques spatiales étaient ensuite converties en sons, pour les deux modes de guidage par sonification, séquentiel ou parallèle. Les métaphores utilisées étaient les mêmes dans les deux modes de sonification. Dans le mode parallèle, les trois principales métriques spatiales étaient exprimées simultanément dans un son composé. Dans le mode séquentiel, une seule métrique était exprimée à la fois, le choix de la métrique spatiale dépendant de l'ampleur de l'écart par rapport à la cible pour chaque mesure : d'abord, l'écart azimutal était transmis, puis lorsque l'angle azimutal était réduit en dessous d'un certain seuil, la distance par rapport à la cible était donnée, et enfin la hauteur de la cible, une fois que la direction et la distance étaient toutes deux inférieures à un certain seuil.

Les conversions des mesures spatiales en sons sont les suivantes (voir tableau 3.4) :

- L'angle azimutal θ_h : la fréquence varie en fonction de θ_h sur un continuum allant de *fmin* = 110 Hz pour $|\theta_h|$ = 90° à *fmax* = 330 Hz pour θ_h = 3°. En mode séquentiel, le son était un son pur généré seulement quand $|\theta_h|$ > 3°. La conversion de θ_h en fréquence sonore suit la loi de puissance de Stevens, dont nous avons donné les équations 3.1, 3.2 et 3.3 plus tôt.
- La profondeur (distance) *d* : le tempo du son varie en fonction de *d*, sur un continuum allant de *Tmin* = 200 BPM pour *dmax* = 6 m, à *Tmax* = 600 BPM pour *dmin* = 0,01 m. En mode séquentiel, le son indiquant la distance était un son pur généré seulement quand $|\theta_h|$ < 3° et *d* > r (le rayon de la cible = 7,5 cm) + 5 cm, ou quand $|\theta_h|$ < 3° et d < r + 5 cm et ΔZ < r. En d'autres termes, en mode séquentiel, le tempo est activé quand la direction pointée est correcte et la distance à la cible est supérieure à 5 cm, ou quand la direction pointée est correcte et la distance à la cible est inférieure à 5 cm et le doigt est à la hauteur de la cible.
- La différence de hauteur ΔZ : si ΔZ < 0, alors la voyelle [o] est générée. Si ΔZ > 0, alors la voyelle [ɑ] est générée. En mode séquentiel, la voyelle est générée à une fréquence de 110 Hz, quand $|\theta_h|$ < 3° et d < (r + 5 cm) et $|\Delta Z|$ > r.

En mode parallèle, le son était toujours une voyelle ([o] ou [ɑ] en fonction de ΔZ), générée

par battements dont le tempo variait en fonction de d, avec une fréquence sonore variant en fonction de θ_h.

Le r + 5 cm a été introduit après que des pré-tests ont montré qu'il était compliqué pour les participants de trouver la hauteur correcte de la cible en mode séquentiel lorsque les voyelles étaient données uniquement lorsque le doigt se trouvait dans la colonne contenant la cible, car ils avaient tendance à la dépasser dans un sens ou l'autre. Par conséquent, ils perdaient souvent l'information sur la hauteur et perdaient du temps pour la retrouver. Avec ces paramètres, en mode séquentiel, les participants trouvaient la bonne direction, avançaient jusqu'à 5 cm de la colonne contenant la cible, trouvaient la bonne hauteur et avançaient ensuite jusqu'à ce qu'ils atteignent la cible.

Des métaphores sonores supplémentaires ont été utilisées pour les deux modes de sonification : Lorsque le point P du bout du doigt dépassait la cible T (|OP| > |OT|), un buzzer (signal carré) remplaçait tous les autres sons. Lorsque le point P entrait dans la cible, un léger bruit blanc était généré et le son principal s'arrêtait. Un tintement distinct annonçait le début de l'essai, un autre lorsque la cible avait été atteinte, et un troisième lorsque le temps limite avait été atteint.

Protocole

L'expérience durait environ une heure. Elle se déroulait dans l'espace de capture de mouvement décrit ci-dessus. Les participants devaient trouver et atteindre du bout de l'index plusieurs cibles présentées successivement.

Pour la comparaison entre sonification séquentielle vs guidage vocal, le guide était informé qu'il devait diriger le participant vers une position 3D dans un monde virtuel, indiquée par une sphère orange dans l'environnement virtuel visualisé. Il était précisé que le point de vue était situé au bout du doigt du participant. Il était mentionné que la cible pouvait être positionnée derrière le participant au début de l'essai.

Phase de familiarisation : Avant de commencer les essais, une phase de familiarisation permettait aux participants d'explorer les métaphores sonores pour chaque condition de sonification autour d'une cible visible. La cible, une sphère virtuelle de 30 cm de diamètre, était matérialisée physiquement par une sphère réelle faite d'anneaux de bois, superposée à la position de la cible virtuelle. Les participants étaient équipés et pouvaient explorer les métaphores sonores et expérimenter la manière dont le retour sonore était lié à leurs mouvements et à leur position autour et à l'intérieur de la cible. L'expérimentateur leur donnait des instructions pour s'assurer que toutes les métaphores sonores avaient été entendues et comprises avant le test.

Tâche d'atteinte de cible : Les participants, les yeux fermés, effectuaient ensuite la tâche d'atteindre des cibles virtuelles invisibles. La tâche comprenait une phase d'entraînement suivie d'une phase de test. Les deux phases étaient similaires, puisque dans les deux cas, les participants

effectuaient des blocs de 6 essais, c'est-à-dire 6 cibles consécutives à atteindre, positionnées de manière semi-aléatoire : entre deux essais successifs, la position de la cible changeait le long de chacune des trois coordonnées de la grille, de sorte que le participant devait couvrir une distance minimale entre deux cibles (environ 1,88 m). La séquence d'un bloc est décrite ci-dessous.

Les participants commençaient par effectuer autant de blocs d'entraînement qu'ils jugeaient nécessaire pour se sentir à l'aise avec chaque mode de guidage, avant de commencer les blocs de test. Pour chaque condition de guidage, les participants effectuaient deux blocs de test de 6 essais. Seuls les blocs de test ont été utilisés dans les analyses. Les deux conditions de guidage alternaient d'un bloc de test à l'autre, et leur ordre de passage était contrebalancé entre les participants. Les participants pouvaient effectuer des blocs d'entraînement supplémentaires avant chaque bloc de test s'ils le jugeaient nécessaire.

Pour chaque bloc (entraînement ou test), les participants commençaient au centre de la pièce, les yeux fermés. Lorsque l'expérimentateur indiquait que le bloc avait commencé, ils pouvaient appuyer sur le bouton de démarrage du mini-clavier. Un son de démarrage indiquait le début de l'essai, puis le feedback sonore était émis, changeant selon l'un ou l'autre mode de sonification assigné au bloc, et selon le mouvement et la position du participant, comme décrit ci-dessus. La cible était considérée comme atteinte lorsque le point P (bout du doigt) restait à l'intérieur de la cible (indiquée par le bruit blanc) pendant 400 ms, ce qui déclenchait un son de victoire et mettait fin à l'essai. Lorsque le participant n'avait pas atteint la cible dans un délai de 120 secondes, un autre signal sonore de défaite indiquait la fin de l'essai. Après chaque essai, le participant restait sur place, d'où il pouvait lancer l'essai suivant à l'aide du mini-clavier lorsqu'il était prêt. Cette procédure se poursuivait jusqu'à ce que les six essais du bloc de test aient été réalisés.

Les sons de contrôle décrits ici étaient utilisés pour les deux conditions de guidage avec sonification, parallèle ou séquentiel. Pour la condition de guidage vocal, c'était le guide qui lançait les essais et qui informait le participant de la réussite et de l'échec pour chaque essai.

Questionnaire sur l'ergonomie Après la tâche d'atteinte de cible, les participants remplissaient un questionnaire sur l'expérience. Il leur était demandé de donner leur avis général sur l'expérience, les difficultés rencontrées et d'évaluer les qualités ergonomiques des deux modes de guidage, à l'aide d'une échelle de Likert à sept niveaux, sur trois aspects : l'efficacité, la facilité et le confort.

Analyses

Les analyses ont été effectuées à l'aide de RStudio (RStudio 2023.12.1+402).

Le score objectif de performance des participants était le temps d'atteinte de cible, qui a une distribution asymétrique. Nous avons donc utilisé une analyse de Cox pour comparer les conditions. Un modèle de Cox analyse les mesures répétées sans calculer la moyenne des données pour chaque participant, en tenant compte de la variabilité intra- et inter-participants (Letué

et al. 2018). Nous avons utilisé la fonction coxph du paquet Survival du logiciel RStudio pour réaliser cette analyse.

Avant d'effectuer l'analyse de Cox, les données ont été nettoyées en supprimant les temps d'atteinte de cible supérieurs à la moyenne plus deux écarts-types. Cela représentait 4,0 % des données des participants voyants (les données du participant malvoyant ne sont pas incluses dans les analyses, mais sont décrites et comparées à celles des participants voyants) pour la comparaison des deux sonifications, et 3,3 % des données pour la comparaison entre guidage par sonification séquentielle et guidage vocal.

Les participants ont fourni des évaluations subjectives sur trois critères (efficacité, facilité et confort) pour chaque condition de guidage. Chaque évaluation est une variable catégorielle ordinale allant de 1 à 7. Étant donné que les scores d'ergonomie pour les trois critères sont corrélés dans chaque condition, nous avons inclus toutes les évaluations dans un seul modèle de régression ordinale plutôt que d'analyser chaque critère séparément. Cette approche nous permet d'évaluer l'effet global de la condition d'orientation sur les évaluations subjectives sans les agréger en un score unique. L'analyse a été réalisée à l'aide d'un modèle de régression ordinale avec la fonction clm() du package ordinal de RStudio.

Nous avons également analysé les enregistrements audios des commandes vocales des guides afin de déterminer si ces commandes sont séquentielles. Si elles sont séquentielles, il devrait y avoir une majorité de commandes indiquant une seule direction ("avance", "tourne à droite", etc.) et peu de commandes indiquant plusieurs directions à la fois ("descends vers la gauche", "en haut à droite", etc.). Nous avons utilisé le logiciel Whisper pour transcrire les enregistrements audios. Whisper (A. Radford et al. 2023) est un système de reconnaissance automatique de la parole (ASR) développé par OpenAI. Pour l'alignement et la segmentation des phrases, il segmente dynamiquement les transcriptions en fonction des pauses, de la ponctuation et de la structure syntaxique. Il supprime également les hésitations et les pauses de remplissage. Nous avons ensuite puisé dans le vocabulaire présent dans l'état de l'art pour lister des mots-clés permettant de catégoriser les commandes dans un (ou plusieurs) sens en fonction des mots-clés qu'elles contiennent (Choutri et al. 2022 ; Contreras, Ayala et Cruz 2020 ; Fayjie et al. 2017 ; Poncela et Gallardo-Estrella 2015)

Voici les catégories de direction suivies des mots-clés qui déterminent si une commande appartient à chaque catégorie :

- **Avant** : avance, tout droit, devant, en face.
- **Arrière** : recule, derrière.
- **Gauche** : gauche.
- **Droite** : droite.
- **Bas** : bas, descend.
- **Haut** : haut, monte, lever.

Ces mots-clés expriment des mouvements selon les 3 translations dans l'espace 3D. Certaines

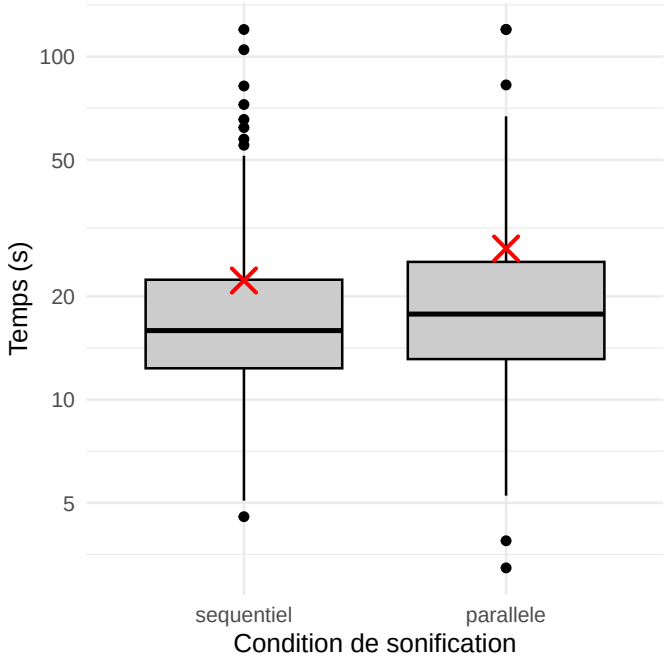


Figure 3.14 – Temps d'atteinte de la cible (échelle log10) pour les deux modalités de sonification séquentiel et parallèle.
Chaque boîte représente la médiane (ligne horizontale), l'intervalle interquartile (boîte) et les valeurs extrêmes (moustaches). Les croix rouges représentent les temps médians du participant malvoyant.

commandes vocales ne correspondent à aucune catégorie de direction : dans ce cas, elles sont classées dans la catégorie "autres" (commandes telles que "un peu plus", "stop", etc.). Nous décrirons ces autres commandes dans la section des résultats.

3.4.3 Résultats

Temps d'atteinte de cible

Guidage par sonification séquentielle vs parallèle Les participants normo-voyants étaient significativement plus rapides avec le mode séquentiel (M = 17,82 s ; SD = 8,38 s) qu'avec le mode parallèle (M = 19,57 s ; SD = 9,23 s) (z = 2,00, p = 0,045) (voir Fig 3.14). Le participant malvoyant a obtenu des résultats similaires, avec un temps moyen d'atteinte de la cible de 27,09 s (SD = 11,66 s) pour le mode parallèle, et de 24,43 s (SD = 13,27 s) pour le mode séquentiel. Sa performance est également représentée sur la Figure 3.14.

Guidage par sonification séquentielle vs guidage vocal Les participants atteignaient les cibles significativement plus rapidement en condition de sonification séquentielle (M = 24,38 s ; SD = 15,40 s) qu'en condition de guidage vocal (M = 34,56 ; SD = 12,89 s) (z = -5,84, p < .001) (voir Fig 3.15).

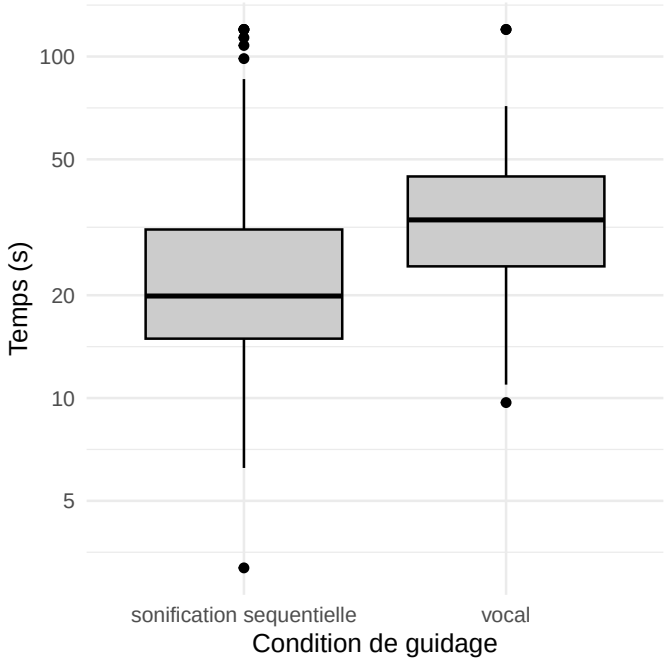


Figure 3.15 – Temps d'atteinte de cible (échelle log10) pour les deux conditions de guidage : sonification séquentielle et guidage vocal.
Chaque boîte représente la médiane (ligne horizontale), l'intervalle interquartile (boîte) et les valeurs extrêmes (moustaches).

Scores d'ergonomie

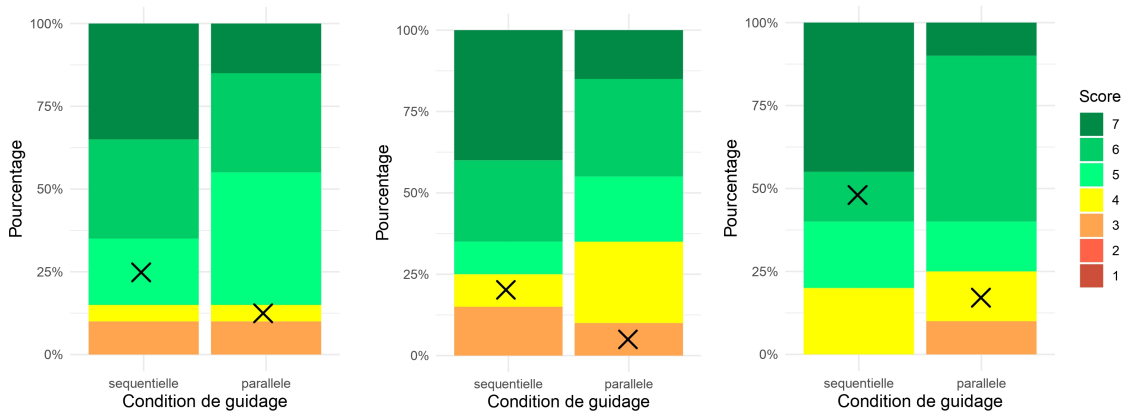
Guidage par sonification séquentielle vs parallèle Les résultats de la régression ordinale révèlent que les participants ont généralement préféré le mode séquentiel de sonification (z = -2,37, p = 0,018). Les notes attribuées par les participants à chaque condition sont représentées sur la Fig 3.16. Le participant malvoyant a également exprimé une préférence pour le mode séquentiel ; les notes qu'il a attribuées pour chaque critère sont représentées sur la Fig 3.16.

Guidage par sonification séquentielle vs guidage vocal Les résultats de la régression ordinale n'ont révélé aucune préférence pour l'une ou l'autre condition. Les scores d'ergonomie donnés par les participants pour chaque condition sont représentés sur la Fig 3.17.

Commandes vocales

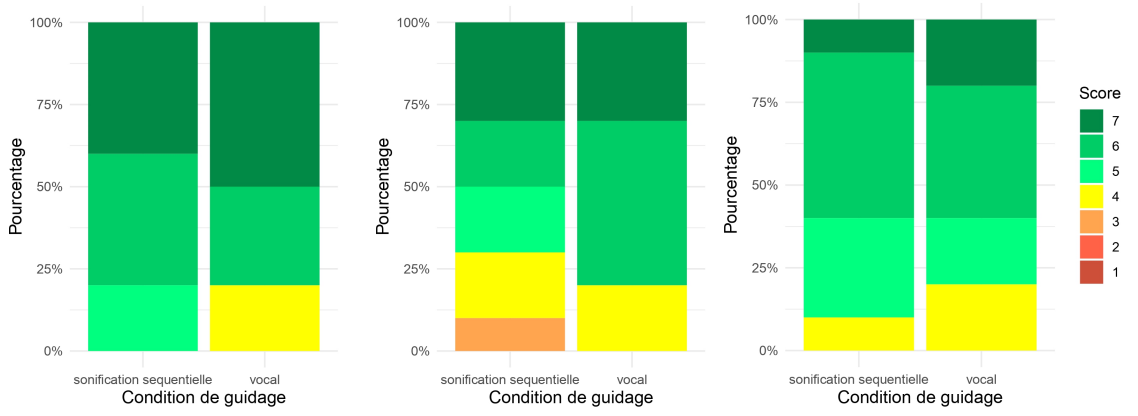
L'analyse des commandes vocales (voir Fig. 3.18) montre que sur un total de 1470 énoncés segmentés et transcrits à l'aide de Whisper, 53,9 % comprenaient des mots-clés exprimant un mouvement dans au moins l'une des six directions.

Une grande majorité (89,3 %) de ces énoncés impliquant des instructions de mouvement ne comprenaient qu'une seule direction, confirmant l'hypothèse de la séquentialité des instructions vocales, même dans le cas d'un guidage spontané par des personnes naïves. Seules quelques



(a) Efficacité **(b)** Confort **(c)** Facilité

Figure 3.16 – Pourcentage de participants ayant donné chaque score pour les deux conditions de sonification, séquentiel et parallèle.
Les trois critères sont pris en compte (efficacité, confort, facilité). Le gradient de couleur représente les niveaux de notation, le rouge correspondant à 1 et le vert foncé à 7, de sorte que les zones plus vertes correspondent à des jugements plus positifs. Les croix noires indiquent le score attribué par le participant malvoyant.



(a) Efficacité **(b)** Confort **(c)** Facilité

Figure 3.17 – Pourcentage de participants ayant donné chaque score pour les deux conditions : sonification séquentielle et guidage vocal.
Les trois critères sont pris en compte (efficacité, confort, facilité). Le gradient de couleur représente les niveaux de notation, le rouge correspondant à 1 et le vert foncé à 7, de sorte que les zones plus vertes correspondent à des jugements plus positifs.

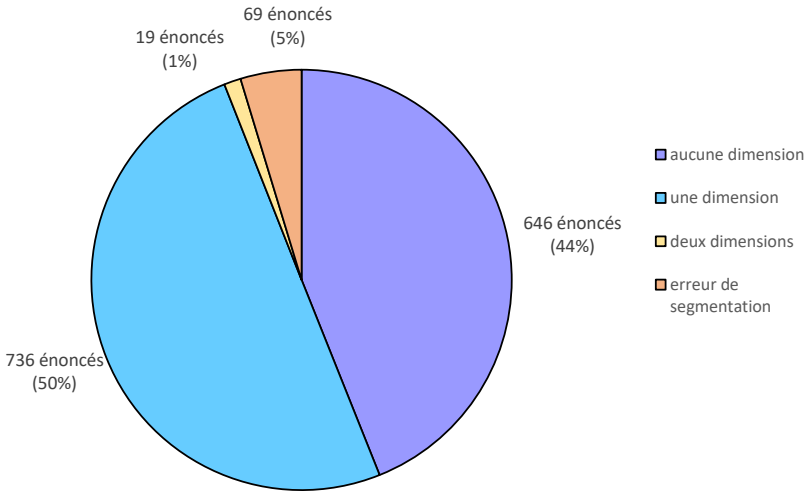


Figure 3.18 – Pourcentage d'énoncés dans chaque catégorie de direction.
1470 énoncés ont été catégorisés. La première catégorie correspond aux déclarations ne comportant aucune catégorie de direction (par exemple, "stop", "encore", etc.). La deuxième correspond aux énoncés ne contenant qu'une seule des six catégories de direction (par exemple « avance », « un peu plus à gauche », etc.). La troisième correspond aux énoncés contenant deux catégories de directions (par exemple « va en haut à droite », etc.). La dernière catégorie représente les erreurs de segmentation de Whisper.

déclarations combinent deux instructions directionnelles telles que "vers la gauche et vers le bas" ou "tout en bas à droite".

Les 69 énoncés restants ont été classés comme exprimant trois instructions de mouvement ou plus. Une vérification manuelle de tous ces énoncés révèle que cette classification est due à des erreurs de segmentation de Whisper qui a agrégé plusieurs instructions successives, bien que de longs silences soient clairement présent.

Il convient de noter le grand nombre d'énoncés qui n'expriment aucune instruction directionnelle. Parmi ces énoncés, environ 40 % correspondent à des ordres simples tels que "stop". Le reste consiste principalement en des réitérations (telles que "encore" ou "plus"), des adverbes d'intensité (tels que "un peu") ou des expressions de validation (telles que "ok", "parfait" ou "génial").

Pour une même classification, les formulations varient d'un participant à l'autre. Par exemple, pour la catégorie direction vers l'avant : "avance", "tout droit", "avance un peu". Pour les affirmations de validation : "super", "ok", "parfait", "c'est bien". La Fig 3.19 représente la distribution des affirmations en fonction du nombre d'occurrences. Nous pouvons constater qu'une grande partie des affirmations sont uniques : la partie la plus basse (à droite) du graphique indique qu'une grande partie des différentes affirmations ne sont prononcées qu'une seule fois. La commande "stop" est celle qui a été répétée le plus souvent, et elle a été dite par tous les guides.

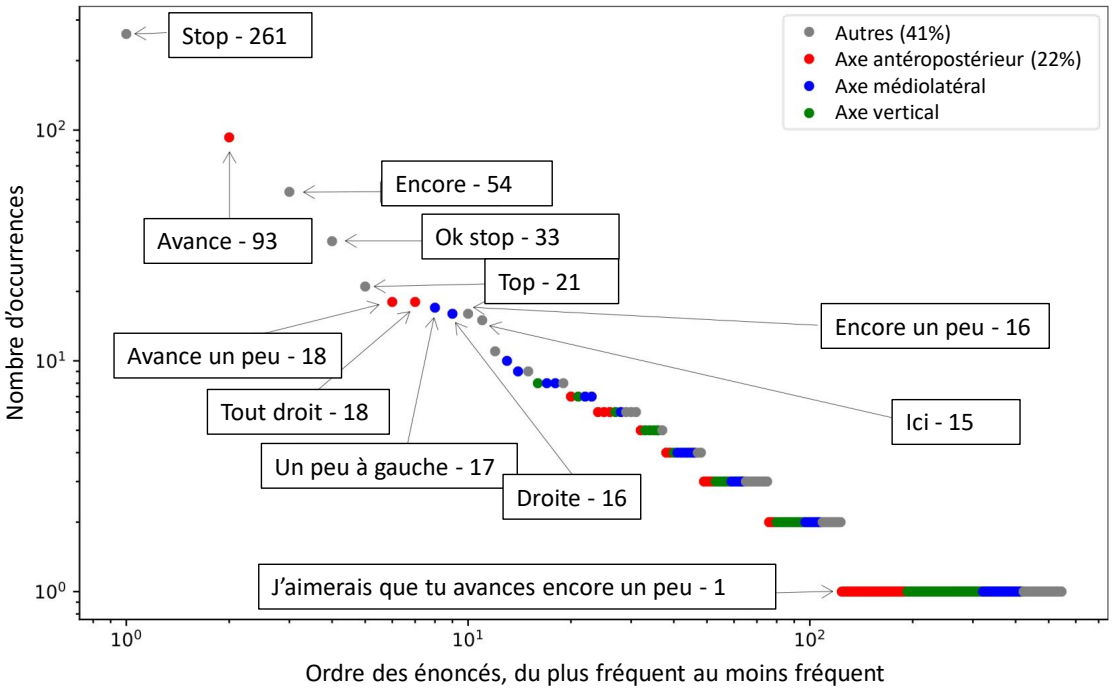


Figure 3.19 – Distribution cumulative des énoncés répétés en fonction de leur occurrence, étiquetés par catégorie de mouvement.
Chaque point représente un énoncé, sa position sur l'axe des y correspondant au nombre de fois qu'il est répété dans l'ensemble de données (100 sont des énoncés dits une seule fois). L'axe des x représente l'ordre des énoncés, du plus fréquent au moins fréquent (100 est l'énoncé répété le plus souvent). La couleur de chaque point indique la catégorie de direction pré-assignée à l'énoncé. L'axe antéropostérieur (en rouge) contient les catégories de direction « avant » et « arrière ». L'axe médio-latéral (en bleu) contient les catégories de direction « gauche » et « droite ». L'axe vertical (en vert) contient les catégories de direction « haut » et « bas ». La catégorie « autre » (en gris) contient divers énoncés sans direction comme « stop », des répétitions comme « encore », des adverbes de degré comme « un peu plus », etc. Les étiquettes contiennent l'énoncé et le nombre de fois qu'il apparaît.

3.4.4 Discussion

Guidage par sonification séquentielle vs parallèle

Nous avons comparé une sonification dissociant les dimensions spatiales à la fois dans l'espace et dans le temps (mode séquentiel) à une sonification dissociant les dimensions dans l'espace uniquement, les dimensions étant présentées simultanément (mode parallèle). Les résultats des tests sur une tâche d'atteinte de cible avec la main ont montré une atteinte plus rapide de la cible avec le mode séquentiel par rapport au mode parallèle. L'analyse des scores subjectifs attribués aux deux modes de sonification par les participants a également montré une préférence pour le mode séquentiel par rapport au mode parallèle, en accord avec les mesures objectives.

Ces résultats démontrent l'avantage d'une stratégie qui délivre une dimension à la fois : le mode séquentiel est à la fois plus efficace (plus rapide) et jugé plus performant, confortable et facile à utiliser par les participants. Il réduit la quantité d'informations à interpréter, ce qui allège la charge cognitive et permet d'atteindre la cible plus rapidement. Le fait d'entendre les variations d'un seul paramètre sonore pour une seule dimension permet de réagir plus efficacement aux variations. Ce mode est donc plus facile à comprendre et à apprendre.

Bien que les résultats globaux montrent un avantage et une préférence pour le mode séquentiel, sur les 20 participants normo-voyants, 5 ont indiqué une préférence pour le mode parallèle. Ces participants ont trouvé le mode parallèle plus précis et plus confortable que le mode séquentiel. Une participante a même trouvé qu'avec le mode parallèle, elle avait l'impression de voir par ses yeux. Le mode parallèle permet d'accéder à toutes les informations simultanément. Il ne restreint pas la stratégie pour atteindre la cible. Ainsi, les participants qui ont apprécié ce mode de sonification pourraient être ceux qui ont su tirer profit des avantages de cette sonification qui leur permet d'avancer vers la cible avec une trajectoire directe s'ils parviennent à tout interpréter en même temps, ou de corriger un écart à la fois comme avec le mode séquentiel, mais dans un ordre décidé par le participant, qui peut être différent de celui imposé par le mode séquentiel.

Cependant, sur ces 5 participants qui ont préféré le mode parallèle, seuls 2 étaient effectivement plus rapides en mode parallèle qu'en mode séquentiel. Cela peut signifier que les participants n'ont pas réussi à se déplacer en ligne droite vers la cible mais qu'ils ont naturellement adopté une stratégie similaire à celle imposée par le mode séquentiel. Cette proposition semble plausible si l'on examine les réponses au questionnaire.

Le questionnaire demandait aux participants de décrire la stratégie utilisée pour trouver la cible avec les deux modes de sonification. Les réponses montrent que sur les 21 participants, 16 (y compris le participant malvoyant) ont indiqué qu'ils adoptaient une stratégie similaire pour les deux modes, c'est-à-dire qu'ils se concentraient sur une dimension à la fois. Seuls 2 participants (dont un qui préférait le mode parallèle) ont indiqué qu'ils essayaient d'interpréter toutes les informations en même temps en mode parallèle. Les trois autres participants ont donné des réponses peu claires ou n'ont pas répondu à la question. Cela signifie que sur 21 participants, un

seul a préféré le mode parallèle ET a obtenu de meilleurs résultats avec ce mode en atteignant la cible par la trajectoire la plus courte, c'est-à-dire en exploitant réellement l'information parallèle contenue dans le son.

Guidage par sonification séquentielle vs guidage vocal

En termes de performances, l'analyse des temps d'atteinte des cibles montre un avantage important de la sonification séquentielle par rapport au guidage vocal par un guide humain. Cette différence peut s'expliquer par le fait que le guidage vocal est beaucoup plus lent. Il y a nécessairement un délai entre deux commandes vocales qui n'existe pas avec la sonification qui peut transmettre les informations de manière continue et fine en utilisant la fréquence sonore par exemple. La sonification est beaucoup plus rapide et plus précise. L'utilisation d'une fréquence qui varie continuellement permet un guidage très précis grâce à la grande capacité du système auditif à discriminer les fréquences (Zwicker et Fastl 2013). Il est donc facile pour le participant de déterminer la direction exacte de la cible. Le fait que la fréquence varie continuellement permet au participant de savoir immédiatement s'il s'est écarté de la bonne direction lorsqu'il se déplace vers la cible, et de corriger rapidement la direction si nécessaire. En revanche, avec le guidage vocal, le participant ne dispose pas d'un retour d'information continu, de sorte que s'il s'écarte de sa trajectoire alors qu'il se dirige vers la cible, il ne le saura qu'à la prochaine instruction du guide. Comme nous le verrons plus loin, cela explique l'utilisation plus fréquente des commandes "stop", "encore", "ok stop", etc. L'utilisation de la stéréo permet au participant de savoir immédiatement et intuitivement si la cible est à droite ou à gauche lorsque l'essai commence. L'utilisation d'une stéréo discrète (droite, gauche ou tout droit) permet de surmonter l'imprécision du système auditif dans la localisation d'une source sonore (Carlini, Bordeau et Ambard 2024 ; Middlebrooks 2015 ; Ziemer et Schultheis 2019a) qui pourrait survenir si on utilise de la HRTF. Il y a donc peu de risques de confusion entre la droite et la gauche.

L'analyse des commandes vocales des guides montre que 50 % de tous les énoncés (et 97,4 % de ceux qui contiennent des directions) ne comprennent qu'une seule catégorie de direction. Cela confirme notre hypothèse selon laquelle un guide qui n'a pas reçu d'instructions spécifiques sur la manière de guider une personne vers une cible utilisera spontanément un mode de guidage séquentiel, comme dans la grande majorité des études qui utilisent des instructions verbales prédéfinies (Ahmetovic et al. 2020 ; Hild et Cheng 2014 ; Hu et al. 2022 ; Manduchi et Coughlan 2014 ; Troncoso Aldas et al. 2020 ; Zientara et al. 2017). La meilleure performance avec la sonification qu'avec le guidage vocal n'est donc pas due au fait que notre sonification est séquentielle (contrairement à la différence de performance observée avec la sonification parallèle), puisque le guidage vocal est également séquentiel.

Outre les commandes vocales directionnelles, les guides ont unanimement utilisé la commande "stop", une manière rapide d'annoncer la fin de la commande directionnelle précédente et de faire la transition avec la commande suivante. Cet arrêt peut être comparé à la transition d'un son à un autre lors d'un changement de dimension en sonification séquentielle, qui est plus

rapide qu'une commande vocale "stop". "Stop" est également un indicateur de séquentialité. Nous ne disons pas "avance" puis "monte", ce qui vous amènerait à continuer d'avancer tout en montant. Nous disons plutôt "avance", "stop", "monte". Les participants ont également utilisé beaucoup d'adverbes d'intensité pour indiquer l'ampleur du mouvement à effectuer et éviter ainsi de dépasser la position cible sur une dimension ou une autre. Dans notre sonification, le tempo et la hauteur indiquent la quantité de mouvement nécessaire pour corriger la déviation en distance et en direction respectivement. Le codage de la hauteur indiquait la direction (vers le haut, vers le bas) grâce aux voyelles, mais pas la distance, ce qui n'est pas un problème car le changement de voyelle d'un côté à l'autre de la cible indique rapidement la hauteur de la cible et permet de changer rapidement de direction si la personne dépasse la bonne hauteur.

En termes de satisfaction perçue, l'analyse de régression ordinale des scores s'ergonomie n'a pas révélé de préférence pour le guidage vocal ou le guidage par sonification. L'avantage de la sonification en termes de rapidité d'atteinte de la cible n'a pas suffi à faire pencher la balance en sa faveur au niveau du ressenti des participants. Le caractère plus naturel et intuitif des commandes vocales a pu gagner la préférence de certains participants, malgré leurs moins bonnes performances. En effet, les notes données par les participants et leurs commentaires qualitatifs reflètent une préférence : ceux qui ont favorisé une condition lui ont généralement donné des notes légèrement plus élevées, même si la différence n'était pas importante. Parmi les 10 participants, 4 ont préféré la condition de sonification, 5 ont préféré la condition de guidage vocal, et 1 seul a donné des notes identiques aux deux conditions et n'a pas exprimé de préférence. Ces préférences se sont reflétées dans les discussions avec l'expérimentateur. Les avis étaient donc partagés. Comparée à la simplicité des commandes vocales, qui ne nécessitent que peu ou pas d'apprentissage de la part du participant, la sonification a pu sembler plus complexe et plus longue à maîtriser pour certains participants, malgré les meilleures performances. Notamment, les voyelles [o] et [ɑ] synthétisées par Pure Data ont nécessité un certain temps d'apprentissage et certains participants avaient du mal à les distinguer. Ce point pourra être amélioré dans les études futures.

3.4.5 Conclusion

Les deux comparaisons menées dans cette étude mettent clairement en évidence les avantages du guidage par sonification séquentielle. Par rapport au mode parallèle, la présentation séquentielle des informations favorise une atteinte de cible plus rapide, plus intuitive et moins coûteuse cognitivement. Comparée au guidage vocal assuré par un guide humain, la sonification séquentielle s'est révélée plus précise et plus efficace, notamment grâce à son retour d'information continu et immédiat. Bien que les instructions vocales reposent elles aussi sur une structure séquentielle, elles restent limitées par leur latence, leur variabilité et l'absence de feedback constant. Ces limites expliquent en partie les performances inférieures du guidage vocal, même si l'inexpérience des guides dans notre étude pourrait également avoir joué un rôle. Des comparaisons avec des guides expérimentés (guides sportifs, instructeurs de locomotion) permettraient

d'affiner cette analyse.

Ces résultats confirment qu'un système de guidage sonore efficace doit respecter les limites perceptives et cognitives humaines, en limitant la quantité d'informations transmises tout en exploitant les atouts du système auditif, comme la discrimination fine des fréquences.

La sonification séquentielle que nous proposons combine la précision et la rapidité des sonifications non spatialisées (fréquence, tempo) avec l'intuitivité inspirée du son spatialisé (stéréo discrète) et du guidage vocal (présentation séquentielle, voyelles). Elle constitue ainsi une solution équilibrée, adaptée aux exigences de la tâche tout en préservant les limites cognitives de l'utilisateur. Plus économique et autonome que le recours à une assistance humaine (ex. applications comme Be My Eyes), elle ouvre la voie à des systèmes de guidage pratiques, intuitifs et validés empiriquement.

3.4.6 Résumé de l'étude A3

Cette étude s'inscrit dans le développement de dispositifs sonores de substitution (DSS) pour guider la main vers une cible en 3D. L'objectif est de proposer un système précis, rapide et économe cognitivement, face aux limites relevées dans la littérature : peu de validation empirique des sonifications et complexité de certains dispositifs. Les résultats de l'étude A1 ont montré qu'il est plus efficace de dissocier les dimensions verticale et horizontale en deux paramètres sonores distincts plutôt que de coder une direction globale, et que les participants privilégient spontanément une stratégie séquentielle, cherchant la cible une dimension après l'autre. De plus, les systèmes de guidage verbal reposent systématiquement sur des instructions séquentielles. Dans ce contexte, nous avons comparé un mode de sonification séquentiel (présentant les écarts dimension par dimension) à un mode parallèle (présentant les trois dimensions simultanément), ainsi qu'à un guidage vocal spontané. Les métaphores sonores retenues proviennent des résultats de l'étude A2 : voyelles [ɑ]/[o] pour la hauteur, fréquence continue + stéréo pour la direction, et tempo pour la distance.

Les résultats montrent un avantage net du mode séquentiel : il permet une atteinte de cible plus rapide et intuitive que le mode parallèle, en réduisant la charge cognitive. Comparée au guidage vocal, la sonification séquentielle s'avère également plus rapides grâce à son feedback immédiat et continu, tandis que les instructions verbales, bien qu'intuitives, souffrent de latences et d'un manque de retour constant.

Les résultats de cette étude ont été soumis à PLOS One et les données sont accessibles sur Zenodo (Fons, Henry et al. 2025).

Synthèse : l'atteinte de cible avec la main guidée par le son

Cette thèse vise à répondre à la question suivante : comment concevoir et évaluer des métaphores sonores pour guider une personne vers une cible sans la vue ? Les trois études de chapitre portaient sur la conception des métaphores sonores de guidage, et avaient pour ambition de répondre à la sous-question : quels métaphores sonores sont les plus efficaces et intuitives pour encoder les trois dimensions spatiales sans dépasser les capacités d'interprétation des utilisateurs ?

Dans les trois études menées pour répondre à ces questions, nous avons pris en compte les limites récurrentes de la littérature sur les dispositifs sonores de guidage vers une cible :

— Le manque de validation empirique des sonifications.
— La surcharge cognitive engendrée par des systèmes trop complexes.

Nous avons fait le choix de concevoir des sonifications précises mais parcimonieuses, en tenant compte à la fois des atouts et des limites du système auditif et des capacités cognitives humaines. Nous avons également cherché à renforcer leur intuitivité en nous appuyant sur des correspondances cross-modales bien établies, mais aussi sur deux sources d'inspiration extérieures : le son spatialisé, via l'usage de la stéréo, et le guidage vocal, via la présentation séquentielle des écarts à la cible et l'utilisation de voyelles.

Pour valider empiriquement ces sonifications, nous les avons comparé paire à paire. Après chaque expérience, la meilleure sonification était testée contre une nouvelle, dans le but d'augmenter la précision tout en conservant l'intuitivité. Les expériences ont été menées sous forme d'un jeu virtuel, où les participants, voyants yeux bandés et déficients visuels, se déplaçaient puis touchaient de la main la cible virtuel.

L'étude A1 a montré qu'il était plus efficace de dissocier les dimensions verticale et horizontale en deux paramètres sonores distincts, plutôt que de coder une direction globale. Ce résultat allait dans le sens de la stratégie spontanée des participants, qui cherchaient la cible une dimension après l'autre, et il confirmait l'intérêt de réduire la complexité de la sonification pour diminuer la charge cognitive. Dans l'étude A2, nous avons validé empiriquement plusieurs métaphores sonores testées séparément en 1D. Les voyelles [a]/[o] pour la hauteur et la combinaison fréquence continue + stéréo pour la direction se sont révélées particulièrement efficaces et intuitives. Ces métaphores ont ensuite servi de base aux sonifications 3D utilisées dans l'étude A3. Enfin, l'étude A3 a permis de comparer deux modes de sonification - séquentiel et parallèle - ainsi qu'un guidage vocal spontané. Les résultats ont montré un avantage net du mode séquentiel, plus rapide, plus intuitif et plus précis que le mode parallèle, et plus efficace que les instructions verbales grâce à son retour immédiat et continu.

En conclusion, dans la conception d'un système de guidage par le son, il y a un inté-

rêt à dissocier les trois dimensions spatiales, à la fois par l'utilisation de métaphores sonores distinctes, et par la présentation séquentielle de ces métaphores plutôt que parallèle. Par rapport à la première sonification testée, cette dernière permet de réduire les temps d'atteinte de cible de plus de moitié. Le guidage est aussi rapide que celui d'un guide vocal humain dans les mêmes conditions.

Au-delà de leur intérêt théorique, ces résultats ouvrent des perspectives d'applications pratiques. Avec les progrès rapides des technologies de détection et de localisation, on peut envisager que ce type de guidage sonore trouve place dans des applications embarquées sur smartphone. Celles-ci pourraient assister une personne non-voyante dans des tâches simples mais fréquentes du quotidien : localiser un produit précis dans un rayon de supermarché, trouver le bon bouton d'interphone, ou encore identifier une entrée de bâtiment une fois le GPS ayant indiqué de manière peu précise que la destination est atteinte.

Enfin, ce chapitre souligne une limite méthodologique importante : la difficulté de comparer directement nos résultats à ceux de la littérature. Si le temps d'atteinte de cible est une mesure largement partagée, la variété des configurations de tâches et de cibles rend les comparaisons difficiles. Le prochain chapitre s'attaque à cette limite, en proposant d'utiliser la loi de Fitts comme métrique standardisée, permettant de comparer les dispositifs testés avec diverses configurations de cibles.

4. Série expérimentale B : Validation empirique de sonifications avec la loi de Fitts pour le guidage de la main vers des cibles 3D

Sommaire

4.1 Introduction 88

4.2 Étude B1 : Comparaison de sonifications parcimonieuses et explicites avec la loi de Fitts, sur une tâche d'atteinte de cible 3D avec la main . . 90

4.2.1 Introduction 90

4.2.2 Matériel et Méthodes 92

4.2.3 Résultats 97

4.2.4 Discussion 99

4.2.5 Conclusion 101

4.2.6 Résumé de l'étude B1 102

4.3 Étude B2 : Comparaison d'un mode de sonification séquentiel vs parallèle avec la loi de Fitts, sur une tâche d'atteinte de cible 3D avec la main simplifiée 103

4.3.1 Introduction 103

4.3.2 Matériel et Méthodes 106

4.3.3 Résultats 108

4.3.4 Discussion 109

4.3.5 Conclusion 110

4.3.6 Résumé de l'étude B2 110

4.1 Introduction

Cette thèse vise à répondre à la question suivante : comment concevoir et évaluer des métaphores sonores pour guider une personne vers une cible sans la vue ? Les deux études de ce chapitre portaient principalement sur l'évaluation du guidage sonore, mais également sur la conception des métaphores sonores. Ce chapitre permet donc de répondre à deux sous-questions :

1. Quels paramètres sonores sont les plus efficaces et intuitifs pour encoder les trois dimensions spatiales, sans dépasser les capacités d'interprétation des utilisateurs ?
2. Comment évaluer un dispositif de guidage ?

Les études précédentes ont permis d'identifier des principes de conception efficaces pour le guidage sonore de la main, mais une limite demeure : l'absence de standardisation dans l'évaluation des dispositifs. En effet, la plupart des travaux utilisent le temps d'atteinte comme indicateur, mais les tâches expérimentales diffèrent trop (types de cibles, configurations) pour permettre des comparaisons directes.

Ce chapitre vise à répondre à cette limite en appliquant la loi de Fitts comme cadre standardisé. Déjà validée pour le pointage visuel, cette loi relie la difficulté de la tâche (Indice de Difficulté, ID) au temps de mouvement (MT, le temps d'atteinte de cible), et permet de calculer une mesure d'évaluation, le *throughput* (TP). Son application à la sonification non spatialisée est inédite et ouvre la voie à une comparaison des résultats entre études.

Deux objectifs guident ce chapitre :

1. Introduire une métrique standardisée (TP de Fitts) pour évaluer les dispositifs de guidage sonore et faciliter la comparaison avec la littérature.
2. Répliquer les résultats des comparaisons de sonifications du chapitre précédent, afin de tester leur robustesse avec une tâche adaptée au modèle de Fitts.

Ce chapitre s'attaque donc aux trois limites des DSS identifiés dans la littérature :

— Le manque de validation empirique systématique des sonifications : nous comparons systématique plusieurs modes de guidage sonore pour identifier les plus efficaces et intuitifs (réplication série expérimentale A)
— La surcharge cognitive induite par des systèmes trop complexes : nous partons du codage le plus simple puis nous l'enrichissons au fur et à mesure, en testant à chaque étape l'intérêt de cet ajout d'information. Les métaphores sonores sont choisies pour leur simplicité (codage binaire plutôt que continu), leur intuitivité (e.g. utilisation de la stéréo et de sons verbaux), et leur lien avec des associations crosse-modales connues (e.g. une fréquence montante = il faut monter la main) (réplication série expérimentale B).
— L'absence de standardisation de l'évaluation des DSS, qui complique la comparaison entre études : nous proposons d'utiliser une métrique issue de la loi de Fitts, le *throughput* (TP) pour évaluer et comparer les dispositifs.

La tâche expérimentale sur laquelle sont testés les sonifications est ici simplifiée par rapport aux études de la première série expérimentale : l'utilisateur est assis, les cibles sont à portée de main, et leur taille ainsi que leur distance varient pour manipuler l'Indice de Difficulté et calculer le TP (voir Figure 4.1).

Deux études expérimentales sont menées dans ce chapitre.

— Étude B1 (section 4.2) : Comparaison de sonifications parcimonieuses et explicites avec la loi de Fitts, sur une tâche d'atteinte de cible 3D avec la main. Cette étude compare la sonification AP qui encode l'écart angulaire en fréquence sonore et la sonification DVH qui dissocie les écarts horizontaux et verticaux en deux paramètres distincts.
— Étude B2 (section 4.3) : Comparaison d'un mode de sonification séquentiel vs parallèle avec la loi de Fitts, sur une tâche d'atteinte de cible 3D avec la main. Cette étude compare un mode de sonification séquentiel, où les écarts sont présentés les uns après les autres, à un mode parallèle, où ils sont transmis simultanément.

4.2 Étude B1 : Comparaison de sonifications parcimonieuses et explicites avec la loi de Fitts, sur une tâche d'atteinte de cible 3D avec la main

4.2.1 Introduction

Cette première expérience visait à vérifier si la loi de Fitts (définie sur des trajectoires 1D ou 2D) pouvait s'appliquer à l'atteinte d'une cible en 3D à l'aveugle guidée le son, en comparant les sonifications testées dans l'étude A1 (section 3.2). Si la loi de Fitts peut s'appliquer, elle permettrait de comparer les dispositifs de manière objective et reproductible, indépendamment de la taille ou de la distance des cibles. Pour cela nous avons adapté la tâche de l'expérience A1 pour pouvoir utiliser la loi de Fitts, en la limitant à l'atteinte de cible avec la main, sans phase de navigation. Nous faisons varier la taille des cibles et la distance à la cible.

La loi de Fitts a déjà été utilisée pour décrire les performances d'un système de guidage sonore vers des cibles 3D (Charoenchaimonkon, Janecek et al. 2010 ; Hu et al. 2022 ; J. C. Lock et al. 2020 ; Marentakis et S. A. Brewster 2006). Cependant, il n'est pas certain que la loi de Fitts s'applique à l'atteinte d'une cible 3D si on utilise uniquement une sonification non spatialisée pour transmettre les écarts entre la position de l'utilisateur et celle de la cible. En effet, dans Charoenchaimonkon, Janecek et al. 2010, le retour auditif n'était activé que lorsque le pointeur entrait ou sortait de la cible (signal binaire). J. C. Lock et al. 2020 et Marentakis et S. A. Brewster 2006 n'ont utilisé que la sonification spatialisée. Hu et al. 2022 ont testé l'algorithme de sonification non spatialisée "the vOICe". Cependant, l'expérience n'a pas montré que la loi de Fitts pouvait être appliquée à cette sonification, car les participants ne parvenaient généralement pas à trouver les cibles. On ne sait toujours pas si la loi de Fitts s'applique à l'atteinte d'une cible 3D à la fois en l'absence de retour visuel et en utilisant des indices auditifs non spatialisés pour convertir la déviation vers la cible. Si la loi s'applique, alors les nombreuses expériences effectuées indépendamment les unes des autres pourraient être comparées entre elles.

Les deux sonifications non spatialisées que nous testons ici sont similaires aux deux sonifications comparées dans notre première étude, AP (Angle-to-Pitch) et DVH (Dissociated Vertical and Horizontal). La seule modification est que nous avons ajouté un codage de la distance, qui pour les deux modes de sonification est convertie en intensité sonore. Dans l'étude précédente les participants avaient pour seul codage de la distance un signal sonore indiquant quand ils avaient dépassé la cible. L'ajout d'un codage continu de la distance en intensité peut être exploité par les participants pour ajuster la vitesse de leur mouvement à la distance restante pour éviter de dépasser la cible et de devoir retourner en arrière.

Avant d'effectuer la comparaison entre les deux modes de sonification, nous avons effectué

une expérience préliminaire pour vérifier que la loi de Fitts pouvait bien être utilisée pour caractériser la relation entre la difficulté de la tâche et le temps mis pour atteindre la cible dans une tâche d'atteinte de cible utilisant un guidage par sonification non-spatialisée. La sonification DVH était testée par les participants :

(1) La position horizontale est exprimée par l'écart angulaire entre la direction pointée par la main et la direction de la cible, projeté sur le plan horizontal. Cet angle est transcodé en hauteur du son sur une échelle continue (plus l'angle est aigu, plus la hauteur est élevée, et vice versa) ;
(2) La position verticale est exprimée par l'écart de distance entre la hauteur de l'index et la hauteur de la cible. Un bruit blanc est superposé au flux auditif principal (dans le même canal) lorsque le doigt est à la même hauteur que la cible ;
(3) La profondeur est exprimée par l'écart de distance entre le doigt et la cible. Cette distance est traduite en intensité sonore sur une échelle continue (plus la distance est courte, plus le son est fort).

Ensuite, lors de l'étude de comparaison, les participants testaient également AP :

(1) La position horizontale et verticale est exprimée par l'écart angulaire global entre la direction pointée par la main et la direction de la cible. Cet angle est transcodé en hauteur du son sur une échelle continue (plus l'angle est aigu, plus le son est aigu, et vice versa) ;
(2) La bonne direction est exprimée par un bruit blanc qui est superposé au flux auditif principal lorsque le doigt pointe la bonne direction ;
(3) La profondeur est exprimée par l'écart de distance entre le doigt et la cible. Cette distance est traduite en intensité sonore sur une échelle continue (plus la distance est courte, plus le son est fort).

Pour tester ces sonifications et les évaluer grâce à la loi de Fitts, nous avons utilisé une tâche similaire à celle utilisée dans la série expérimentale A, mais simplifiée à la seule atteinte de cible avec la main, sans la phase de marche vers la cible. Nous avons remarqué dans la première série expérimentale que les participants n'avaient aucun mal à approcher de la cible, et que la complexité de la tâche se trouvait dans cette deuxième phase de tâtonnement pour toucher la cible. Cette phase est donc la plus pertinente pour comparer des modes de sonification. De plus, elle nous permet ici d'appliquer la loi de Fitts, qui dit que la rapidité d'un mouvement de pointage d'une cible dépend de la taille de la cible et la distance entre le point de départ et le centre de la cible. Dans notre tâche, la cible était donc toujours à portée de main, et nous avons fait varier sa taille et la distance au point de départ de l'index du participant.

Comme pour l'étude A1 (voir section 3.2), nous nous attendions à trouver de meilleures performances pour la sonification DVH qui dissocie les dimensions verticales et horizontales. Ces meilleures performances devraient se refléter à la fois dans le temps d'atteinte de cible et dans le *throughput* (TP) calculé grâce à la loi de Fitts, par l'équation 2.5. Nous avons utilisé la version

originale de la loi de Fitts (équation 2.2), pas celle de Mackenzie (pour discussion sur ce sujet, voir section 2.5.2).

4.2.2 Matériel et Méthodes

Participants

Expérience préliminaire Six normo-voyants (22,0 ± 3,6 ans) ont participé à l'expérience préliminaire. Deux d'entre eux étaient des étudiants en psychologie et ont reçu un bonus d'un point sur un cours de leur choix en échange de leur participation. Les quatre autres participants étaient des étudiants d'autres disciplines recrutés par le laboratoire sur une base volontaire. Nous avons également recruté deux participants malvoyants (57,0 ± 9,9 ans) en dehors du laboratoire. Tous les participants ont donné leur consentement éclairé à l'écrit avant de prendre part à l'étude.

Comparaison entre AP et DVH Cinq normo-voyants ont participé à l'expérience (19,2 ± 0,9 ans). Quatre d'entre eux étaient des étudiants en psychologie qui ont reçu un point de bonus pour un cours de leur choix en échange de leur participation. Le dernier a été recruté en dehors du laboratoire. Tous ont donné leur consentement éclairé à l'écrit avant de participer à l'étude.

Une analyse a priori a été réalisée à l'aide de G*Power version 3.1.9.7 pour l'estimation de la taille de l'échantillon, sur la base des données de l'expérience préliminaire. Avec un critère de significativité de α = 0,05 et une puissance de 0,95, la taille minimale de l'échantillon nécessaire avec la taille de l'effet obtenue lors de l'expérience préliminaire (voir section 4.2.3) est N = 5. La taille de notre échantillon est donc adéquate.

Matériel

Les tests prenaient la forme d'un jeu virtuel dans lequel les participants devaient atteindre des sphères dans un espace 3D avec leur index.

Comme pour les études 2 (section 3.3) et 3 (section 3.4), nous avons utilisé un système de capture de mouvement Vicon, et le même logiciel de contrôle C++ que pour les études précédentes (voir Figure 1.1). La position du doigt et du torse étaient acquises par le système, et la position de ces points anatomiques par rapport à la position de la cible étaient utilisés pour calculer les métriques spatiales.

Chaque cible était une sphère de 5,33 cm, 8 cm ou 12 cm de diamètre, selon la condition. Les cibles pouvaient être placées à 12 cm, 18 cm ou 27 cm du bouton déclenchant le début d'essai (voir figure 4.1) ce qui donne un total de 9 combinaisons de distance et de taille de cible (voir tableau 4.1. Il y avait 32 positions différentes pour chaque distance. Chaque participant effectuait 3 blocs d'essai différents pour chaque taille de cible, soit un total de 9 blocs d'essai. Un bloc de test consistait en une série de 16 cibles consécutives à atteindre, sélectionnées de manière semi-

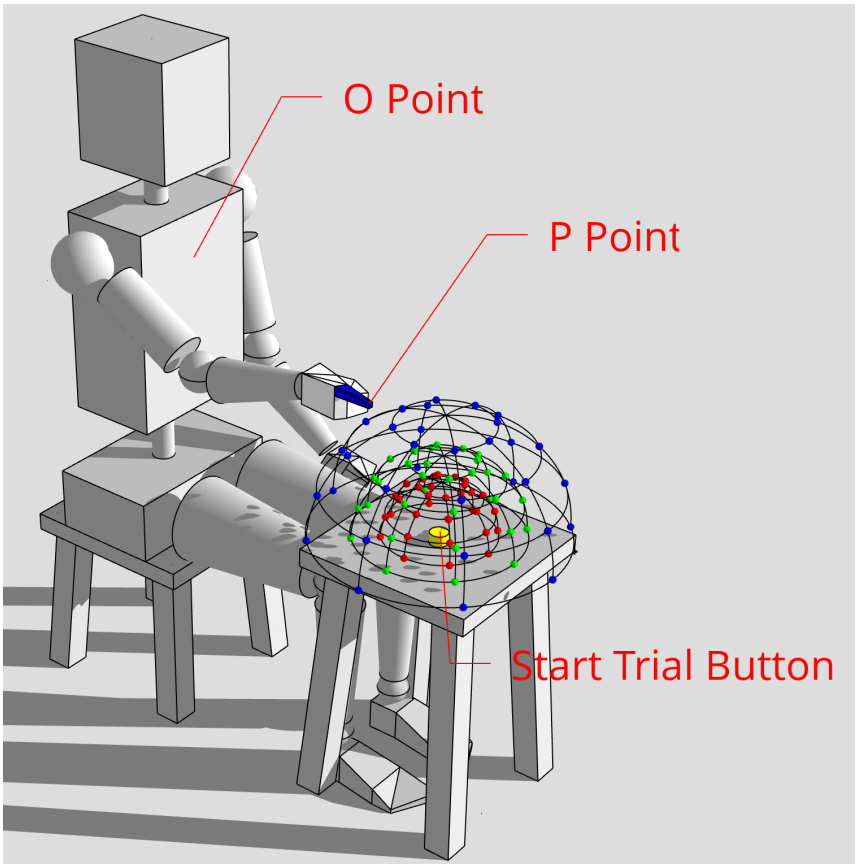


Figure 4.1 – Illustration du dispositif expérimental utilisé pour la tâche d'atteinte de cibles de type Fitts.
Le vecteur $\overrightarrow{OP}$ définit la direction de pointage du participant. La distance à la cible est calculée de P au centre de la cible T. Les cibles peuvent être placées à 12 cm (sphères rouges), 18 cm (sphères vertes) ou 27 cm (sphères bleues) du point de départ (« start trial button »). Le diamètre des cibles varie entre les essais : 5,33 cm ; 8 cm ; 12 cm.

aléatoire afin que le participant atteigne la moitié de toutes les cibles possibles, avec un nombre égal de cibles à chaque distance, sans répéter la même cible deux fois pour une même taille de cible. La séquence garantissait que la distance à la cible variait entre chaque essai.

Le participant était équipé de deux jeux de capteurs fixés sur un harnais dans le dos et sur le doigt, permettant de capter la position du milieu son torse et du bout de son index. Il portait également un casque à conduction osseuse comme dans les études précédentes. Le mini clavier utilisé dans les études précédentes pour lancer les essais était également utilisé ici, mais il était fixé sur le tabouret en face duquel était assis le participant. Le participant appuyait sur le bouton pour démarrer chaque essai, ce qui permettait de fixer le point de départ de son doigt (voir figure 4.1) et donc de fixer la distance à parcourir jusqu'à la cible au début de l'essai.

Tableau 4.1 – Conditions de difficulté de la tâche de Fitts. A représente la distance à la cible (en centimètres), W le diamètre de la cible (en centimètres), et ID l'indice de difficulté (en bits). MT représente le temps (durée) de mouvement ("Movement Time") ou temps d'atteinte de cible (en secondes) obtenu pour chaque condition de difficulté.

A (cm)	W (cm)	ID (bits)	MT (s)	
			AP	DVH
12	5,33	2,17	12,06	9,03
18	5,33	2,76	7,89	6,78
27	5,33	3,34	5,70	5,27
12	8	1,58	13,28	10,15
18	8	2,17	8,91	7,47
27	8	2,75	6,72	6,04
12	12	1,00	17,88	11,50
18	12	1,58	10,66	9,24
27	12	2,17	9,07	7,30

Métriques spatiales

Plusieurs caractéristiques spatiales ont été extraites de la scène pour être transcodées en sons (Figure 4.2). Le vecteur OP définissait la direction du pointage du participant et la distance par rapport à la cible T.

Les métriques spatiales utilisées étaient pour la sonification AP (Figure 4.2(a)) :

- L'angle θ qui correspond à l'angle formé par les lignes OP et OT (écart angulaire).
- La distance d = [PT] projetée sur le plan horizontal parallèle au sol (écart de distance).

Et pour DVH (Figure 4.2(b)) :

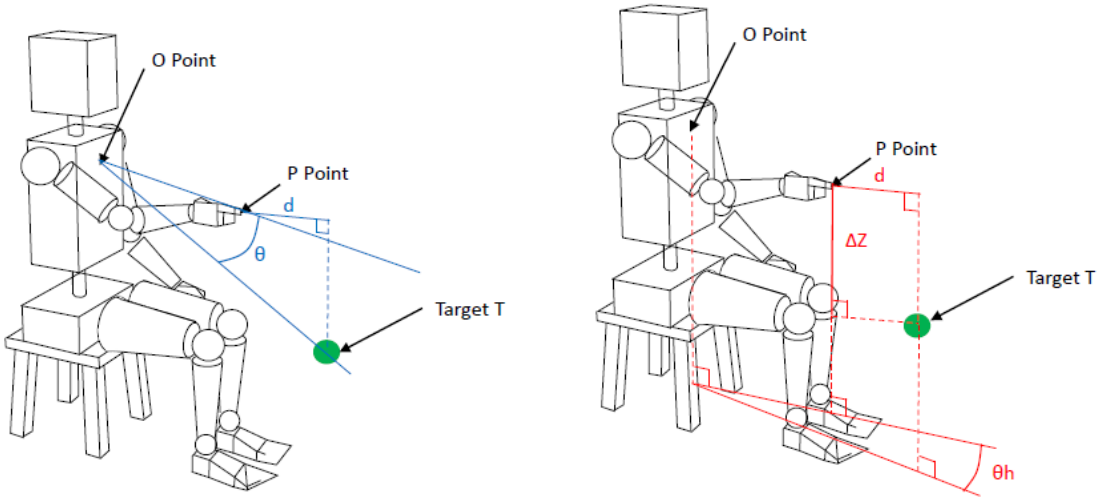
- L'angle θ_h qui correspond à la projection de l'angle formé par les lignes OP et OT sur le plan horizontal parallèle au sol (écart angulaire).
- La différence de hauteur ΔZ = |zP - zT|, qui correspond à la projection de la distance [PT] sur l'axe vertical Z (écart de distance).
- La distance d = [PT] projetée sur le plan horizontal parallèle au sol (écart de distance).

Sonifications

Un son pur était généré pour les deux modes de sonification.

Pour la sonification AP (voir tableau 4.2) :

- La fréquence f du son pur varie sur un continuum allant de *fmin* = 110 Hz pour θmax = 45° à *fmax* = 440 Hz pour θmin = 2°, de sorte que f se situe dans le spectre audible tout en évitant les fréquences désagréables.



(a) AP **(b)** DVH

Figure 4.2 – Métriques spatiales utilisées pour les sonifications AP et DVH testées sur une tâche de Fitts.
θ_h est la projection de l'angle formé par les droites OP et OT sur le plan horizontal parallèle au sol. d est la projection de la distance [PT] sur le plan horizontal parallèle au sol. ΔZ est la projection de la distance [PT] sur l'axe vertical Z.

Tableau 4.2 – Métaphores sonores utilisées pour la sonification AP sur la tâche de Fitts

	Horizontale	**Verticale**	**Distance**
Fréquence	θmin = 2°, θmax = 45° *fmax* = 440 Hz, *fmin* = 110 Hz		
Intensité			*dmin* = 0,01 m, *dmax* = 0,5 m *Imax* = 100 dB, *Imin* = 85 dB
Bruit Blanc	OP intersecte la cible 80 dB		P entre dans la cible 85 dB
Buzzer			P dépasse la cible

- La bonne direction est codée par l'activation d'un bruit blanc qui se superpose à ce son pur quand le participant pointe dans la direction de la cible.

Pour la sonification DVH (voir tableau 4.3) :

- La fréquence f du son pur variait sur un continuum allant de *fmin* = 110 Hz pour $\theta_h max$ = 45° à *fmax* = 440 Hz pour $\theta_h min$ = 2°.
- La dimension verticale est codée par l'activation d'un bruit blanc qui se superpose à ce son pur quand le participant pointe à la même hauteur que la position de la cible, c'est-à-dire lorsque ΔZ < r.

Pour les deux sonifications, la conversion de θ ou θ_h suit la loi psychophysique de Stevens (S. Stevens 1957) comme décrit dans les études précédentes.

Pour les deux sonifications, la distance à la cible est codée par l'intensité du son pur, qui varie sur un continuum avec une différence de 15 dB entre *Imin* (avec *dmax* = 0,5 m) et *Imax* (avec

Tableau 4.3 – Métaphores sonores utilisées pour la sonification DVH sur la tâche de Fitts

	Horizontale	Verticale	Distance
Fréquence	$\theta_h min$ = 2°, $\theta_h max$ = 45° *fmax* = 440 Hz, *fmin* = 110 Hz		
Intensité			*dmin* = 0,01 m, *dmax* = 0,5 m *Imax* = 100 dB, *Imin* = 85 dB
Bruit Blanc		$\|\Delta Z\|$ <r 80 dB	P entre dans la cible 85 dB
Buzzer			P dépasse la cible

dmin = 0,01 m). La conversion de la distance en intensité suit la fonction de conversion suivante :

$$I = Ai * log(d) + Bi \tag{4.1}$$

Avec :

$$Ai = \frac{Imin - Imax}{log(\frac{dmax}{dmin})} = -8,82 \tag{4.2}$$

$$Bi = Imin - Ai * log(dmax) = 82,34 \tag{4.3}$$

Des indices sonores supplémentaires permettant de préciser la trajectoire du doigt pendant les essais sont expliqués ci-dessous.

Protocole

L'expérience se déroulait dans l'espace de capture de mouvement décrit ci-dessus (section 4.2.2). Les participants, les yeux fermés, devaient trouver plusieurs cibles présentées l'une après l'autre.

L'expérience commençait par une phase d'entraînement, consistant en un bloc de 4 essais par taille de cible, en commençant par les plus grandes cibles et en terminant par les plus petites. Le participant pouvait ensuite faire autant de blocs d'essais qu'il jugeait nécessaire pour se familiariser avec la sonification et développer une stratégie de recherche qui lui convenait. Durant cette phase, le participant recevait des explications de l'expérimentateur sur les difficultés qui pouvaient survenir et sur la manière de les surmonter. La session se poursuivait par la réalisation de neuf blocs expérimentaux alternant la taille des cibles, dans un ordre contrebalancé entre les participants. Seuls les blocs expérimentaux étaient utilisés pour les statistiques. De l'arrivée du participant à son départ, la session durait une heure, incluant toutes les étapes mentionnées ci-dessus.

Pour l'expérience préliminaire, les participants réalisaient cette session avec la sonification

DVH. Pour l'étude de comparaison des sonifications AP et DVH, le déroulement de la session était identique, mais les participants réalisaient chacun deux sessions, l'une avec AP, et l'autre avec DVH, dans un ordre qui était contrebalancé entre les participants. Les deux sessions étaient espacées d'au moins 24 heures.

Chaque bloc se déroulait comme suit : le participant s'asseyait sur une chaise, un tabouret étant placé devant lui. Il fermait les yeux et commençait le premier essai en appuyant sur le bouton de démarrage du mini-clavier accessible par le toucher, placé sur le tabouret. Un son de début d'essai confirmait le début de l'essai. Comme décrit en détail précédemment, à partir de là, le participant entendait le retour sonore qui changeait en fonction de ses mouvements. Lorsque le doigt du participant passait trop loin derrière de la cible ([OP] > [OT]), une quinte parfaite était superposée au son pur. Le participant savait donc qu'il avait dépassé la cible, mais il conservait l'accès aux variations de hauteur et donc aux informations directionnelles. Lorsque P (le bout du doigt) entrait dans la cible, le son pur était interrompu et un fort bruit blanc était déclenché. Ce bruit blanc puissant était clairement distinct de celui indiquant la bonne hauteur pour la sonification DVH, ou la bonne direction pour la sonification AP, qui est plus faible. Lorsque le doigt P du participant restait à l'intérieur de la cible pendant 100 ms (s'assurant que la cible n'était pas atteinte par hasard), un son de victoire était joué ; le retour sonore était désactivé et l'essai s'achevait. L'essai était également interrompu si le participant n'atteignait pas la cible dans les 60 secondes, et un son de cloche indiquait la défaite. Après chaque essai, le participant commençait l'essai suivant en appuyant sur le bouton de démarrage avec son index P.

4.2.3 Résultats

Expérience préliminaire

Ici, cinq ID ont été calculés selon la formulation originale de la loi de Fitts (voir équation 2.2). Pour déterminer si une réponse de type Fitts se produit, nous effectuons une régression linéaire entre MT (temps d'atteinte de cible) et ID. La moyenne des temps d'atteinte de cible est calculée pour chaque combinaison de taille et de distance (9 combinaisons), comme Fitts l'a fait dans son étude (Fitts 1954).

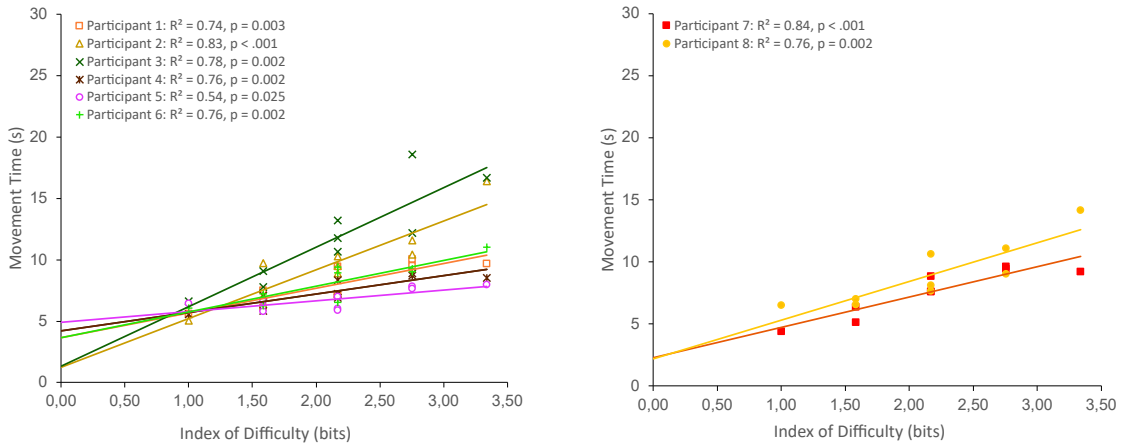
L'adéquation du modèle à nos données est confirmée, avec un R^2 de 0,93 pour les participants voyants ($F(1,7) = 96{,}07$, $p < .001$), et un R^2 de 0,93 pour les participants malvoyants ($F(1,7) = 89{,}74$, $p < .001$).

Les droites de régression des participants sont représentées sur la figure 4.3.

L'équation complète obtenue était, pour les participants normo-voyants :

$$MT = 3,17(\pm 0,59) + 2,55(\pm 0,26)ID \tag{4.4}$$

Et pour les participants déficients visuels :



(a) Normo-voyants **(b)** Malvoyants

Figure 4.3 – Droites de régression linéaire pour les temps de mouvement moyens (MT) en fonction de l'indice de difficulté (ID) pour chaque participant normo-voyants et malvoyant.

$$MT = 2,29(\pm 0,66) + 2,75(\pm 0,29)ID \tag{4.5}$$

Une analyse de puissance post hoc a été réalisée à l'aide de G*Power version 3.1.9.7 (Faul et al. 2007) pour l'estimation de la puissance sur la base de nos résultats pour les participants voyants. Une taille d'effet importante (13,29) a été calculée à partir du R^2 (.93) de l'analyse de régression. Avec un critère de significativité de α = .05, la puissance statistique obtenue est de .99. La taille de l'échantillon utilisé est donc adéquate pour tester l'hypothèse de l'étude.

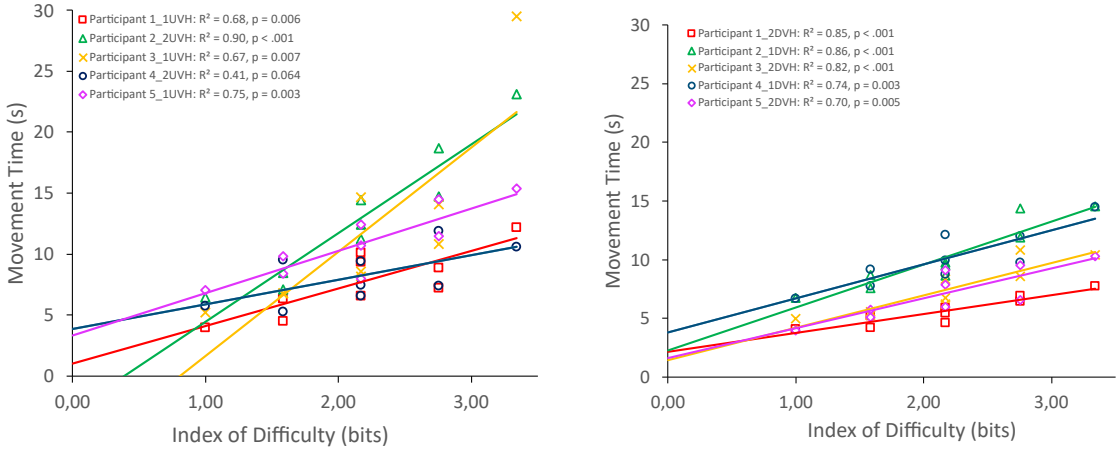
Enfin, nous avons calculé la moyenne du *throughput* (TP) avec l'équation 2.5 pour les participants voyants et malvoyants. Nous avons trouvé un TP moyen de 0,39 bits/s pour les participants voyants et de 0,36 bits/s pour les participants malvoyants.

Comparaison entre AP et DVH

Fitts Nous avons effectué une régression linéaire de MT en fonction de ID pour chaque condition de sonification, de la même manière que lors de l'expérience préliminaire. Les lignes de régression individuelles tracées pour les participants dans chaque condition de sonification sont représentées dans la figure 4.4. Les résultats montrent que dans les deux conditions, le MT augmente linéairement en fonction de l'ID. L'analyse de régression de la MT moyenne a donné des valeurs R^2 significatives de 0,83 pour AP ($F(1, 7) = 35,30$, $p < .001$), et R^2 = 0,91 pour DVH ($F(1, 7) = 73,27$, $p < .001$).

L'équation complète obtenue était, pour AP :

$$MT = -0,33(\pm 1,87) + 4,88(\pm 0,82)ID \tag{4.6}$$



(a) AP **(b)** DVH

Figure 4.4 – Droites de régression linéaire pour les temps de mouvement moyens (MT) en fonction de l'indice de difficulté (ID) pour chaque participant, pour les deux conditions de sonification, AP et DVH.

Et pour DVH :

$$MT = 2,24(\pm 0,72) + 2,70(\pm 0,32)ID \tag{4.7}$$

Le TP est donc de 1/4,88 = 0,20 bits/s pour AP, et de 1/2,70 = 0,37 bits/s pour DVH.

Temps d'atteinte de cible Les participants ont mis en moyenne 10,21 secondes (± 8,26 s) pour atteindre la cible avec la sonification AP et en moyenne 8,09 secondes (± 5,09 s) avec la sonification DHV. Pour comparer les MT pour les deux types de sonification, nous avons utilisé un modèle de Cox. Pour ce faire, nous avons utilisé la fonction coxph du package Survival du logiciel R. Les participants ont été significativement plus rapides à atteindre les cibles avec DVH (z = 6,14, p <.001) qu'avec AP.

4.2.4 Discussion

Notre étude visait à tester l'applicabilité de la loi de Fitts pour caractériser l'atteinte d'une cible sans la vue dans un environnement virtuel 3D avec un feedback auditif utilisant une sonification non spatialisée, et si applicable, l'utiliser pour comparer différents modes de sonification du guidage.

Nous avons effectué une expérience préliminaire sur une sonification déjà testée par avant, DVH, sur une tâche de type Fitts où le participant doit toucher du doigt plusieurs cibles successives, ici guidé par la sonification. Nous avons effectué des régressions linéaires entre les temps de mouvement moyens (MT) et l'indice de difficulté (ID). Comme prévu, les résultats montrent qu'il y a une forte corrélation entre MT et ID. Par conséquent, la loi de Fitts peut être utilisée

pour évaluer ce type de système de guidage sonore. Les résultats des participants malvoyants étaient similaires à ceux des participants normo-voyants.

Nous avons ensuite utilisé le *throughput* calculé grâce à la loi de Fitts, dans une étude comparative, pour comparer les deux modes de sonification que nous avions déjà comparé dans l'étude A1 sur la base du temps d'atteinte de cible. Nous nous attendions à retrouver de meilleures performances pour la sonification DVH, qui dissocie l'écart horizontal et l'écart vertical en deux paramètres sonores différents, comparé à la sonification AP, qui utilise un seul paramètre sonore pour coder un écart directionnel global.

Plusieurs points intéressants ressortent des résultats :

- Comme pour l'expérience préliminaire, les résultats ont montré que la loi de Fitts est un modèle valable pour évaluer l'atteinte de la cible, même avec une sonification non spatialisée (à la fois AP et DVH) comme seul feedback ;
- En utilisant la loi de Fitts, nous avons montré que les performances avec DVH dans l'expérience comparative sont similaires aux performances avec DVH lors de l'expérience préliminaire.
- En utilisant la loi de Fitts, nous avons montré que les performances d'atteinte de cible avec la sonification DVH sont meilleures que les performances avec la sonification AP.
- Cette différence de performance est confirmée par une comparaison des temps d'atteinte de cible en utilisant un modèle de Cox.

En effectuant une analyse de régression linéaire entre MT et ID, nous avons pu montrer que la loi de Fitts modélise correctement l'atteinte de cibles non visuelles en 3D avec un feedback sous forme de sonification non spatialisé. En effet, dans les deux conditions de sonification, les temps de mouvement augmentent significativement avec la difficulté de la tâche (ID), ce qui reproduit les résultats de l'expérience préliminaire. Nous trouvons des résultats similaires pour la sonification DVH dans cette expérience que pour les participants normo-voyants lors de l'expérience préliminaire, avec un R^2 de 0,91 dans l'expérience comparative et de 0,93 lors de l'expérience préliminaire, et une pente de 2,70 s/bit dans l'expérience comparative et de 2,55 s/bit lors de l'expérience préliminaire.

Nous observons que le TP pour la sonification DVH (0,37 bits/s) est presque deux fois plus élevé que le TP pour la sonification AP (0,20 bits/s). Cela signifie que les participants ont mis beaucoup plus de temps à atteindre les cibles avec AP qu'avec DVH au fur et à mesure que la difficulté de la tâche augmentait. La sonification DVH semble être plus efficace que AP en tant que retour d'information, car elle permet des temps de déplacement plus rapides pour des ID de plus en plus élevés. La supériorité de DVH sur AP est confirmée par la comparaison des MT avec un modèle de Cox, qui montre des mouvements significativement plus rapides avec DVH qu'avec AP.

La différence de performance entre AP et DVH est en accord avec nos études antérieures (Fons, Huet, Pellerin, Gerber et al. 2023) (voir section 3.2.3). AP intègre les dimensions spa-

tiales dans un flux auditif unique en utilisant des échelles continues. Cela permet un guidage sonore précis et rapide, mais le fait d'avoir un seul paramètre sonore qui code un écart 3D peut rendre compliquée l'interprétation des sons, une difficulté qui est amplifiée par le fait que les dimensions verticale et horizontale ne sont pas orthogonales. En revanche, DVH utilise deux flux sonores distincts (dans le même canal) pour coder les dimensions verticale et horizontale. Les dimensions sont donc orthogonales, ce qui permet au participant de les interpréter séparément. Malgré l'utilisation de deux flux sonores, le fait que l'un d'entre eux soit binaire évite aux participants de devoir passer de l'un à l'autre, ce qui réduit la charge cognitive (Gao et al. 2022). Par conséquent, les performances dans la tâche d'atteinte de la cible sont meilleures avec DVH qu'avec AP.

4.2.5 Conclusion

Les systèmes de guidage sonore sont traditionnellement évalués à l'aide du temps nécessaire pour atteindre des cibles dans des environnements réels ou virtuels en 3D. Dans ce chapitre, nous proposons d'aller au-delà de cette mesure classique en utilisant la loi de Fitts comme métrique standardisée, permettant de comparer les performances des dispositifs de manière objective et reproductible, indépendamment de la taille ou de la distance des cibles.

Dans une première expérience préliminaire, des participants normo-voyants et malvoyants ont atteint des cibles virtuelles en 3D avec leur index, guidés par la sonification qui dissocie les écarts horizontaux et verticaux en deux flux sonores séparés, la même que celle qui avait montré les meilleures performances dans l'étude A1 du chapitre précédent. Les résultats confirment que la loi de Fitts constitue un modèle valide pour évaluer l'atteinte de cibles, même lorsque le retour sensoriel est purement auditif avec une sonification non spatialisée, et sur une tâche 3D, et permettent de répliquer les résultats précédemment obtenus, validant l'efficacité de cette sonification dissociée.

Dans une deuxième expérience comparative, nous avons utilisé la loi de Fitts pour mesurer le *throughput* (TP) et comparer deux modes de guidage sonore déjà testés dans le chapitre précédent (étude A1) : la sonification dissociée (écarts horizontaux et verticaux séparés) et une sonification utilisant un seul paramètre sonore pour coder l'écart directionnel global. Les résultats confirment l'avantage de la sonification dissociée, qui permet une atteinte de cible plus rapide et efficace, tout en respectant les limites cognitives et perceptives des utilisateurs. Cette comparaison réplique et renforce les observations du chapitre précédent, montrant que la dissociation des dimensions verticales et horizontales est robuste et bénéfique pour la performance et l'intuitivité.

Cette étude démontre que la loi de Fitts peut être utilisée comme métrique standardisée pour évaluer et comparer différents systèmes de guidage sonore dans les mêmes conditions expérimentales. Contrairement au simple temps d'atteinte, influencé par les paramètres expérimentaux comme la taille et la distance des cibles, l'application de la loi de Fitts permet d'obtenir une mesure comparable entre études et dispositifs variés. Pour favoriser la reproductibilité et la comparai-

son dans la littérature, nous recommandons également que les auteurs publient les équations de régression complètes lorsqu'ils utilisent la loi de Fitts.

4.2.6 Résumé de l'étude B1

Les systèmes de guidage sonore sont traditionnellement évalués en mesurant le temps nécessaire pour atteindre des cibles, les appréciations subjectives des utilisateurs, et parfois la trajectoire de la main. Cependant, la diversité des paramètres expérimentaux (taille et position des cibles) rend difficile la comparaison des dispositifs entre études. Pour pallier cette limitation, nous proposons d'utiliser la loi de Fitts comme métrique standardisée, permettant de relier la difficulté de la tâche au temps de mouvement et de calculer une mesure d'efficacité comparable, le *throughput* (TP).

Dans cette première étude du chapitre, nous avons d'abord vérifié que la loi de Fitts pouvait être appliquée à une tâche d'atteinte de cible sans la vue guidée par une sonification non spatialisée, en environnement 3D. Les participants normo-voyants et malvoyants devaient atteindre des cibles virtuelles à portée de main, en utilisant un geste simple de la main, ce qui permet de rapprocher la tâche des conditions classiques de Fitts et de limiter l'atteinte à un mouvement du bras. Cette simplification rend l'application de la loi de Fitts pertinente, contrairement aux études précédentes où le geste combinait marche et tâtonnement avec la main.

Ensuite, nous avons comparé deux modes de sonification testés dans l'étude A1 du chapitre précédent :

— AP : conversion de l'écart directionnel global en fréquence sonore continue.
— DVH : dissociation des écarts horizontaux et verticaux en deux flux sonores distincts, la fréquence continue pour la dimension horizontale et un bruit blanc binaire indiquant la bonne hauteur pour la dimension verticale.

Les résultats confirment et répliquent ceux de l'étude A1 : la sonification DVH, qui dissocie les dimensions verticale et horizontale, permet une atteinte de cible plus rapide et efficace que la sonification AP, tant en termes de *throughput* calculé via la loi de Fitts que de temps d'atteinte de cible. Cette étude démontre que la dissociation des dimensions est un principe robuste et reproductible, et valide l'utilisation de la loi de Fitts comme métrique standardisée pour comparer les performances des dispositifs de guidage sonore, tout en facilitant la comparaison avec la littérature

Les résultats de cette étude ont été publiés dans (Fons, Huet, Pellerin et Graff 2024).

4.3 Étude B2 : Comparaison d'un mode de sonification séquentiel vs parallèle avec la loi de Fitts, sur une tâche d'atteinte de cible 3D avec la main simplifiée

4.3.1 Introduction

Nous avons montré dans l'étude A3 (section 3.4) qu'une sonification séquentielle (qui sépare les écarts sur chaque dimension en paramètres sonores différents et les présente l'une après l'autre) était plus efficace (plus rapide) et plus ergonomique (meilleurs scores d'ergonomie donnés par les participants) qu'une sonification parallèle (qui sépare les écarts sur chaque dimension en paramètres sonores différents et les présentes simultanément). Ces deux sonifications ont été testées sur une tâche d'atteinte de cible à la main, où les participants devaient se déplacer dans la salle avant de pouvoir "attraper" la cible avec la main.

Ici, nous voulons répliquer ces résultats sur une tâche d'atteinte de cible simplifiée à l'atteinte avec la main, en utilisant une métrique de la loi de Fitts, le *throughput* (TP). Cette étude a deux objectifs :

1. Renforcer la robustesse des résultats obtenus dans l'étude A3, en montrant que les différences observées entre les deux modes de sonification, séquentiel et parallèle, se retrouvent à la fois en comparant les temps d'atteinte de cible, et en comparant les TP mesurés grâce à la loi de Fitts.
2. Faciliter la comparaison avec d'autres études, puisque l'utilisation de la loi de Fitts permet de comparer nos résultats à ceux obtenus dans d'autres études, même dans des configurations expérimentales différentes.

Pour cela, les participants réalisent une tâche d'atteinte de cible de type Fitts, avec la cible toujours à portée de main, qu'ils doivent toucher du doigt le plus rapidement possible, comme dans l'étude précédente.

Pour rappel, le choix des métaphores sonores utilisées pour créer les deux modes de sonification séquentiel et parallèle s'était fait sur la base d'une étude ou nous avions testé différentes métaphores pour coder l'écart de distance sur l'axe vertical, et différentes métaphores pour coder l'écart angulaire horizontal (voir étude A2 du chapitre précédent, section 3.3). Ces métaphores étaient testées sur deux tâches d'atteinte de cible en une dimension : trouver une hauteur cible ou une direction cible.

Dans cette étude, nous avions fait varier la hauteur de la cible et son épaisseur pour la tâche verticale, et la direction de la cible. Nous avons donc pour chaque tâche plusieurs distances au point de départ (A), et plusieurs tailles de cible (W) (voir Figure 4.5), ce qui nous permet de

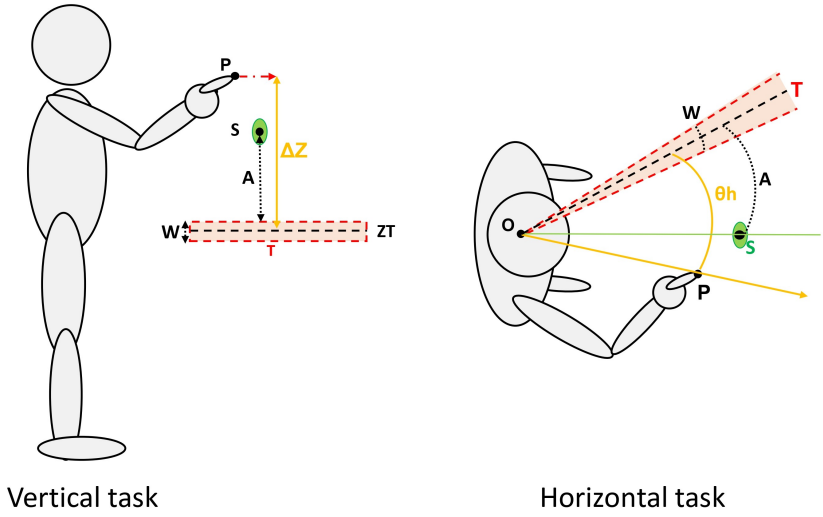


Figure 4.5 – Pour les deux tâches, S représente le bouton sur lequel les participants doivent appuyer pour commencer un essai. A représente l'écart entre la cible et S, et W la largeur de la cible. ΔZ est l'écart en distance converti en son pour la tâche verticale, et θ_h est l'écart angulaire converti en son pour la tâche horizontale.

calculer plusieurs indices de difficulté (ID) (selon le calcul de l'ID avec la loi originale de Fitts, équation 2.1, plus la cible est petite est loin, plus l'ID est élevé) et tester le lien entre ID et temps de mouvement (MT) pour chacune de ces tâches. Si la loi de Fitts s'applique ici, on devrait trouver une corrélation positive entre ID et temps d'atteinte de cible. Les tableaux 4.4 et 4.5 contiennent toutes les conditions de difficulté des deux tâches, les ID, et les temps d'atteinte de cible pour chacune des conditions de difficultés pour chacune des métaphores sonores.

Pour évaluer si la loi de Fitts s'appliquait dans ces conditions de métaphores sonores, nous avons effectué une analyse de régression linéaire pour chaque métaphore sur Jamovi, ainsi qu'un test F pour attester de la significativité de la relation linéaire entre MT et ID. Les MT sont moyennés pour chaque condition de taille et de distance à la cible. L'inverse de la pente a été utilisé pour calculer le TP pour chaque métaphore. Tous ces résultats sont représentés dans le tableau 4.6.

L'adéquation du modèle à nos données est confirmée, avec des corrélations significatives entre ID et MT pour chaque métaphore, pour les deux tâches. Dans l'étude A2, nous avions comparé les métaphores sur la base des temps d'atteinte de cible, des scores d'ergonomie donnés par les participants et des trajectoires du doigt. Nous avions gardé, pour l'étude A3 comparant la sonification séquentielle et la sonification parallèle, les voyelles pour coder la verticale, et la stéréo et la fréquence continue pour coder la direction horizontale, sur la base de ces résultats. Ici, en utilisant la loi de Fitts, la validité de nos choix est confirmée. En effet, pour la tâche verticale, nous voyons que la métaphore *Formant* a le TP le plus élevé, et pour la tâche horizontale, la métaphore *Continuous + Stereo* a le TP le plus élevé. La comparaison des métaphores avec la loi de Fitts est conforme aux résultats trouvés en comparant les temps d'atteinte de cible, les scores d'ergonomie et les trajectoires. L'utilisation de la loi de Fitts pourra permettre aux futurs auteurs utilisant la loi de Fitts pour évaluer leur dispositif de comparer leurs résultats aux nôtres.

Tableau 4.4 – Conditions de difficulté de la tâche de Fitts verticale et temps d'atteinte de cible. A représente la distance à la cible (en centimètres), W l'épaisseur de la cible (en centimètres), et ID l'indice de difficulté (en bits). MT représente le temps de mouvement ou temps d'atteinte de cible (en secondes) obtenu pour chaque condition de difficulté, pour chaque métaphore.

A (cm)	W (cm)	ID (bits)	MT (s)			
			Shepard	*Continuous*	*Formant*	*Discrete*
15	3,75	3	6,06	4,51	4,75	5,17
15	7,5	2	4,56	3,54	3,35	3,69
15	15	1	2,29	3,13	2,86	2,48
30	3,75	4	5,61	5,58	4,74	4,87
30	7,5	3	4,71	5,18	3,48	3,78
30	15	2	2,93	3,06	2,30	2,78
45	3,75	4,58	5,83	5,38	4,56	5,90
45	7,5	3,58	3,90	4,80	3,62	4,45
45	15	2,58	4,26	3,99	3,39	4,53

Tableau 4.5 – Conditions de difficulté de la tâche de Fitts horizontale et temps d'atteinte de cible. A représente la distance à la cible (en degrés), W la largeur de la cible (en degrés), et ID l'indice de difficulté (en bits). MT représente le temps de mouvement ou temps d'atteinte de cible (en secondes) obtenu pour chaque condition de difficulté, pour chaque métaphore.

A (°)	W (°)	ID (bits)	MT (s)		
			Continuous	*Stereo*	*Continuous + Stereo*
30	3,75	4,00	3,75	3,95	2,99
30	7,5	3,00	3,07	3,75	2,41
30	15	2,00	2,61	2,15	1,99
55	7,5	3,87	3,33	3,58	2,89
55	15	2,87	2,77	2,51	2,34
80	3,75	5,42	5,54	5,16	3,95
80	7,5	4,42	3,75	4,05	3,13
80	15	3,42	4,30	3,53	3,07

Tableau 4.6 – Equations de régression et résultats du test de régression entre MT et ID pour chaque métaphore pour les deux tâches verticale et horizontale. Le TP est présenté, en bits/s.

Métaphore	Equation de régression	R^2	F	p	TP (bits/s)
Tâche verticale					
Shepard	MT = 1,89(±0,84) + 0,90(±0,28) ID	0,60	10,54	0,014	1,11
Continuous	MT = 2,12(±0,40) + 0,79(±0,13) ID	0,84	35,67	<.001	1,27
Formant	MT = 2,00(±0,56) + 0,59(±0,18) ID	0,60	10,46	0,014	1,69
Discrete	MT = 1,53(±0,54) + 0.90(±0,18) ID	0,82	25,25	0,002	1,11
Tâche horizontale					
Continuous	MT = 0,81(±0,70) + 0,78(±0,19) ID	0,74	17,35	0,006	1,33
Stereo	MT = 0,58(±0,49) + 0,83(±0,13) ID	0,87	40,27	<.001	1,20
Continuous + Stereo	MT = 0,83(±0,24) + 0.56(±0,06) ID	0,93	77,28	<.001	1,79

Nous allons maintenant comparer la sonification séquentielle à la sonification parallèle, sur une tâche de Fitts similaire à celle de l'étude B1 (section 4.2). L'étude vise à répliquer nos résultats précédents tout en montrant que la loi de Fitts permet d'évaluer ces systèmes de manière robuste et comparable à d'autres travaux qui utiliseraient le TP pour évaluer le système, même s'ils utilisent une configuration de cibles différente. Les métaphores utilisées pour créer ces deux modes de sonification seront les mêmes que pour l'étude A3 (section 3.4), celles qui ont donné lieu aux meilleures performances lors de l'étude de comparaison de métaphores sur une dimension (section 3.3) à l'exception de la distance, qui ici sera codée par l'intensité et non par le tempo :

- Azimut : l'écart entre la direction pointée et la direction de la cible est converti en fréquence sur une échelle continue. Plus la direction pointée est proche de la direction de la cible, plus le son est aigu. Les sons sont présentés dans l'oreille droite lorsque la cible est à droite, et dans l'oreille gauche lorsque la cible est à gauche, et dans les deux oreilles lorsque le sujet pointe dans la direction de la cible.
- Profondeur : la distance entre le doigt et la cible est continuellement convertie en intensité sonore. À distance, l'intensité est faible, et plus le doigt se rapproche de la cible, plus l'intensité est forte.
- Hauteur : la voyelle [ɑ] est générée lorsque la cible est en dessous de la main du participant. La voyelle [o] est produite lorsque la cible est au-dessus de la main du participant.

Ces métriques spatiales sont présentées soient simultanément (mode parallèle), soit une à une, d'abord la direction (azimut) puis la hauteur quand le doigt pointe dans la bonne direction (mode séquentiel). Ici, seules deux dimensions sont présentées séquentiellement, la profondeur étant toujours donnée par l'intensité. Nous nous attendons à trouver un avantage de la sonification séquentielle à la fois au niveau des temps d'atteinte de cible, des scores d'ergonomie, et des TP.

4.3.2 Matériel et Méthodes

Participants

Dix participants (M = 29,7 ± 10,8 ans) ont été recrutés pour cette étude. Tous les participants ont donné leur consentement éclairé par écrit avant de participer à l'étude.

Matériel

L'expérience a été réalisée dans le même espace de capture de mouvement que celui décrit dans les études précédentes. La configuration, la taille et la distance de la cible sont identiques à celles de l'étude B1 (voir figure 4.1). Les participants étaient équipés d'un harnais et d'un pointeur sur l'index, comme dans l'étude précédente.

Tableau 4.7 – Métaphores sonores utilisées pour les deux modes de sonification séquentiel et parallèle sur la tâche de Fitts.

	Horizontale	Verticale	Distance
Fréquence	$\theta_h min$ = 2°, $\theta_h max$ = 45° *fmax* = 330 Hz, *fmin* = 110 Hz		
Stéréo	Cible à droite : oreille droite Cible à gauche : oreille gauche Bonne direction : deux oreilles		
Timbre		ΔZ <0 =>[o] ΔZ >0 =>[a]	
Bruit Blanc			P entre dans la cible 85 dB
Buzzer			P dépasse la cible

Sonification

Les métriques spatiales sont identiques à celles utilisées dans l'étude B1 pour la sonification DVH, soit l'angle azimutal θ_h, la différence de hauteur entre le doigt et la cible ΔZ, et la distance entre le doigt et la cible d (voir figure 4.2(b)). Les métaphores sonores utilisées pour les deux modes de sonification sont présentées dans le tableau 4.7.

- Sonification séquentielle : lorsque $|\theta_h|$ > 3°, un son pur est généré, dont la fréquence varie en fonction de θ_h sur un continuum allant de *fmin* = 110 Hz pour $|\theta_h|$ = 90°, à *fmax* = 330 Hz pour $|\theta_h|$ = 0 °. Lorsque $|\theta_h|$ < 3°, la sinusoïde s'arrête et une voyelle est générée, d'une hauteur constante de 330 Hz. Si ΔZ < 0, la voyelle [o] est générée. Si ΔZ > 0, la voyelle [ɑ] est générée.
- Sonification parallèle : une voyelle est générée, dont la fréquence varie en fonction de θ_h sur un continuum allant de *fmin* = 110 Hz pour $|\theta_h|$ = 90°, à *fmax* = 330 Hz pour $|\theta_h|$ = 0°. Si ΔZ < 0, la voyelle [o] est générée. Si ΔZ > 0, la voyelle [ɑ] est générée.

Pour les deux sonifications, un léger bruit blanc est généré et superposé au flux auditif principal lorsque $|\theta_h|$ < 3 ° et ΔZ < r. Un buzzer remplace tous les autres sons si le doigt (point P) dépasse la cible.

D'autres signaux sonores sont utilisés pour indiquer qu'un essai a commencé, qu'une cible a été atteinte et que le temps imparti est écoulé.

Protocole

L'expérience durait environ 45 minutes. Les participants effectuaient 3 blocs de 12 essais pour chaque sonification, un bloc pour chaque taille de cible. Au sein d'un bloc, il y avait 4 essais pour chaque distance par rapport au point de départ, dans un ordre aléatoire. Les participants effectuaient les trois blocs d'abord pour une sonification, puis pour l'autre. L'ordre dans lequel les sonifications étaient testées était contrebalancé entre les participants. Tous les participants

commençaient par le bloc avec les cibles les plus grandes, puis les cibles moyennes et enfin les cibles les plus petites. La phase de test était précédée d'une phase d'entraînement au cours de laquelle les participants devaient atteindre les cibles de la même manière que pendant la phase de test. L'entraînement durait aussi longtemps que le participant avait besoin pour se sentir à l'aise avec les sonifications.

Pour chaque bloc, le participant était assis sur une chaise, le bouton de démarrage étant placé sur un tabouret devant lui. Il appuyait sur le bouton pour commencer l'essai. Un feedback sonore était déclenché et continuait jusqu'à ce que le participant atteigne la cible ou que le délai de 60 secondes pour atteindre la cible se soit écoulé. Le participant appuyait à nouveau sur le bouton pour commencer l'essai suivant, jusqu'à la fin du bloc de cibles. À la fin des six blocs d'essais, le participant répondait à un questionnaire identique à celui des études précédentes.

Analyses

Les analyses ont été réalisées avec RStudio (RStudio 2023.12.1+402) et Jamovi (Jamovi 2.3.29).

Les valeurs aberrantes ont été supprimées : les temps d'atteinte de cible supérieurs à 1,5 fois l'intervalle interquartile ont été exclus des analyses.

Comme pour l'étude précédente, une analyse de Cox a été réalisée pour comparer les temps d'atteinte de cible entre les sonifications, ainsi qu'une analyse de régression entre ID et MT et une régression ordinale pour comparer les scores attribués à chaque sonification.

4.3.3 Résultats

Temps d'atteinte de cible

Le temps d'atteinte de cible moyen était de 8,09 ± 5,27 s pour la sonification parallèle et de 7,62 ± 4,98 s pour la sonification séquentielle. Une analyse de Cox n'a pas montré de différence significative entre les temps d'atteinte de cible moyens des deux sonifications (z = 1,19, p = 0,23).

Cinq IDs ont été calculés et une régression linéaire entre le temps d'atteinte cible (MT) et l'indice de difficulté (ID) a été réalisée pour chaque sonification. La moyenne du MT a été calculée pour chaque condition de taille et de distance par rapport à la cible, ce qui donne 9 conditions. Les conditions de difficulté et les temps d'atteinte cible pour chaque condition de difficulté pour chacune des sonifications sont représentés dans le tableau 4.8.

L'adéquation du modèle a été confirmée par des corrélations significatives entre ID et MT pour les deux modes de sonification. Le TP a été calculé pour chacune des deux sonifications. Ces résultats sont reportés dans le tableau 4.9.

Tableau 4.8 – Temps d'atteinte de cible moyen pour chaque condition de difficulté des sonifications parallèle et séquentielle.

A (cm)	W (cm)	ID (bits)	MT (s)	
			par	*seq*
12	5,33	2,17	9,13	8,78
12	8	1,58	7,30	7,73
12	12	1,00	6,07	6,20
18	5,33	2,75	8,28	7,38
18	8	2,17	6,99	6,51
18	12	1,58	6,68	5,58
27	5,33	3,34	10,42	10,20
27	8	2,75	8,51	8,05
27	12	2,17	9,37	8,20

Tableau 4.9 – Résultats de l'analyse de régression entre ID et MT pour les deux conditions de sonification séquentiel et parallèle.

Sonification	Équation de régression	R^2	F	p	TP (bits/s)
parallèle	$1,64(\pm 0,43) + 4,52(\pm 0,97)ID$	0,68	14,75	0,006	0,61
séquentiel	$1,45(\pm 0,51) + 4,49(\pm 1,16)ID$	0,54	8,06	0,025	0,69

Scores d'ergonomie

Nous avons effectué une régression ordinale pour comparer les scores entre les deux modes de sonification, ce qui a révélé un meilleur score pour la sonification séquentielle pour le critère d'efficacité (z = -2,25, p = 0,025). Il n'y a pas de différence significative entre les deux sonifications pour les critères de facilité et de confort.

4.3.4 Discussion

Nous avons comparé deux modes de sonification : l'un dissociant les dimensions verticales et horizontales, codées par deux paramètres sonores distincts et présentés l'un après l'autre (mode séquentiel), et l'autre dissociant les dimensions uniquement par des paramètres sonores différents, mais présentées simultanément (mode parallèle). Nous nous attendions à ce que la sonification séquentielle soit plus performante, conformément aux résultats obtenus dans une autre tâche d'atteinte de cible (étude A3, section 3.4). L'étude visait à la fois à répliquer ces résultats tout en montrant que la loi de Fitts permet d'évaluer ce type de dispositif et facilite la comparaison avec d'autres travaux de la littérature.

Les résultats ont confirmé une préférence pour le mode séquentiel, jugé plus efficace par la majorité des participants. Ces derniers ont rapporté que la sonification parallèle exigeait une « gymnastique cognitive » pour interpréter plusieurs sons en même temps, tandis que la sonification séquentielle leur permettait de se concentrer sur une dimension après l'autre, ce qui facilitait l'apprentissage.

Cependant, cette préférence subjective n'a pas conduit à une différence significative en termes de temps d'atteinte ou de *throughput* (TP), les deux modes produisant des performances très proches. Sur les 10 participants, 2 ont même indiqué préférer la sonification parallèle, perçue comme plus directe pour guider la trajectoire vers la cible, alors que la sonification séquentielle imposait un traitement dimension par dimension.

Cette absence de différence nette par rapport à l'étude précédente peut s'expliquer par la tâche utilisée ici, une tâche de type Fitts où la cible est atteinte directement avec la main, qui est beaucoup plus simple et rapide que la tâche de locomotion et de pointage utilisée auparavant. Cette simplicité a probablement conduit à un effet plafond, limitant la sensibilité de la tâche pour détecter les avantages du guidage séquentiel. L'avantage du mode séquentiel pourrait ainsi se manifester plus fortement dans des tâches plus complexes, où la gestion progressive des dimensions devient plus pertinente.

4.3.5 Conclusion

Cette étude a comparé deux modes de sonification pour le guidage de la main dans une tâche de type Fitts : un mode séquentiel, présentant les écarts vertical et horizontal successivement, et un mode parallèle, les présentant simultanément. Les résultats confirment que la majorité des participants perçoit la sonification séquentielle comme plus intuitive et moins exigeante cognitivement, car elle permet de traiter une dimension à la fois. En termes de performances objectives (temps d'atteinte et TP), aucun des deux modes ne s'est distingué, probablement en raison de la simplicité de la tâche et d'un effet plafond. Une minorité de participants a préféré la sonification parallèle pour sa trajectoire plus directe vers la cible. Cette étude permet de répliquer les résultats précédents et confirme que la loi de Fitts constitue une métrique pertinente pour évaluer et comparer les dispositifs de guidage sonore. Enfin, elle souligne que le choix du mode de sonification doit tenir compte de la complexité de la tâche et du niveau d'expérience de l'utilisateur.

4.3.6 Résumé de l'étude B2

L'objectif de cette étude était à la fois de confirmer nos découvertes sur la conception d'un dispositif de guidage sonore (étude A3, section 3.4), et de tester une nouvelle manière d'évaluer ce type de dispositif. Pour cela, nous avons comparé deux modes de sonification pour le guidage de la main vers une cible dans une tâche de type Fitts : un mode séquentiel, où les écarts vertical et horizontal sont codés par des paramètres sonores distincts et présentés successivement, et un mode parallèle, où ces mêmes écarts sont présentés simultanément. Les métaphores sonores utilisées pour chaque dimension avaient été sélectionnées sur la base de leur efficacité et de leur intuitivité dans l'étude A2 du chapitre précédent (section 3.3).

Les participants devaient atteindre des cibles virtuelles à portée de main en utilisant uniquement le geste de la main, ce qui permettait d'appliquer la loi de Fitts pour calculer le *throughput* (TP) et comparer les performances de manière standardisée. Cette simplification de la tâche rend la mesure comparable à d'autres études utilisant la loi de Fitts.

Les résultats montrent que la majorité des participants trouve la sonification séquentielle plus intuitive et moins exigeante cognitivement, car elle permet de se concentrer sur une dimension à la fois. La sonification parallèle, en revanche, demande d'interpréter plusieurs flux d'information simultanément, ce qui peut générer une surcharge cognitive pour certains participants. Malgré cette différence subjective, les performances objectives (temps d'atteinte et TP) sont très proches entre les deux modes, probablement en raison de la simplicité de la tâche (comparé à la tâche utilisée dans le chapitre précédent), qui induit un effet plafond. Une minorité de participants a préféré la sonification parallèle pour la trajectoire plus directe qu'elle permet si on arrive à traiter le surplus d'informations simultanées.

Cette étude permet de répliquer en partie les résultats de l'étude A3, tout en confirmant que la loi de Fitts constitue une métrique pertinente pour évaluer et comparer les dispositifs de guidage sonore. Elle souligne également que le choix du mode de sonification doit prendre en compte la complexité de la tâche et l'expérience de l'utilisateur : la sonification séquentielle peut favoriser les utilisateurs les moins expérimentés grâce à son nombre d'informations à traiter simultanément qui est réduit, tandis que la sonification parallèle pourrait être plus pertinente pour des utilisateurs expérimentés qui sauront profiter de toutes les informations disponibles simultanément.

Les résultats présentés ici ont été publiés dans (Fons, Huet, Pellerin et Graff 2025a).

Synthèse : l'évaluation des DSS de guidage avec la loi de Fitts

Cette thèse vise à répondre à la question suivante : comment concevoir et évaluer des métaphores sonores pour guider une personne vers une cible sans la vue ? Les deux études de ce chapitre portaient principalement sur l'évaluation du guidage sonore, mais aussi sur la conception des métaphores sonores. En effet, cette série expérimentale propose une métrique issue de la loi de Fitts pour évaluer les métaphores sonores, le *throughput* (TP), et l'utilise pour répliquer les résultats des comparaisons de sonifications du chapitre précédent. Nous répondons donc ici à deux sous-questions :

1. Quelles métaphores sonores sont les plus efficaces et intuitives pour encoder les trois dimensions spatiales, sans dépasser les capacités d'interprétation des utilisateurs ?
2. Comment évaluer un dispositif de guidage ?

Dans les deux études menées pour répondre à ces questions, nous avons donc pris en compte les limites récurrentes de la littérature sur les dispositifs sonores de guidage vers une cible :

— Le manque de validation empirique des sonifications (réplication de la série A)
— La surcharge cognitive engendrée par des systèmes trop complexes (réplication de la série A)
— L'absence de standardisation de l'évaluation des dispositifs, qui rend difficile la comparaison entre études (nouveauté de la série B)

Les deux études présentées dans ce chapitre utilisent la loi de Fitts comme métrique standardisée pour évaluer les systèmes de guidage sonore. Cette approche répond à une limite importante identifiée dans la littérature : la diversité des protocoles expérimentaux empêche de comparer les performances des DSS, car la taille, la position et la configuration des cibles varient d'une étude à l'autre. Les modes de sonification sont testés dans une tâche d'atteinte de cible avec la main, de type Fitts, ce qui permet à la fois de répliquer les résultats du chapitre précédent sur la conception des sonifications de guidage et de proposer une métrique standardisée facilitant la comparaison avec d'autres travaux.

Dans l'étude B1, la tâche d'atteinte de cible a été simplifiée : les participants devaient toucher directement les cibles avec la main, sans se déplacer, afin d'appliquer la loi de Fitts. Deux sonifications déjà testées dans le chapitre précédent ont été comparées : AP (Angle-to-Pitch), qui convertit l'écart directionnel global en fréquence continue, et DVH (Dissociated Vertical-Horizontal), qui dissocie les écarts horizontaux et verticaux en deux paramètres sonores distincts (fréquence continue pour l'horizontal, bruit blanc binaire pour le vertical).

Les résultats confirment la validité de la loi de Fitts pour ce type de tâche : l'indice de difficulté (ID) est corrélé au temps d'atteinte, et le *throughput* (TP) calculé avec la loi reflète

les différences observées dans les temps d'atteinte. On retrouve également les résultats du chapitre précédent : la sonification DVH permet d'atteindre les cibles plus rapidement et avec un TP plus élevé que la sonification AP. Les équations complètes de Fitts pour chaque condition sont fournies, ce qui permet de comparer ces performances à d'autres études, même avec des configurations de cibles différentes.

Dans l'étude B2, nous utilisons la même tâche de type Fitts. Deux modes de sonification déjà testés dans le chapitre précédent sont comparés : le mode séquentiel, où les écarts dimension par dimension sont présentés successivement, et le mode parallèle, où ces mêmes écarts sont présentés simultanément. Les métaphores sonores pour chaque dimension (fréquence continue et stéréo pour l'horizontale, voyelles binaires pour la verticale, tempo/intensité pour la distance) sont celles retenues dans l'étude A2 du chapitre précédent, et sont les mêmes pour les deux modes de sonifications.

Les résultats montrent que la majorité des participants juge le mode séquentiel plus facile et moins exigeant cognitivement, car il permet de traiter les dimensions une par une. Le mode parallèle exige de traiter plusieurs flux d'information simultanément, ce qui peut générer une surcharge cognitive. Cependant, aucune différence significative n'a été observée en termes de TP ou de temps d'atteinte, probablement en raison de la simplicité de la tâche et d'un effet plafond. Une minorité de participants a préféré la sonification parallèle pour la trajectoire plus directe qu'elle offre.

Ces deux études permettent de répliquer et confirmer les résultats des chapitres précédents, tout en introduisant une métrique standardisée pour comparer les systèmes de guidage sonore. Les résultats soulignent que les dispositifs les plus efficaces sont ceux qui réduisent la complexité cognitive grâce à des métaphores intuitives (voyelles, stéréo, correspondances cross-modales), tout en respectant les contraintes perceptives et cognitives des utilisateurs (séparation des dimensions spatialement et temporellement). En fournissant à la fois les TP calculés, les équations complètes de Fitts et les temps d'atteinte associés à chaque ID, ces travaux facilitent les comparaisons futures, y compris avec des configurations expérimentales différentes ou des variantes de la loi de Fitts.

En résumé, ce chapitre montre que la loi de Fitts est une métrique pertinente et standardisée pour évaluer les systèmes de guidage sonore, permettant de confirmer les différences de performances entre modes de sonification observées précédemment et de dépasser les limites méthodologiques identifiées dans l'état de l'art. Les applications visées par ces métaphores sonores restent pratiques : trouver un produit dans un supermarché, trouver le bon bouton d'interphone, etc.

Le chapitre suivant explore une application plus ludique : le guidage sonore de la visée du tir, dans le cadre d'un sport récent, le laser-run. L'adaptation du laser-run aux personnes handicapées visuelles est l'objectif final du projet Sam-Guide dans lequel s'inscrit cette thèse. Les sonifications testées dans ce nouveau chapitre sont celles validées précédemment dans

des tâches 3D, mais aussi celle utilisée en compétition de biathlon lors des épreuves de tir à la carabine. Ce chapitre permettra de vérifier si ces sonifications restent efficaces dans une tâche de tir en 2D, tout en fournissant des données nouvelles sur le guidage sonore pour le tir, un domaine peu exploré.

5. Série expérimentale C : Application et validation empirique des sonifications améliorées en 3D pour le guidage de la visée 2D d'une cible au pistolet

Sommaire

5.1 Introduction 116

5.2 Etude C1 : Comparaison de sonifications pour l'atteinte d'une cible 2D au tir au pistolet, sans contrainte de temps 119

5.2.1 Introduction 119

5.2.2 Matériel et Méthode 120

5.2.3 Résultats 124

5.2.4 Discussion 127

5.2.5 Conclusion 128

5.2.6 Résumé de l'étude C1 129

5.3 Etude C2 : Comparaison de sonifications pour l'atteinte d'une cible 2D au tir au pistolet, avec contrainte de temps 131

5.3.1 Introduction 131

5.3.2 Matériel et Méthode 132

5.3.3 Analyses 134

5.3.4 Résultats 137

5.3.5 Discussion 147

5.3.6 Conclusion 149

5.3.7 Résumé de l'étude C2 150

5.1 Introduction

Cette thèse vise à répondre à la question suivante : comment concevoir et évaluer des métaphores sonores pour guider une personne vers une cible sans la vue ? Après avoir mené deux séries d'expérience pour déterminer comment concevoir des métaphores sonores pour atteindre une cible avec la main, et comment évaluer les dispositifs de guidage sonore, nous nous penchons sur la troisième sous-question de cette thèse : Les métaphores optimisées pour l'atteinte de cible 3D avec la main sont-elles transposables à la visée au pistolet, où la précision et la rapidité sont critiques ? Les activités ludiques sont un terrain intéressant pour développer des procédures qui pourront ensuite servir dans d'autres contextes du quotidien.

L'accès à l'activité physique est souvent entravé par plusieurs facteurs pour les personnes malvoyantes. Cependant, la participation à des activités physiques a un impact significatif sur la qualité de vie et le bien-être physique (Diaz et al. 2019), ainsi que sur le bien-être psychologique et émotionnel (Bauman et al. 2012). Les sports offrent des opportunités de développement de compétences, de réseaux de soutien et d'indépendance (Darcy, D. Lock et Taylor 2017 ; Kiuppis 2018). Cet aspect social permet de lutter contre l'isolement que connaissent trop souvent les personnes handicapées.

Une revue (Elipe-Lorenzo et al. 2025) met en évidence plusieurs obstacles à la participation sportive des personnes handicapées : manque de personnel formé et de programmes accessibles, mauvaise compréhension de ce qu'est l'inclusion, approche trop compétitive du sport, attitudes discriminatoires et stéréotypes négatifs, manque d'accès à l'information, insuffisance d'installations et d'équipements accessibles, absence de politiques inclusives, obstacles financiers, etc.

Dans (Hillan et al. 2023), un groupe de discussion composé de personnes malvoyantes explore les obstacles et les facilitateurs de l'activité physique. Les participants critiquent l'approche "unique", qui suppose que tous les déficients visuels veulent pratiquer les mêmes activités. Les activités qu'ils recherchent sont souvent inaccessibles. En outre, dans les clubs sportifs, ils sont souvent les seuls participants malvoyants et estiment qu'ils prennent trop de temps et nécessitent trop d'aide de la part des instructeurs. Cette situation est à nouveau liée au manque d'installations et d'équipements accessibles.

Pour rendre plus accessibles les tâches difficiles ou impossibles à réaliser en raison d'une déficience visuelle, la substitution sensorielle est couramment utilisée. Certains sports adaptés utilisent des signaux auditifs, comme le hockey pour aveugles, qui utilise un palet sonore, le football pour aveugles, qui se joue avec un ballon contenant des clochettes, ou le paracyclisme, où les athlètes courent en tandem avec un guide voyant. En para-biathlon, le système de tir utilisé par les malvoyants est très réglementé et le seul système approuvé est le "système de tir EcoAims E00739 E-BSS2006 Biathlon", qui fournit un retour sonore. Comparé au para biathlon, où les spécifications du système de tir utilisé sont fixes, le laser-run est un sport nouveau, et les spécifications pour l'adapter aux malvoyants peuvent encore varier. Le laser-run consiste à courir et à tirer avec un pistolet laser sur une cible fixe.

Nous nous concentrons ici sur le tir à l'aveugle. Notre objectif est de développer le système de guidage de la visée basé sur la sonification le plus efficace et le plus ergonomique possible. Le système EcoAims utilisé en compétition repose sur la conversion de la distance entre le point de visée et le centre de la cible en tempo à longue distance, puis en fréquence à proximité du centre. Plus le point de visée est proche du centre de la cible, plus la fréquence sonore est élevée. Cependant, très peu d'études ont examiné les meilleurs paramètres de sonification pour ce type de tâche. Dans l'étude de Apavou, Bouchara et Bourdot 2024, les auteurs ont comparé quatre types de sonification, l'un étant la sonification de référence utilisée pour le tir à l'aveugle, c'est-à-dire la fréquence seule, à laquelle ils ont ajouté un son de référence. Deux conditions utilisent la fréquence pour la verticale (avec une condition continue et une condition binaire), et le tempo continu pour l'horizontale. La dernière condition s'inspire des travaux de Ziemer (Ziemer et Schultheis 2018), et utilise la chroma, les battements et la rugosité qui varient de manière continue pour coder les trois dimensions. Aucune différence de précision n'a été constatée entre les conditions. Les résultats montrent que les deux modes de sonification qui dissocient les axes verticaux et horizontaux permettent une stratégie de recherche axe par axe, ce qui n'est pas le cas pour la fréquence seule. La sonification inspirée de (Ziemer et Schultheis 2018), qui a fourni le plus d'informations (à la fois la direction et la distance à la cible pour les deux dimensions) a également nécessité le plus de temps pour tirer et a imposé une charge cognitive plus élevée.

Dans (Apavou, Bouchara et Bourdot 2024), les résultats montrent donc que la sonification qui fournit le plus d'informations est également la plus compliquée, que les participants mettent plus de temps à tirer et que la charge cognitive est plus élevée. Les résultats de nos études antérieures suggèrent que les participants ont tendance à utiliser une stratégie "étape par étape" dans leur recherche de cible (Fons, Huet, Pellerin, Gerber et al. 2023 ; Fons, Huet, Pellerin et Graff 2024) avec la main. Nous avons également montré qu'une sonification séquentielle était plus rapide qu'une sonification parallèle sur une tâche d'atteinte de cible 3D avec la main (Fons, Henry et al. 2025 ; Fons, Huet, Pellerin et Graff 2025a), et qu'elle était de plus jugée plus efficace par les participants.

Dans les chapitres précédents, nous avons montré que les sonifications dissociant les dimensions (en métaphores sonores différentes et présentées séquentiellement), testées pour guider la main vers une cible 3D, améliorent la performance tout en restant intuitives et économes cognitivement. Ce chapitre s'inscrit dans la continuité de ces travaux en appliquant ces métaphores à une tâche de visée 2D, dans le cadre du projet SAM-Guide, visant à permettre aux personnes privées de vue d'accéder au laser-run, qui nécessite à la fois rapidité et précision.

L'objectif de ce chapitre est donc double : premièrement, adapter et valider empiriquement les sonifications développées pour la 3D sur une tâche de visée 2D, afin de vérifier si les bénéfices observés précédemment se reproduisent ; deuxièmement, améliorer le guidage sonore utilisé en compétition, pour fournir aux athlètes malvoyants un retour sonore plus performant et confortable, qui les encouragera à pratiquer ces activités.

Deux études sont présentées dans ce chapitre :

— Etude C1 (section 5.2) : Comparaison de sonifications pour l'atteinte d'une cible 2D au tir au pistolet, sans contrainte de temps. Dans cette étude nous testons les sonifications dans une tâche de visée avec un pistolet factice et une cible virtuelle fixe, en comparant quatre modes de sonification dont les modes séquentiel et parallèle inspirés des chapitres précédents, comparés au mode de sonification de référence utilisé en biathlon au tir à la carabine. Le protocole est similaire à celui des expériences précédentes, avec un nombre d'essais (de tirs) défini et des mesures de temps mis pour tirer. Nous ajoutons ici des mesures de précision, un aspect essentiel du tir au pistolet.

— Etude C2 (section 5.3) : Comparaison de sonifications pour l'atteinte d'une cible 2D au tir au pistolet, avec contrainte de temps. L'étude C2 reprend les mêmes sonifications que dans l'étude C1, mais dans une tâche se rapprochant des conditions réelles du laser-run, avec une contrainte temporelle en plus d'une contrainte de précision, afin de vérifier si les résultats observés sur les tâches précédentes se maintiennent dans un contexte plus proche de la discipline sportive que nous voulons rendre accessible.

5.2 Etude C1 : Comparaison de sonifications pour l'atteinte d'une cible 2D au tir au pistolet, sans contrainte de temps

5.2.1 Introduction

Notre étude compare le système existant avec d'autres sonifications que nous avons développées, sur la base de la littérature sur le guidage auditif vers une cible, et des résultats de nos études précédentes (Fons, Henry et al. 2025 ; Fons, Huet, Pellerin, Gerber et al. 2023 ; Fons, Huet, Pellerin et Graff 2024, 2025a) (voir les deux chapitres précédents, sections 3 et 4). Nous avons créé et comparé quatre types de sonification pour guider la visée du tir au pistolet :

- PO (Pitch Only) : L'écart au centre de la cible est continuellement converti en fréquence sonore. C'est la sonification de référence utilisée dans les systèmes de tir à l'aveugle, que nous souhaitons améliorer (condition contrôle).
- PWN (Pitch and White Noise). L'une des difficultés du Pitch Only est de reconnaître la fréquence la plus élevée indiquant le centre de la cible. Dans cette sonification, lorsque le centre de la cible est pointé, un bruit blanc est ajouté de façon binaire à la fréquence. PWN est similaire à la sonification AP utilisée dans l'étude A1 sur une tâche 3D (section 3.2).
- PSTseq (Pitch, Stereo and Timbre en mode séquentiel) : les déviations selon les dimensions verticale et horizontale sont dissociées (fréquence continue et stéréo binaire pour l'horizontale, voyelles binaires pour la verticale) et délivrées de manière séquentielle, de la même manière que le mode séquentiel de l'étude A3 (section 3.4).
- PSTpar (Pitch, Stereo and Timbre en mode parallèle) : identique à PSTseq, mais les deux dimensions verticale et horizontale sont présentées simultanément, de la même manière que le mode parallèle de l'étude A3 (section 3.4).

Ces deux dernières sonifications sont plus complexes que les deux premières. L'utilisation d'encodages binaires (stéréo droite/gauche, timbre haut/bas) vise à simplifier le signal et à alléger la charge cognitive. De plus, avec PO et PWN, lorsque l'essai commence, nous ne savons pas si le centre de la cible est plus vers le haut ou vers le bas, vers la droite ou vers la gauche. La personne doit initier un mouvement et exploiter les variations de sons pour situer le centre de la cible. Les signaux binaires suppriment cette difficulté en fournissant directement la direction de la cible. L'utilisation de la stéréo binaire nous rapproche de l'intuitivité d'une sonification spatialisée, en se libérant des limites du système auditif à localiser une source sonore. L'utilisation de voyelles est plus proche des instructions verbales, simples et naturelles, mais en plus rapides à interpréter que des mots entiers.

Ces quatre sonifications ont été testées dans le cadre d'une tâche de visée et de tir au pistolet sur une cible fixe. La tâche se déroule dans un espace de capture de mouvement qui permet de

mesurer les écarts entre la position du point de visée et la position de la cible et de les convertir en sons. Nous comparons la précision de tir (distance globale à la cible, distance sur l'axe vertical et distance sur l'axe horizontal), le temps mis pour tirer et l'ergonomie de ces quatre sonifications.

Nous nous attendons à ce que les trois dernières sonifications (PWN, PSTseq, PSTpar) offrent une meilleure précision de tir que la sonification PO, ainsi qu'une meilleure ergonomie. Nous nous attendons également à ce que les métaphores dissociant l'horizontal et le vertical (PSTseq et PSTpar) soient plus efficaces que celles donnant une déviation globale (PO et PWN) et plus ergonomiques. Enfin, nous pensons qu'une sonification dissociant les dimensions temporellement (PSTseq) sera plus ergonomique qu'une sonification identique dans laquelle les dimensions sont présentées simultanément (PSTpar).

5.2.2 Matériel et Méthode

Participants

Chaque participant n'a effectué qu'une seule des deux comparaisons de sonification (pour des raisons de fatigue due à la durée de l'expérience) : PO vs PWN ou PSTseq vs PSTpar.

12 étudiants en psychologie normo-voyants ont participé à la comparaison PO vs PWN (M = 19,5 ans, SD = 0,8 ans). Un étudiant déficient visuel a également participé à cette comparaison (18 ans).

12 étudiants en psychologie normo-voyants ont participé à la comparaison PSTseq vs PSTpar (M = 20 ans, SD = 1,6 ans).

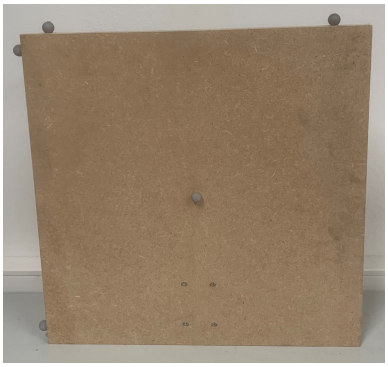
Tous les participants ont donné leur consentement éclairé par écrit avant de participer à l'étude.

Matériel

L'expérience s'est déroulée dans un espace de capture de mouvement. Le participant était équipé d'un pistolet factice imprimé en 3D (voir Fig. 5.1(a)), inspiré d'un modèle de pistolet de laser run, sur lequel étaient fixés des réflecteurs le repérant dans l'espace virtuel de capture de mouvement. Une cible carrée (voir figure 5.1(b)) était également localisée par le système. Un paramétrage à partir du système de capture de mouvement nous permettait de bouger à volonté la cible virtuelle (un cercle de 6 cm de diamètre) ancrée dans l'espace de capture de mouvement par cet objet cible, et ainsi de dépasser les limites de l'espace réel de la salle. Nous avons ainsi pu reproduire les dimensions standard du pas de tir employé en compétition de laser-run. Le dispositif de capture de mouvement et de conversion des écarts en son via Pure Data était similaire à celui des autres expériences (Figure 1.1). La gâchette du pistolet factice était connectée au PC pilote, ce qui permettait de détecter les tirs des participants.



(a) Pistolet factice imprimé en 3D.
Un câble relie la gâchette à l'ordinateur pilote et détecte le tir. Des tiges supplémentaires soutiennent les réflecteurs sphériques captés par le système de capture de mouvement.



(b) Objet cible physique.
Cette cible visible était utilisée pendant la phase de familiarisation. Le réflecteur central marque le centre de la cible. Elle était placée à un mètre du pas de tir. Pendant les phases d'entraînement et de test, les participants visaient et tiraient sur une cible virtuelle invisible, liée à la position de la cible visible mais décalée à 5 mètres du pas de tir et à 1,40 mètre au-dessus du sol en modifiant les paramètres du système de capture de mouvement.

Figure 5.1 – Objets équipés de réflecteurs sphériques et localisés par le système de capture de mouvement.

Sonification

Les métriques spatiales extraites de la scène étaient déduites de la position du point de visée de l'arme sur le plan vertical contenant la cible. Trois métriques ont ainsi été calculées :

- ΔD : distance globale entre le point de visée et le centre de la cible ;
- ΔDx : distance horizontale entre le point de visée et le centre de la cible.
- ΔDz : distance verticale entre le point de visée et le centre de la cible.

Ces mesures ont été converties en sons dans quatre conditions de sonification :

- PO (Pitch Only, voir tableau 5.1) : la fréquence sonore son pur varie entre 110 Hz (pour ΔDmax = 2 m) et 1000 Hz (pour ΔDmin = 0,01 m). La conversion de ΔD en fréquence suit la loi de puissance de Stevens (S. Stevens 1957).
- PWN (Pitch + White Noise, voir tableau 5.2) : identique à PO, mais avec un bruit blanc superposé au son pur quand $\Delta D <= 3$ cm (quand le participant vise bien le centre de la cible).
- PSTseq (Pitch, Stereo and Timbre en mode séquentiel, voir tableau 5.3) : quand $\Delta Dx > 3$ cm, la fréquence du son pur varie entre 110 Hz et 300 Hz. Le son pur est transmis dans l'oreille droite quand $\Delta Dx < -3$ cm, et dans l'oreille gauche quand $\Delta Dx > 3$ cm. Quand $\Delta Dx <= 3$ cm, le son pur est remplacé par la voyelle [a] quand $\Delta Dz > 0$ cm, et par la voyelle [o] quand $\Delta Dz < 0$ cm. La voyelle est transmise dans les deux oreilles. Un bruit blanc est superposé à la voyelle quand $\Delta D <= 3$ cm.

Tableau 5.1 – Métaphores sonores utilisées pour la sonification PO

	Horizontale	**Verticale**
Fréquence	$\Delta Dmax$ = 2 m, $\Delta Dmin$ = 0,01 m *fmax* = 1000 Hz, *fmin* = 110 Hz	

Tableau 5.2 – Métaphores sonores utilisées pour la sonification PWN

	Horizontale	**Verticale**
Fréquence	$\Delta Dmax$ = 2 m, $\Delta Dmin$ = 0,01 m *fmax* = 1000 Hz, *fmin* = 110 Hz	
Bruit Blanc	ΔD <3 cm	

- PSTpar (Pitch, Stereo and Timbre en mode parallèle, voir tableau 5.3) : une voyelle est toujours générée, en continu : [a] quand $\Delta Dz > 0$ cm, et [o] quand $\Delta Dz < 0$ cm. La fréquence de cette voyelle varie entre 110 Hz et 330 Hz. Elle est transmise dans l'oreille droite quand $\Delta Dx < -3$ cm, et dans l'oreille gauche quand $\Delta Dx > 3$ cm.

Protocole

L'expérience durait environ 45 minutes et se déroulait dans l'espace de capture de mouvement décrit précédemment. Le protocole était similaire à celui des expériences précédentes, avec deux blocs de cinq essais pour chaque condition de guidage. Les participants visaient et tiraient plusieurs fois sur une cible virtuelle fixe de 6 cm de diamètre, centrée sur une cible physique enregistrée dans le système de capture de mouvement (voir Fig. 5.1(b)). La taille de la cible virtuelle était similaire à celle des cibles utilisées en laser-run, et la distance à la cible fixée à 5 m du pas de tir et à 1,4 m du sol, les conditions utilisées dans certaines conditions de laser-run (selon l'âge des joueurs). Contrairement aux expériences en 3D présentées dans les séries précédentes, la cible était donc fixe, et un bloc contenait cinq essais (cinq tirs), ce qui correspond au nombre de tirs que les joueurs de laser-run doivent effectuer avant de pouvoir continuer à courir.

Les participants avaient pour consigne d'être aussi précis que possible lorsqu'ils tiraient sur la cible. Un temps limite de trois minutes était fixé pour chaque tir, afin d'éviter que certains

Tableau 5.3 – Métaphores sonores utilisées pour les deux modes de sonification séquentiel et parallèle de la tâche de tir

	Horizontale	**Verticale**
Fréquence	$\Delta Dxmax$ = 2 m, $\Delta Dxmin$ = 0,01 m *fmax* = 330 Hz, *fmin* = 110 Hz	
Stéréo	Cible à droite : oreille droite Cible à gauche : oreille gauche Bonne direction : deux oreilles	
Timbre		ΔDx <0 =>[o] ΔDx >0 =>[a]
Bruit Blanc	ΔD <3 cm	



Figure 5.2 – Tâche de tir au pistolet.
Le participant vise la cible avec le pistolet factice. La cible physique utilisée pour la phase de familiarisation est située à un mètre de la table derrière laquelle le participant est positionné pour tirer. La cible virtuelle que le participant vise pendant le test (visible sur l'écran de contrôle à gauche) est située à 5 mètres du stand de tir et à 1,4 mètre au-dessus du sol. Sur l'ordinateur pilote (à gauche), l'expérimentateur visualise la visée en temps réel et les impacts des tirs virtuels.

essais ne s'éternisent quand le participant perdait complètement la cible. En pratique cette limite a été très rarement atteinte. Chaque participant confrontait deux conditions de sonification : PO et PWN pour les uns, et PSTseq et PSTpar pour les autres. Positionnés derrière une table, ils tiraient sur une cible placée sur une table devant le pas de tir (voir Fig. 5.2).

Dans la phase de familiarisation, les participants visaient et tiraient en gardant les yeux ouverts pour entendre tous les sons des deux conditions de sonification et comprendre comment les mouvements de leur arme influençaient le retour sonore. Ils pouvaient s'approcher librement de la cible, en plaçant l'extrémité du pistolet en son centre pour entendre les sons correspondants. Le participant déficient visuel était assisté dans cette phase, où l'expérimentateur bougeait le pistolet que le participant tenait en main, et lui indiquait où il se situait par rapport au centre de la cible et les modifications de son que sa position et ses mouvements entraînaient.

Une phase d'entraînement et une phase de test se suivaient dans les mêmes conditions. Les participants pouvaient avoir un feedback sur leurs performances pendant l'entraînement. Ils s'entraînaient autant que nécessaire avant de commencer les blocs de test. Ils effectuaient quatre blocs de test (deux par condition), chacun contenant cinq essais (c'est-à-dire cinq tirs), en alternant les conditions dans un ordre contrebalancé. Un entraînement supplémentaire était autorisé entre les blocs de test si nécessaire.

Chaque bloc suivait la séquence suivante : le participant fermait les yeux et une petite musique indiquait le début du bloc. Pour "charger" le pistolet, il le levait verticalement, ce qui déclenchait

un son imitant le son de recharge d'une arme conventionnelle, marquant le début de la sonification, et donc de l'essai de tir. Cette disposition répond aux exigences des épreuves de tir qui ne permettent pas de conserver sa position après un tir réussi ou pour réajuster le précédent. Après avoir tiré (ou échoué à tirer dans le délai imparti), il rechargeait l'arme pour commencer l'essai suivant, jusqu'à ce que le bloc se termine par un autre signal sonore indiquant la fin. L'expérimentateur pouvait alors afficher la performance du participant sur l'ordinateur pilote, en montrant les points de visée sur la cible.

À la fin de la session, le participant remplissait un questionnaire. Il évaluait chacun des deux mode de sonification sur une échelle de Likert de 1 à 7 selon trois critères : l'efficacité, le confort et la facilité.

Analyses

Les comparaisons de la précision globale (ΔD), de la précision horizontale (ΔDx) et de la précision verticale (ΔDz) ont été effectuées avec des tests de Wilcoxon pour les comparaisons entre PO et PWN, et entre PSTseq et PSTpar (échantillons appariés essai par essai). Les comparaisons de précision entre PO et PSTseq, PO et PSTpar, PWN et PSTseq, PWN et PSTpar, ont été effectuées avec des tests de Mann-Whitney (échantillons indépendants). Les analyses de précision verticale et horizontale ont été effectuées sur les valeurs absolues des écarts. Toutes les comparaisons de précision ont été effectuées sur Jamovi 2.3.26.0.

En ce qui concerne la rapidité du tir, le temps nécessaire pour tirer a été comparé entre les conditions, à l'aide de tests de Wilcoxon ou de Mann-Whitney en fonction de l'indépendance des échantillons.

Afin de comparer les scores d'ergonomie entre les différentes conditions, des tests de régression ordinale ont été effectués pour chaque paire de conditions. Des comparaisons ont été effectuées entre les conditions PO, PWN, PSTseq et PSTpar. Les p-values ont été corrigées avec la méthode Bonferroni pour limiter le risque de faux positifs lié aux comparaisons multiples.

5.2.3 Résultats

Précision

La précision globale (ΔD), horizontale (ΔDx) et verticale (ΔDz) du tir sont représentées graphiquement sur la figure 5.3.

Les résultats ont montré une différence significative dans la précision globale (ΔD) entre PO et PWN (W = 2450, p = 0,014), les sujets étant en moyenne plus précis avec PWN (M = 55 mm, SD = 39 mm) qu'avec PO (M = 71 mm, SD = 47 mm). Les résultats pour le participant malvoyant étaient similaires, avec une distance moyenne au centre de la cible de 27 mm pour la condition PWN contre 67 mm pour la condition PO.

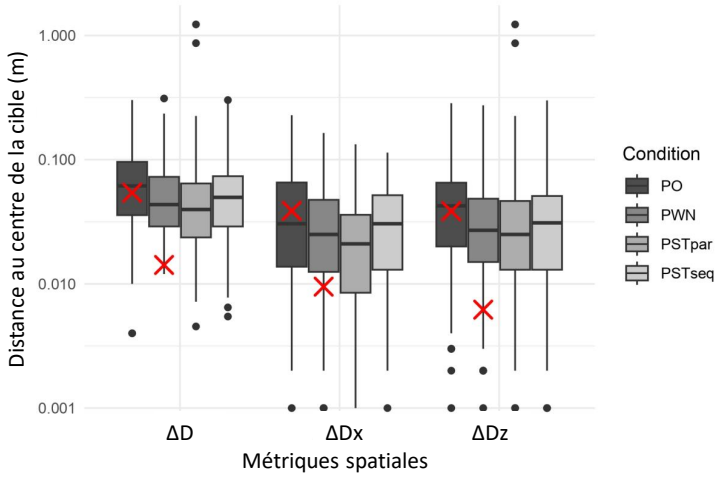


Figure 5.3 – Distance du point de visée par rapport au centre de la cible au moment du tir (échelle log10).
Distance globale (ΔD), distance horizontale (ΔDx), distance verticale (ΔDz), en mètres, pour chacune des quatre conditions de sonification. Chaque boîte représente la médiane (ligne horizontale), l'intervalle interquartile (boîte) et les valeurs extrêmes aberrantes (moustaches). Les croix rouges représentent la précision médiane du participant malvoyant, qui a testé les modes PO et PWN.

La comparaison entre PSTseq et PSTpar a montré une différence significative de précision, uniquement pour la dimension horizontale ΔDx (W = 2226, p = 0,002), les sujets étant plus précis dans la condition PSTpar (M = 24 mm, SD = 20 mm) que dans la condition PSTseq (M = 32 mm, SD = 23 mm).

Temps d'atteinte de cible

Le temps nécessaire pour tirer sur la cible dans chaque condition de sonification est représenté sur la Fig. 5.4.

Les résultats montrent que les participants ont tiré plus rapidement dans la condition PSTseq (M = 19,88 s, SD = 11,64 s) que dans la condition PSTpar (M = 23,97 s, SD = 13,28 s) (W = 4092, p = 0,007).

Ils étaient également plus rapides dans la condition PSTseq que dans la condition PO (U = 5155, p = 0,002).

Scores d'ergonomie

La proportion de chaque score sur l'échelle de Likert donnée par les participants pour chaque condition de sonification est représentée dans la Fig. 5.5.

Les résultats de la régression ordinale montrent que les participants ont attribué des scores d'ergonomie moins élevés à PO qu'à PWN (z = 4,638, p = 0,002), PSTseq (z = 1,771, p = 0,001) et PSTpar (z = 3,593, p = 0,001). Ils ont également donné de meilleurs scores à la condition PWN

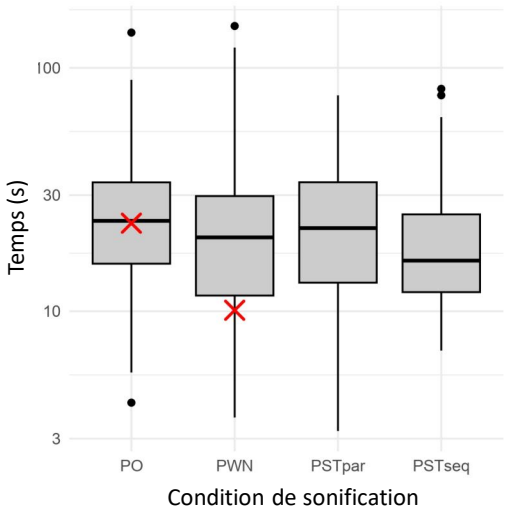
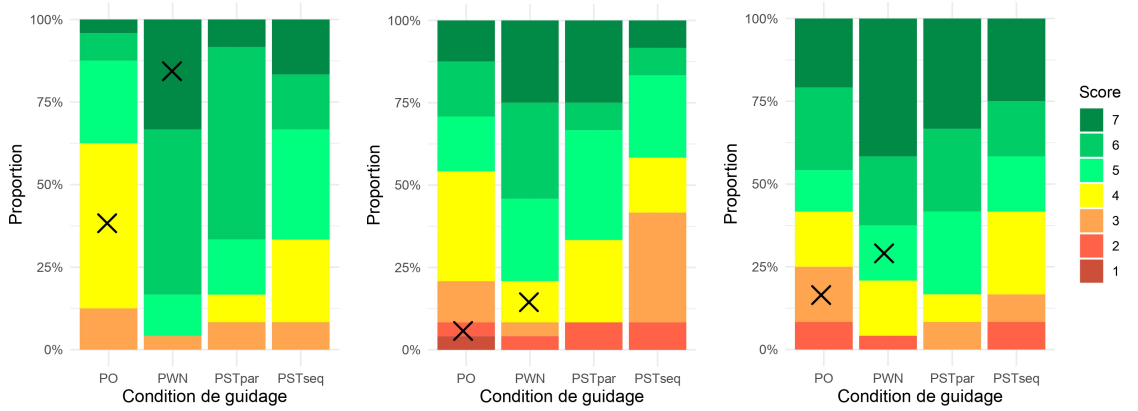


Figure 5.4 – Temps mis pour tirer, en secondes (échelle log10), pour chacune des quatre conditions de sonification.
Chaque boîte représente la médiane (ligne horizontale), l'intervalle interquartile (boîte) et les valeurs extrêmes aberrantes (moustaches). Les croix rouges représentent le temps médian de tir du participant malvoyant, qui a testé les modes PO et PWN.



(a) Efficacité **(b)** Confort **(c)** Facilité

Figure 5.5 – Pourcentage de participants ayant donné chaque score pour les quatre conditions de la tâche de tir.
Les trois critères sont pris en compte (efficacité, confort, facilité). Le gradient de couleur représente les niveaux de notation, le rouge correspondant à 1 et le vert foncé à 7, de sorte que les zones plus vertes correspondent à des jugements plus positifs. Les croix noires indiquent le score attribué par le participant malvoyant, pour chaque critère des deux conditions de sonification qu'il a testées, PO et PWN.

qu'à la condition PSTpar (z = 1,809, SD = 0,047).

5.2.4 Discussion

Comme attendu, les résultats montrent une plus grande précision avec PWN, PSTseq et PSTpar qu'avec PO. Les participants ont également trouvé la sonification PO moins ergonomique que les autres sonifications. En termes de temps mis pour tirer, seule PSTseq était plus rapide que PO. Cela confirme notre première hypothèse.

Nous avons observé que le participant malvoyant, qui a testé les sonifications PO et PWN, a obtenu de meilleurs résultats avec PWN qu'avec PO, comme les autres participants, mais avec une différence plus prononcée entre les deux conditions. Cela peut être dû à sa forte motivation pour l'expérience et à sa volonté de passer de nombreux blocs d'entraînement avant de passer les blocs de test. Pour rappel, la phase d'entraînement permettait aux participants de s'exercer autant qu'ils le jugeaient nécessaire avant de passer au test, afin d'atteindre un plafond de performances.

Nous nous attendions également à ce que les sonifications qui séparent les dimensions verticales et horizontales en deux paramètres sonores (PSTpar et PSTseq) soient plus performantes que celles qui utilisent un seul paramètre pour la distance globale au centre (PO et PWN), et également plus ergonomiques. PSTpar et PSTseq ont été jugés plus ergonomiques que PO, mais PSTpar était considéré comme moins ergonomique que PWN. En termes de précision, PSTpar et PSTseq ont été plus performants que PO, mais pas plus performants que PWN. La condition PSTseq était plus rapide que la condition PO, mais il n'y avait pas de différence de vitesse entre PSTseq et PWN, ou entre PSTpar et les autres conditions. Notre hypothèse n'a été que partiellement confirmée.

Enfin, nous nous attendions à ce que PSTseq, qui sépare temporellement les informations verticales et horizontales, soit jugé plus ergonomique que PSTpar, qui présente les deux simultanément. Cette hypothèse n'a pas été confirmée, car il n'y a pas eu de différence dans les évaluations des participants.

Dans l'ensemble, PO, la sonification conventionnelle utilisée dans les systèmes de tir à l'aveugle, a été jugée moins ergonomique et s'est révélée moins précise que les autres solutions. L'ajout d'un simple signal binaire pour le centre de la cible a considérablement amélioré la précision et l'ergonomie. Cependant, les sonifications séparant les dimensions verticales et horizontales n'offraient pas d'avantage majeur par rapport à PWN, qui utilise la fréquence continue et un bruit blanc de bonne direction. La sonification PWN a également été jugée plus ergonomique que PSTpar. Pour une sonification simple et précise, l'ajout d'un bruit blanc pour indiquer le centre de la cible peut être suffisant, tandis que la séparation des axes pourrait être inutile et ajouter une charge cognitive.

Cependant, PSTseq, qui séquentialise la présentation des écarts verticaux et horizontaux, a permis de tirer plus rapidement. Dans le laser-run, la vitesse est aussi importante que la précision. Les joueurs alternent course et tir, et peuvent repartir du pas de cours soit après avoir réussi

cinq tirs, soit au bout de 50 secondes, le nombre de tirs étant illimité. À précision égale, une sonification qui sépare les dimensions spatiales et temporelles (PSTseq) pourrait être préférable à une métrique spatiale unique convertie en fréquence (PWN) parce qu'elle permet de tirer plus vite pour une précision similaire. Nous pensons que cela est dû au fait que la séparation des axes permet aux participants de se concentrer sur un seul axe à la fois, une stratégie observée dans des études antérieures (Apavou, Bouchara et Bourdot 2024 ; Fons, Huet, Pellerin, Gerber et al. 2023 ; Fons, Huet, Pellerin et Graff 2024). Alors que PSTpar devrait également permettre cela, la présentation simultanée des deux dimensions pourrait avoir surchargé cognitivement les participants, comme le suggèrent ses notes ergonomiques inférieures à celles de PWN.

PSTseq et PSTpar fournissent tous deux un retour d'information directionnel instantané via la stéréo, ce que les participants ont apprécié. Cependant, le codage de l'élévation de la cible pourrait être amélioré, car certains participants ont eu du mal à distinguer les voyelles synthétiques, ce qui leur a demandé un effort cognitif supplémentaire. Nous pourrons tester d'autres timbres plus discriminants et intuitifs. Il est également possible d'adapter la taille de la zone de bruit blanc indiquant le centre de la cible aux capacités du tireur.

Le perfectionnement de ces aspects pourrait améliorer la précision et la vitesse de tir de PSTseq, ce qui en ferait un bon compromis entre les deux - un facteur essentiel dans le laser-run. De futures études pourraient approfondir cet équilibre, en se concentrant séparément sur la vitesse et la précision, et en incorporant une tâche cardiovasculaire pour mieux simuler les conditions réelles de la compétition. Enfin, tester ces sonifications avec des tireurs plus expérimentés permettrait d'obtenir des informations supplémentaires.

Au-delà de l'accessibilité, ces sonifications pourraient également profiter aux tireurs voyants en améliorant la perception sensorielle de la visée, tant dans les jeux vidéo que dans les sports de tir laser. Les indications auditives pourraient compléter les repères visuels, améliorant ainsi la précision et durée d'un tir. En outre, le tir sonore pourrait être utilisé à la fois par les malvoyants et les voyants, ce qui favoriserait une meilleure intégration dans les sports. Contrairement à de nombreuses activités adaptées qui séparent les athlètes aveugles et voyants, cette approche leur permet de s'entraîner et de concourir ensemble, réduisant ainsi les obstacles à la participation.

5.2.5 Conclusion

Nos résultats montrent que les sonifications alternatives améliorent la précision du tir et l'ergonomie par rapport à la sonification classique utilisée dans le tir à l'aveugle. L'ajout d'un simple signal binaire pour indiquer le centre de la cible (PWN) apporte déjà des gains significatifs en termes de précision et de facilité d'utilisation. Les sonifications qui séparent les axes verticaux et horizontaux n'ont pas toujours offert un avantage clair, bien que PSTseq ait permis un tir plus rapide, ce qui est bénéfique dans des contextes sportifs comme le laser-run.

Ces résultats ouvrent la voie à des améliorations dans la conception du guidage auditif pour les sports de tir et les jeux. Des recherches futures pourraient explorer l'impact de ces sonifica-

tions dans des conditions sportives plus réalistes, intégrant des contraintes de temps ou d'effort physique.

5.2.6 Résumé de l'étude C1

L'accès des personnes malvoyantes aux activités sportives est souvent limité par le manque d'équipement adapté et de méthodes d'assistance efficaces. Cette étude explore la sonification pour améliorer le guidage de la visée dans le tir à l'aveugle, qui est très peu explorée dans la littérature. Nous avons comparé quatre sonifications :

1. Pitch Only (PO) : méthode traditionnelle utilisée en biathlon, où la fréquence varie continuellement avec la direction visée.
2. Pitch + White Noise (PWN) : ajout d'un bruit blanc indiquant le centre de la cible.
3. Pitch + Stereo + Timbre en mode séquentiel (PSTseq) : séparation de la direction horizontale (fréquence + stéréo) et verticale (timbre), présentées successivement.
4. Pitch + Stereo + Timbre en mode parallèle (PSTpar) : même principe que PSTseq, mais avec une présentation simultanée des dimensions.

Ces nouvelles métaphores s'inspirent directement de nos travaux antérieurs sur l'atteinte de cibles 3D. PWN reprend le principe d'AP (Angle-to-Pitch) testé dans les études 1 (section 3.2) et 4 (section 4.2). PSTseq et PSTpar adaptent les modes séquentiel et parallèle étudiés dans l'étude A3 (section 3.4), appliqués ici à une tâche de visée 2D.

Nous avons formulé trois hypothèses :

- les trois nouvelles sonifications seraient plus performantes que la condition contrôle PO,
- la séparation des dimensions (PSTseq, PSTpar) améliorerait encore les performances par rapport à PO et PWN (comme dans l'étude A1 sur une tâche 3D, section 3.2, et l'étude B1, section 4.2).
- le mode séquentiel (PSTseq) permettrait de meilleures performances que le mode parallèle (PSTpar) (comme dans l'étude A3, section 3.4 et l'étude B2, section 4.3).

Le but de cette étude était donc à la fois d'améliorer le système de guidage sonore classique utilisé en compétition et vérifier si les résultats obtenus en 3D peuvent être répliqués sur une tâche de visée 2D. Pour cela, nous avons mesuré la précision du tir et le temps de réaction sur une tâche de visée dans un espace de capture de mouvement, ainsi que les préférences des participants.

Les résultats montrent que toutes les nouvelles sonifications améliorent la précision et la facilité d'utilisation par rapport à la méthode traditionnelle. L'ajout de bruit blanc (PWN) augmente considérablement la précision et la facilité d'utilisation, offrant une amélioration simple mais

très efficace. L'amélioration des performances par la séparation des dimensions (PSTseq, PST-par), comparée à PWN, est moins prononcée, bien que PSTseq permette la visée la plus rapide, ce qui est utile dans les sports où le temps est compté, comme le laser-run.

Ces résultats ont été publiés dans (Fons, Huet, Pellerin et Graff 2025b).

5.3 Etude C2 : Comparaison de sonifications pour l'atteinte d'une cible 2D au tir au pistolet, avec contrainte de temps

5.3.1 Introduction

Dans les chapitres précédents, nous avons optimisé des sonifications pour le guidage de la main vers une cible dans un espace 3D. Dans ce chapitre, nous menons des études pour savoir si les métaphores sonores optimisées pour l'atteinte de cible 3D avec la main sont transposables à la visée au pistolet. Dans l'étude C1, nous avons comparé quatre modes de sonification : trois modes inspirés des sonifications optimisées en 3D dans les chapitres précédents, et un mode inspiré du système actuellement utilisé par les athlètes déficients visuels lors des épreuves de tir à la carabine au biathlon. Ces modes de sonification ont été testé sur une tâche de tir au pistolet, avec un protocole similaire à celui des études menées en 3D, c'est-à-dire avec un nombre d'essais (ici, de tir) fixe, où on mesurait un temps d'atteinte de cible moyen et une précision moyenne pour chaque condition de guidage.

Nous avons montré que toutes les sonifications optimisées amélioraient la précision et la facilité d'utilisation par rapport à la méthode actuellement utilisée au para-biathlon. Cependant, la présentation successive des dimensions spatiales (mode séquentiel) n'apportait pas un avantage franc par rapport à la présentation simultanée des dimensions (mode parallèle). En effet, le mode séquentiel permettait de tirer plus rapidement, mais le mode parallèle permettait de tirer plus précisément. De plus, l'avantage de ces deux modes de sonification qui séparaient vertical et horizontal en métaphores sonores distinctes ne ressortait pas, pas rapport à une sonification fournissant un écart global à la cible ajouté à un bruit blanc de bonne direction.

Dans cette nouvelle étude, nous voulons comparer les mêmes modes de sonification, mais en modifiant le protocole pour qu'il se rapproche des conditions réelles du tir lors d'une épreuve de laser-run. Nous voulons améliorer la sonification permettant aux personnes déficientes visuelles de participer à ce sport, dont l'accessibilité est l'objectif final du projet SAM-Guide dans lequel s'inscrivent les travaux de cette thèse. Pour cela, nous avons introduit à la fois une limite de temps et une contrainte de précision. Dans l'épreuve de laser-run, durant la phase de tir, les joueurs doivent toucher une cible fixe cinq fois avant de pouvoir reprendre la course. Ils ont pour cela un nombre d'essais illimité, mais un temps limite de cinquante secondes après leur premier tir, au delà duquel ils peuvent courir à nouveau.

Nous avons donc comparé quatre sonifications, par paires, sur une tâche de tir avec un temps limite de cinquante secondes après le premier tir. Comme nos participants, normo-voyants et déficients visuels, ne sont pas des experts du tir et sont donc peu nombreux à réussir à toucher la cible 5 fois en moins de cinquante secondes, nous avons également comparé les sonifications

sur une tâche en temps illimité, où les participants avaient un nombre de tir illimité pour réussir à toucher la cible 5 fois.

Les quatre types de sonification pour guider la visée du tir au pistolet étaient les mêmes que pour l'étude C1. Lorsque nous avons comparé un mode de sonification séquentiel à un mode parallèle sur une tâche d'atteinte avec la main (étude A3, section 3.4), nous avons discuté du fait que techniquement, le mode parallèle devrait permettre d'atteindre la cible par une trajectoire directe et donc permettre de meilleures performances, mais que la grande quantité d'information à traiter en simultané était difficile à interpréter par les participants. Si le mode séquentiel convient bien à des débutants, des utilisateurs plus expérimentés pourraient tirer parti des avantages du mode parallèle. Ainsi, nous avons également testé l'effet de l'entraînement, uniquement sur cette paire de sonifications, PSTseq et PSTpar. Quatre groupes de participants ont effectué quatre comparaisons : PO vs PWN ; PWN vs PSTseq ; PWN vs PSTpar ; PSTseq vs PSTpar. Pour tester l'effet de l'entraînement, les participants testant PSTseq et PSTpar réalisaient deux sessions expérimentales séparées d'au moins 24 heures. Dans l'étude C1 toutes nos nouvelles sonifications optimisées ont montré un avantage par rapport à la sonification PO. Le but de ces comparaisons est de tester les hypothèses suivantes :

— L'ajout d'un bruit blanc indiquant le centre de la cible (PWN) permettra de meilleures performances et une meilleure ergonomie que la condition contrôle (PO).
— La dissociation des dimensions en deux métaphores sonores distinctes (PSTseq et PSTpar) permettra de meilleures performances et une meilleure ergonomie qu'une sonification ne dissociant pas les dimensions (PWN).
— Une sonification dissociant les dimensions temporellement (PSTseq) permettra de meilleures performances et une meilleure ergonomie qu'une sonification identique dans laquelle les dimensions sont présentées simultanément (PSTpar), pour des utilisateurs débutants.
— L'avantage de la sonification PSTseq sur la sonification PSTpar diminuera chez les utilisateurs expérimentés.

5.3.2 Matériel et Méthode

Participants

En tout, 49 personnes ont participé à l'expérience. Sur ces 49 participants, 11 étaient déficients visuels (4 femmes et 7 hommes, M = 58 ans ; ET = 13 ans) et 38 étaient normo-voyants (32 femmes et 6 hommes, M = 25 ans ; ET = 10 ans).

Ces participants étaient répartis dans quatre conditions expérimentales :

— *Comparaison entre PO et PWN* : N = 12 (9 normo-voyants, 3 déficients visuels)
— *Comparaison entre PWN et PSTpar* : N = 13 (9 normo-voyants, 4 déficients visuels)
— *Comparaison entre PWN et PSTseq* : N = 16 (10 normo-voyants, 6 déficients visuels). Deux

des participants déficients visuels n'ont pas pu finir l'expérience, l'une en raison de la chaleur excessive dans la salle où se déroulait l'expérience, l'autre pour des raisons de temps disponible. Les résultats des phases qu'ils ont pu effectuer ont été conservés, ainsi que leurs scores d'ergonomie.

— *Comparaison entre PSTseq et PSTpar* : N = 14 (10 normo-voyants, 4 déficients visuels). Dans cette condition, les participants effectuaient deux fois l'expérience, en deux sessions espacées de 24h à 72h. Une participante normo-voyante a également passé une seule session de PSTseq-PSTpar en raison de problèmes matériels lors de la deuxième session. Deux des quatre participants déficients visuels ont passé les deux sessions, les deux autres ont réalisé seulement une session.

Un des participants déficients visuels a passé les tests pour les quatre modes de sonification, en deux sessions séparées de 24h. Une autre participante déficiente visuelle a passé les quatre modes de sonification en une seule session, mais n'a pas passé tous les tests pour la sonification PO.

Tous les participants ont donné leur consentement éclairé par écrit avant de participer à l'étude.

Matériel

Le matériel était identique à celui de l'étude C1 : un pistolet factice (voir Figure 5.1(a)) et une cible (voir Figure 5.1(b)) captés par le système de capture de mouvement. La distance à la cible était la même (5 mètres) et la hauteur de la cible également (1,4 mètres).

Sonification

Les métriques spatiales utilisées étaient les mêmes que pour l'étude C1 : la distance globale (ΔD), la distance horizontale (ΔDx) et la distance verticale (ΔDz).

Les quatre modes de sonification étaient les mêmes que pour l'étude C1 (voir tableaux 5.1, 5.2 et 5.3). Une seule modification a été apportée pour les sonifications PSTseq et PSTpar. En effet, dans l'étude précédente, nous avons remarqué que certains participants avaient du mal à distinguer les deux voyelles indiquant la hauteur de la cible. Ainsi, nous avons conservé le [ɑ] pour dire d'aller vers le bas, mais nous avons remplacé le [o] par un [i] afin que le contraste entre les deux timbres soit plus net.

Protocole

L'expérience durait entre 45 minutes et 1h30 selon la rapidité des participants. Chaque participant testait deux modes de sonification l'un après l'autre, dans un ordre contrebalancé entre les participants. Tout au long de l'expérience, le participant devait "charger" le pistolet (le lever

à la verticale) pour déclencher le feedback sonore avant de viser et tirer. Pour chaque mode de sonification, l'expérience se déroulait dans cet ordre :

- Familiarisation : Identique à la phase de familiarisation de l'étude C1. Les participants exploraient librement les métaphores sonores avec les yeux ouverts. Les participants déficients visuels étaient assistés dans cette phase, où l'expérimentateur bougeait le pistolet et expliquait les changements de sons qui survenaient.
- Première phase en temps illimité : Le participant fermait les yeux et devait atteindre la cible cinq fois. Il disposait d'un nombre d'essais illimité pour y arriver. Cette phase s'arrêtait quand la cible était atteinte cinq fois.
- Deux phases de 50 secondes : Le participant devait essayer de toucher la cible cinq fois, en moins de 50 secondes. Le chronomètre des 50 secondes se déclenchait au moment du premier tir (comme au laser-run). Cette phase s'arrêtait soit quand le participant touchait la cible pour la cinquième fois, soit quand le temps était écoulé. Cette phase était répétée deux fois.
- Deuxième phase en temps illimité : identique à la première phase en temps illimité.

Comme dans l'étude C1, une petite musique indiquait le début de la phase de test, et une autre musique en indiquait la fin. A la fin d'une phase de test, l'expérimentateur pouvait montrer au participant normo-voyant sa performance affichée sur l'ordinateur pilote, en montrant les points de visée sur la cible, ou décrire au participant déficient visuel sa performance.

A la fin de la session, le participant remplissait le questionnaire pour évaluer les deux modes de sonification testés sur une échelle de Likert de 1 à 7 selon trois critères : efficacité, confort et facilité.

Les participants effectuant la comparaison entre PSTseq et PSTpar réalisaient la même expérience deux fois : une première session à J0, et une deuxième session entre J+1 et J+3.

5.3.3 Analyses

Nous avons effectué plusieurs mesures quantitatives :

— Temps mis pour réussir cinq tirs, pour la phase de tir en temps illimité.
— Score vitesse-précision (dont nous décrivons le calcul plus bas)
— Nombre de tirs réussis, pour la phase de tir en 50 secondes.
— Temps d'atteinte de cible (temps mis pour tirer, sur les tirs réussis)
— Précision

Et une mesure qualitative : les scores d'ergonomie donnés par les participants.

Ces mesures sont comparées, pour chaque condition (PO - PWN ; PWN - PSTseq ; PWN - PSTpar ; PSTseq - PSTpar) et chaque groupe de participant (normo-voyants ; déficients visuels) selon plusieurs facteurs :

— Sonification (PO ; PWN ; PSTseq ; PSTpar) : deux modalités, selon la condition expérimentale.
— Bloc (bloc 1 ; bloc 2) : chaque phase de test (temps illimité ou temps limité à 50 s) était effectuée deux fois par chaque participant (bloc 1 ; bloc 2). L'effet du bloc est donc un effet d'entraînement.
— Session (session 1 ; session 2) : uniquement pour la comparaison entre PSTseq et PSTpar, les participants réalisaient l'expérience deux fois, en deux sessions espacées d'au moins 24h. L'effet de la session est donc également un effet de l'entraînement.

Temps mis pour atteindre les cinq tirs réussis

Pour chaque comparaison (PO - PWN ; PWN - PSTseq ; PWN - PSTpar ; PSTseq - PSTpar), et chaque groupe de participant (normo-voyants ; déficients visuels), nous avons réalisé une ANOVA à mesures répétées comparant le temps mis pour réussir à toucher la cible cinq fois lors de la phase en tir illimité, en fonction de la sonification (2 modalités) et du bloc (bloc 1 ; bloc 2). Pour la comparaison entre PSTseq et PSTpar, la session (session 1 ; session 2) était ajoutée comme facteur pour tester l'effet de l'entraînement.

Compromis vitesse-précision

Afin de mesurer l'efficacité des sonifications à la fois sur la vitesse et la précision, deux critères aussi important l'un que l'autre au laser-run, nous avons créé un score composite comprenant à la fois le temps mis pour tirer et la distance à la cible.

Avant cela, nous avons enlevé les données aberrantes des temps mis pour tirer et des distances à la cible, à savoir les données supérieures à la moyenne plus deux fois l'écart-type, ce qui a mené à exclure environ 5% des données.

Pour chaque tir, nous avons calculé un score z du temps mis pour tirer et de la distance à la cible (z = (x - moyenne)/écart-type). La moyenne et l'écart-type utilisés sont ceux du temps mis pour tirer et de la distance de tout l'échantillon de participants, pour chaque comparaison de sonification. Le score composite mesurant à la fois la vitesse et la précision était ensuite calculé en ajoutant les scores z du temps et de la distance, et en prenant l'opposé de la somme (-z), de sorte qu'un score élevé corresponde à un temps court et une distance courte, et donc à une vitesse et une précision élevées. Les scores correspondent donc au nombre d'écart-type au-dessus ou en dessous de la moyenne, et peuvent donc être positifs ou négatifs.

Pour chaque comparaison (PO - PWN ; PWN - PSTseq ; PWN - PSTpar ; PSTseq - PSTpar), chaque phase de test (temps illimité ; temps limité à 50 s) et chaque groupe de participant (normo-voyants ; déficients visuels), nous avons réalisé une ANOVA à mesures répétées comparant les scores en fonction de la sonification (2 modalités) et du bloc (bloc 1 ; bloc 2). Pour la comparaison entre PSTseq et PSTpar, la session (session 1 ; session 2) était ajoutée comme facteur pour tester

l'effet de l'entraînement.

Nombre de tirs réussis en 50 secondes

Nous avons calculé le nombre de tirs réussis durant la phase de tir en temps limité à 50 secondes.

Pour chaque comparaison (PO - PWN ; PWN - PSTseq ; PWN - PSTpar ; PSTseq - PSTpar) et chaque groupe de participant (normo-voyants ; déficients visuels), nous avons réalisé une ANOVA à mesures répétées comparant le nombre de tirs réussis en fonction de la sonification (2 modalités) et du bloc (bloc 1 ; bloc 2). Pour la comparaison entre PSTseq et PSTpar, la session (session 1 ; session 2) était ajoutée comme facteur pour tester l'effet de l'entraînement.

Temps d'atteinte de cible

Nous avons comparé les temps d'atteinte de cible entre sonifications, c'est-à-dire les temps mis pour tirer, pour les essais réussis.

Pour chaque comparaison (PO - PWN ; PWN - PSTseq ; PWN - PSTpar ; PSTseq - PSTpar) et chaque groupe de participant (normo-voyants ; déficients visuels), nous avons réalisé une ANOVA à mesures répétées comparant les temps mis pour tirer en fonction de la sonification (2 modalités). Pour la comparaison entre PSTseq et PSTpar, la session (session 1 ; session 2) était ajoutée comme facteur pour tester l'effet de l'entraînement.

Précision

Nous avons comparé la précision entre les sonifications, c'est-à-dire la distance entre le point de visée et le centre de la cible à chaque tir.

Pour chaque comparaison (PO - PWN ; PWN - PSTseq ; PWN - PSTpar ; PSTseq - PSTpar) et chaque groupe de participant (normo-voyants ; déficients visuels), nous avons réalisé une ANOVA à mesures répétées comparant la précision en fonction de la sonification (2 modalités). Pour la comparaison entre PSTseq et PSTpar, la session (session 1 ; session 2) était ajoutée comme facteur pour tester l'effet de l'entraînement.

Scores d'ergonomie

Des tests de régression ordinale ont été effectués pour chaque paire de conditions (PO - PWN ; PWN - PSTseq ; PWN - PSTpar ; PSTseq - PSTpar) et pour chaque groupe de participants (normo-voyants ; déficients visuels).

Pour la comparaison de PSTseq et PSTpar, la session (session 1 ; session 2) a été ajouté comme facteur, pour tester l'effet de l'entraînement.

5.3.4 Résultats

Temps mis pour atteindre les cinq tirs réussis (phase de test en temps illimité)

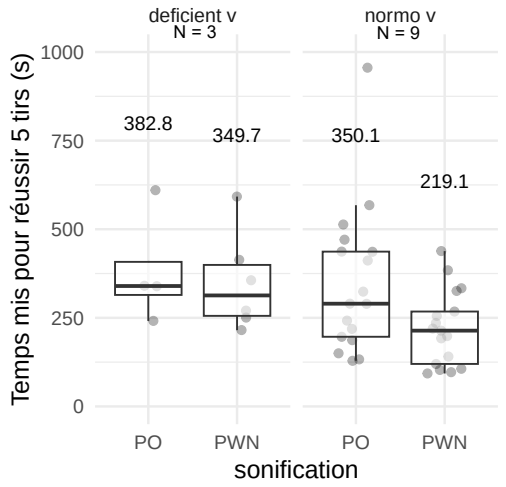
Nous avons comparé le temps mis pour atteindre les cinqs essais réussis dans la phase de tir illimité entre les conditions. Chaque participant a effectué deux essais en temps illimité pour chaque condition testée. Pour la comparaison PSTseq-PSTpar, les participants ont réalisé deux sessions de test : nous comparons également les deux sessions pour tester l'effet de l'entraînement. Les résultats sont tous reportés sur la figure 5.6.

- **PO - PWN** : Il n'y a pas d'effet significatif de la sonification sur le temps mis pour atteindre 5 cibles, ni d'effet du bloc, que ce soit pour les participants normo-voyants ou déficients visuels.
- **PWN - PSTseq** : Il n'y a pas d'effet significatif de la sonification sur le temps mis pour atteindre 5 cibles, ni d'effet du bloc, que ce soit pour les participants normo-voyants ou déficients visuels.
- **PWN - PSTpar**. Participants normo-voyants : Il n'y a pas d'effet significatif de la sonification sur le temps mis pour atteindre 5 cibles, ni d'effet du bloc. Participants déficients visuels : il y a un effet significatif du mode de sonification sur le temps mis pour atteindre les cinq cibles ($F(1,1) = 553{,}8$; $p = 0{,}027$). Les participants sont plus rapides pour atteindre cinq fois la cible avec la sonification PWN qu'avec PSTpar. Il n'y a pas d'effet du bloc.
- **PSTseq - PSTpar**. Participants normo-voyants : il y a un effet significatif de la session sur le temps mis pour atteindre la cible ($F(1,8) = 9{,}11$; $p = 0{,}017$) : les participants étaient plus rapides lors de la deuxième session que lors de la première. Participants déficients visuels : aucun effet significatif de la sonification, de la session ou du bloc.

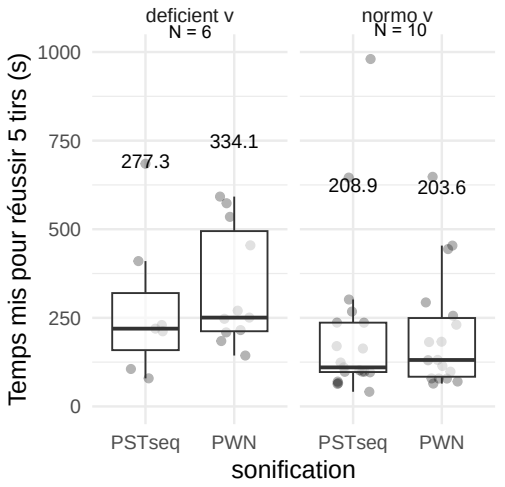
Compromis vitesse-précision

Les scores composites vitesse-précision sont rapportés en figure 5.7.

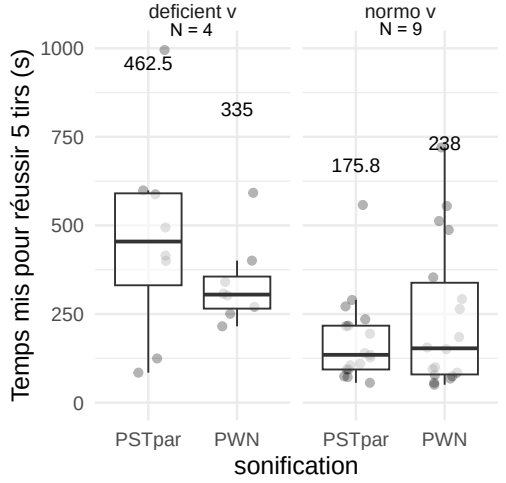
- **PO - PWN** : Les analyses n'ont pas révélé de différence de score vitesse-précision entre les conditions PO et PWN, que ce soit avec temps limité ou temps illimité, pour les participants normo-voyants et malvoyants.
- **PWN - PSTseq** : Pour la phase en temps limité à 50 secondes, les analyses n'ont pas révélé d'effet significatif du mode de sonification ou du bloc, que ce soit pour les participants voyants ou malvoyants. Pour la phase en temps illimité, les analyses ont montré un effet significatif de la sonification ($F(1,9) = 10{,}44$; $p = 0{,}010$), avec de meilleurs scores pour la sonification PSTseq que pour PWN, pour les participants normo-voyants. L'effet du bloc est également significatif ($F(1,9) = 7{,}05$; $p = 0{,}026$), avec de meilleurs scores pour le deuxième bloc, pour les participants normo-voyants. Aucun effet significatif n'a été trouvé pour le groupe des malvoyants.



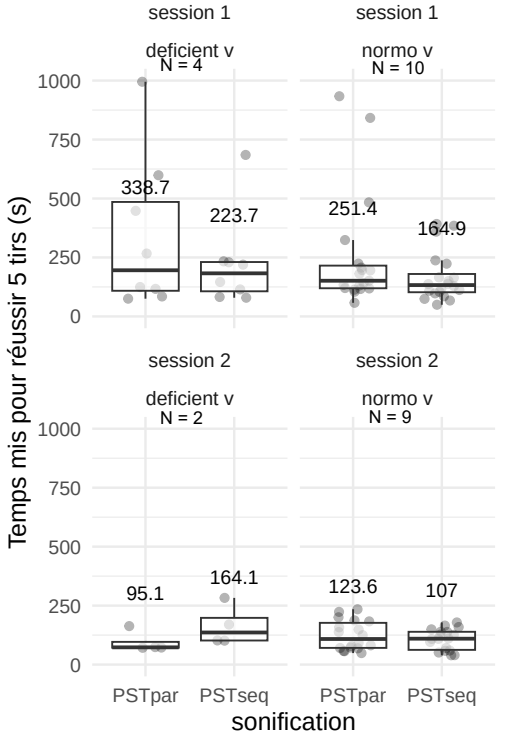
(a) PO - PWN



(b) PWN - PSTseq



(c) PWN - PSTpar



(d) PSTseq - PSTpar

Figure 5.6 – Temps mis pour atteindre la cible 5 fois
Temps mis pour atteindre la cible 5 fois durant la phase de tir en temps illimité, séparé par comparaison de sonification, par groupe de participant et par session. Chaque boîte représente la médiane (ligne horizontale), l'intervalle interquartile (boîte) et les valeurs aberrantes (moustaches). Le temps moyen est affiché au dessus de chaque boîte. Les points représentent les temps de chaque participant.

— **PWN - PSTpar** : Pour la phase en temps limité à 50 secondes, les analyses n'ont pas révélé d'effet significatif du mode de sonification ou du bloc, que ce soit pour les participants voyants ou malvoyants. Pour la phase en temps illimité, un effet du bloc a été trouvé avec de meilleurs scores pour le deuxième bloc pour les participants normo-voyants ($F(1,8) = 41,62$; $p < 0,001$) et pour les participants déficients visuels ($F(1,8) = 39,52$; $p < .001$).

— **PSTseq - PSTpar** : Pour la phase en temps limité à 50 secondes, nous trouvons un effet significatif de la session ($F(1,5) = 24,23$; $p = 0,004$), avec de meilleurs scores vitesse-précision lors de la deuxième session pour les participants normo-voyants. Pour la phase en temps illimité, nous trouvons un effet significatif de la sonification pour les participants normo-voyants ($F(1,8) = 6,79$; $p = 0,031$) et pour les participants malvoyants ($F(1,1) = 164,08$; $p = 0,050$), avec de meilleurs scores pour PSTseq que pour PSTpar. Pour les participants normo-voyants, nous trouvons également un effet de la session ($F(1,8) = 5,77$; $p = 0,043$) avec de meilleurs scores pour la deuxième session, et un effet du bloc ($F(1,8) = 9,07$; $p = 0,017$) avec de meilleurs scores pour le deuxième bloc.

Nombre de tirs réussis en 50 secondes

Les nombres de tirs réussis en moyenne en 50 secondes sont représentés en figure 5.8.

— **PO - PWN** : Pour les participants malvoyants, aucun effet de la sonification ou du bloc n'a été trouvé. Pour les participants normo-voyants, nous trouvons un effet significatif de la sonification ($F(1,7) = 14,91$; $p = 0,006$), avec un nombre de tirs réussis supérieur pour la sonification PWN.

— **PWN - PSTseq** : Pour les participants malvoyants, aucun effet de la sonification ou du bloc n'a été trouvé. Pour les participants normo-voyants, nous trouvons un effet significatif de la sonification ($F(1,9) = 5,16$; $p = 0,049$), avec un nombre de tirs réussis supérieur pour la sonification PSTseq.

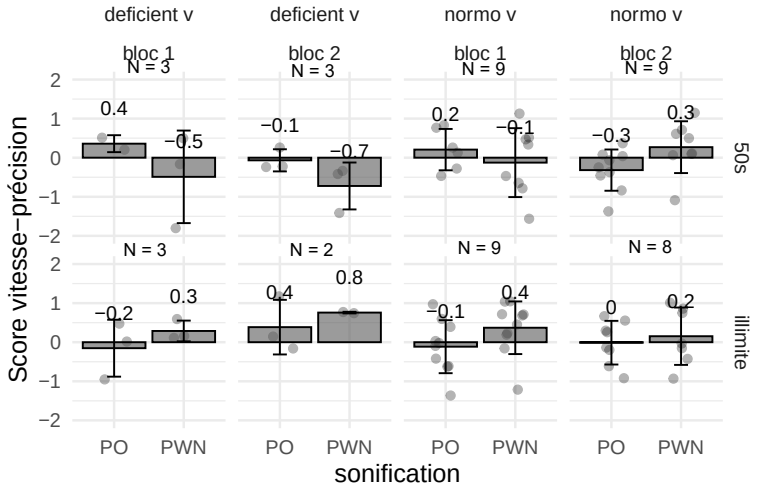
— **PWN - PSTpar** : Pour les participants malvoyants et pour les participants normo-voyants, aucun effet de la sonification ou du bloc n'a été trouvé.

— **PSTseq - PSTpar** : Pour les participants malvoyants, aucun effet de la sonification, du bloc ou de la session n'a été trouvé. Pour les participants normo-voyants, un effet d'interaction a été trouvé entre la sonification et le bloc ($F(1,8) = 7,21$; $p = 0,028$), avec un nombre de tirs réussis supérieur pour la sonification PSTseq au deuxième essai mais pas au premier.

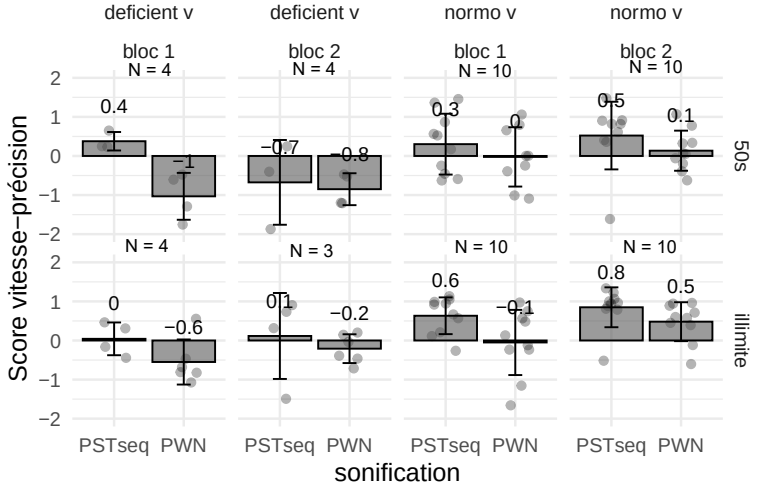
Temps d'atteinte de cible

Les temps mis pour tirer sur la cible (tirs réussis) sont reportés sur la figure 5.9.

— **PO - PWN** : aucun effet significatif de la sonification sur les temps d'atteinte de cible n'a été trouvé, pour les participants normo-voyants ou pour les participants déficients visuels.

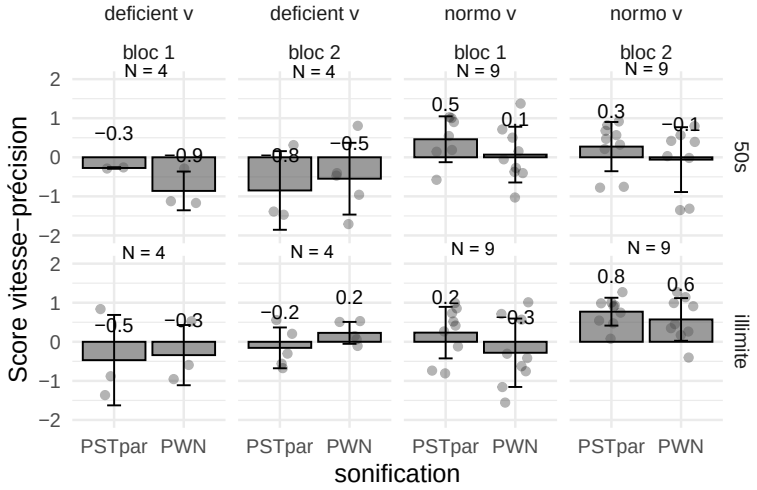


(a) PO - PWN

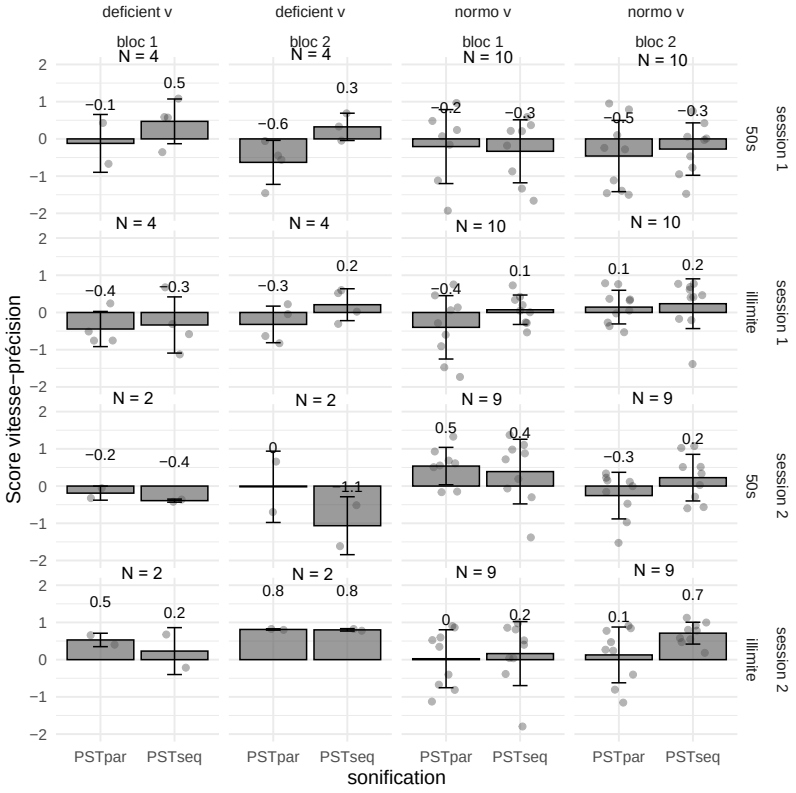


(b) PWN - PSTseq

Figure 5.7 – Scores composites vitesse-précision (partie 1)
Score prenant en compte la vitesse et la précision du tir, séparé par comparaison de sonification, par groupe de participant, par phase (temps illimité ; 50 secondes) par bloc et par session. Les scores sont calculés en additionnant la note z du temps et la note z de la distance à la cible. Les barres d'erreur représentent l'écart-type. Le score moyen est affiché au dessus de chaque barre. Les points représentent les scores de chaque participant.

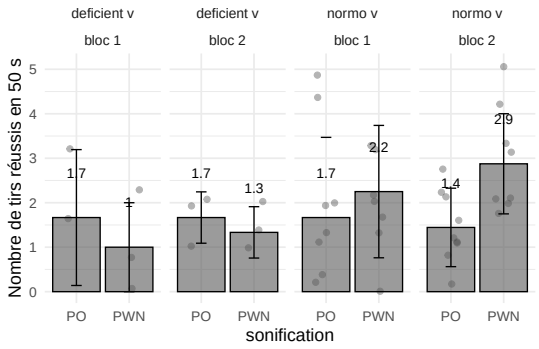


(c) PWN - PSTpar

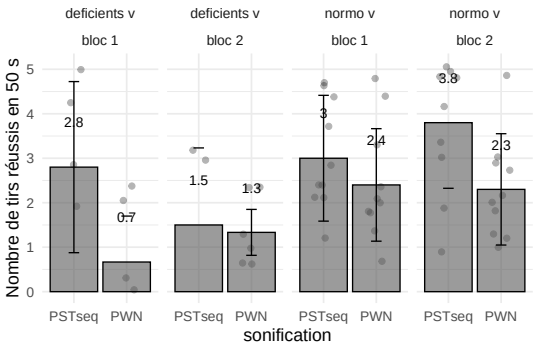


(d) PSTseq - PSTpar

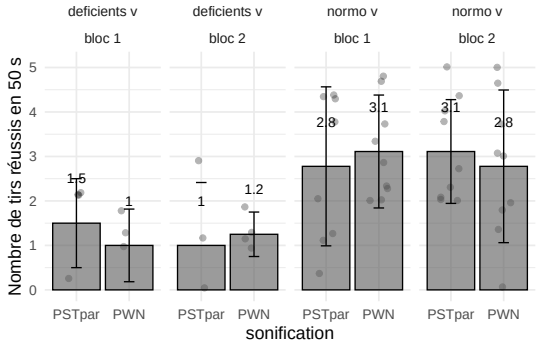
Figure 5.7 – Scores composites vitesse-précision (partie 2)
Score prenant en compte la vitesse et la précision du tir, séparé par comparaison de sonification, par groupe de participant, par phase (temps illimité ; 50 secondes) par bloc et par session. Les barres d'erreur représentent l'écart-type. Le score moyen est affiché au dessus de chaque barre. Les points représentent les scores de chaque participant.



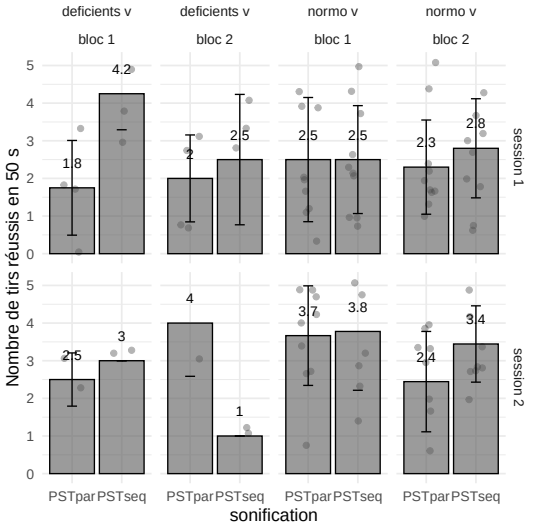
(a) PO - PWN



(b) PWN - PSTseq



(c) PWN - PSTpar



(d) PSTseq - PSTpar

Figure 5.8 – Nombre de tirs réussis en 50 secondes
Nombre de tirs réussis en 50 secondes, séparé par comparaison de sonification, par groupe de participant, par bloc et par session. Les barres d'erreur représentent l'écart-type. Le nombre de tirs réussis moyen est affiché au dessus de chaque barre. Les points représentent le nombre de tir réussi de chaque participant.

— **PWN - PSTseq** : Pour les participants malvoyants, aucun effet de la sonification n'a été trouvé. Pour les participants normo-voyants, nous trouvons un effet significatif de la sonification ($F(1,8) = 11,16$; $p = 0,010$), avec des temps plus faibles pour la sonification PSTseq.
— **PWN - PSTpar** : Pour les participants malvoyants et pour les participants normo-voyants, aucun effet de la sonification n'a été trouvé.
— **PSTseq - PSTpar** : Pour les participants malvoyants, aucun effet de la sonification ou de la session n'a été trouvé. Pour les participants normo-voyants, nous trouvons un effet significatif de la session avec des temps plus faibles pour la deuxième session (F(1,8) = 6,22 ; p = 0,037) et un effet de la sonification avec des temps plus faibles pour PSTseq (F(1,8) = 6,20 ; p = 0,038).

Précision

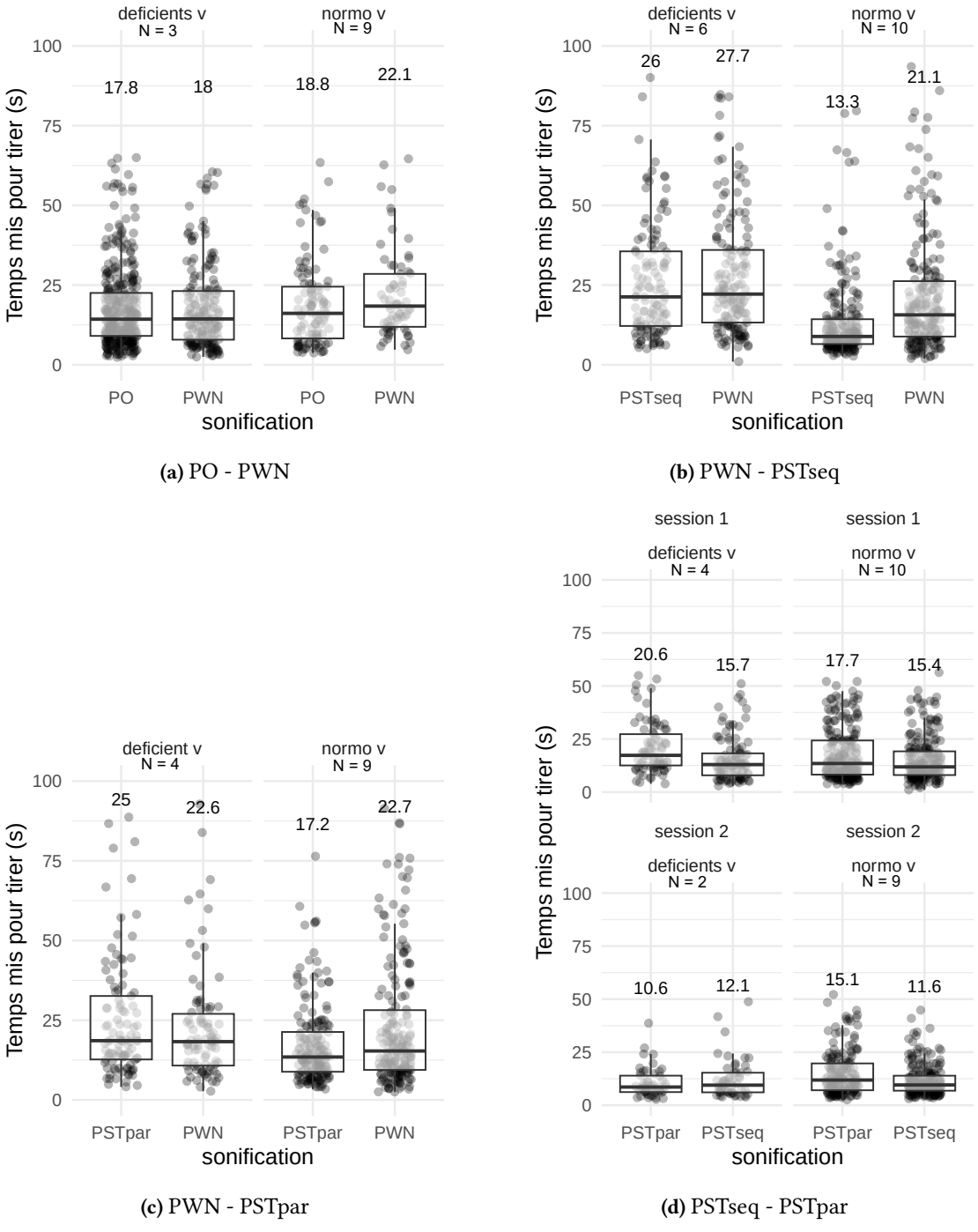
La précision moyenne pour chaque condition est représentée en figure 5.10.

— **PO - PWN** : aucun effet significatif de la sonification sur la précision n'a été trouvé pour les participants déficients visuels. Pour les participants normo-voyants, nous trouvons un effet significatif de la sonification, avec une meilleure précision pour PWN (F(1,7) = 15,35 ; p = 0,006).
— **PWN - PSTseq** : aucun effet significatif de la sonification sur la précision n'a été trouvé, pour les participants normo-voyants ou pour les participants déficients visuels.
— **PWN - PSTpar** : aucun effet significatif de la sonification sur la précision n'a été trouvé, pour les participants normo-voyants ou pour les participants déficients visuels.
— **PSTseq - PSTpar** : aucun effet significatif de la sonification sur la précision n'a été trouvé, pour les participants normo-voyants ou pour les participants déficients visuels.

Scores d'ergonomie

Les scores donnés par les participants normo-voyants sont représentés en Figure 5.11, et ceux des participants malvoyants en Figure 5.12.

— **PO - PWN** : L'analyse de régression ordinale montre un effet significatif de la sonification sur les scores donnés par les participants normo-voyants ($z = 3,82$; $p < .001$), qui ont donné de meilleurs scores à PWN qu'à PO. Il n'y a pas de différence de scores d'ergonomie entre PO et PWN pour les participants malvoyants.
— **PWN - PSTseq** : L'analyse de régression ordinale montre un effet significatif de la sonification sur les scores donnés par les participants normo-voyants (z = 2,37 ; p = 0,018) qui ont donné de meilleurs scores à PSTseq, et par les participants malvoyants (z = -2,30 ; p = 0,021) qui ont donné de meilleurs scores à PWN.



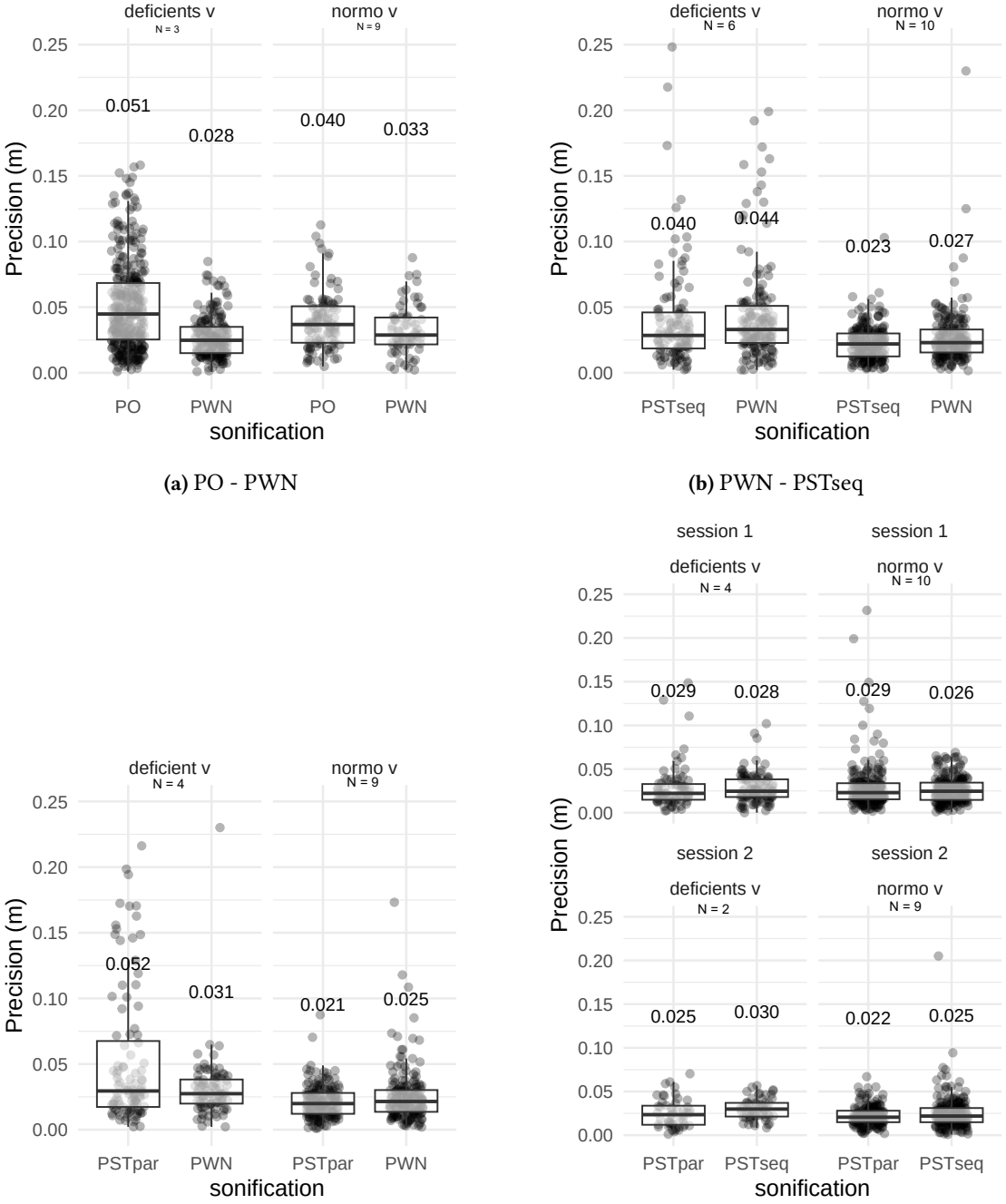
(a) PO - PWN

(b) PWN - PSTseq

(c) PWN - PSTpar

(d) PSTseq - PSTpar

Figure 5.9 – Temps mis pour tirer
Temps mis pour tirer, en secondes, séparé par comparaison de sonification, par groupe de participant et par session. Chaque boîte représente la médiane (ligne horizontale), l'intervalle interquartile (boîte) et les valeurs aberrantes (moustaches). Le temps moyen est affiché au dessus de chaque boîte. Les points représentent les temps de chaque participant.



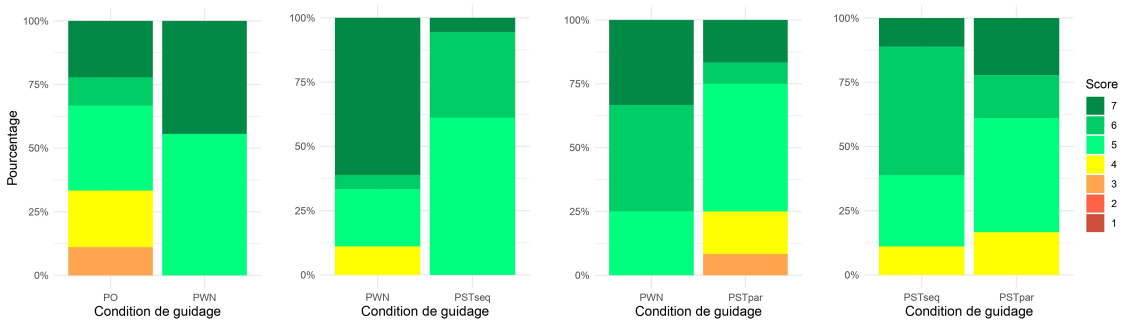
(a) PO - PWN

(b) PWN - PSTseq

(c) PWN - PSTpar

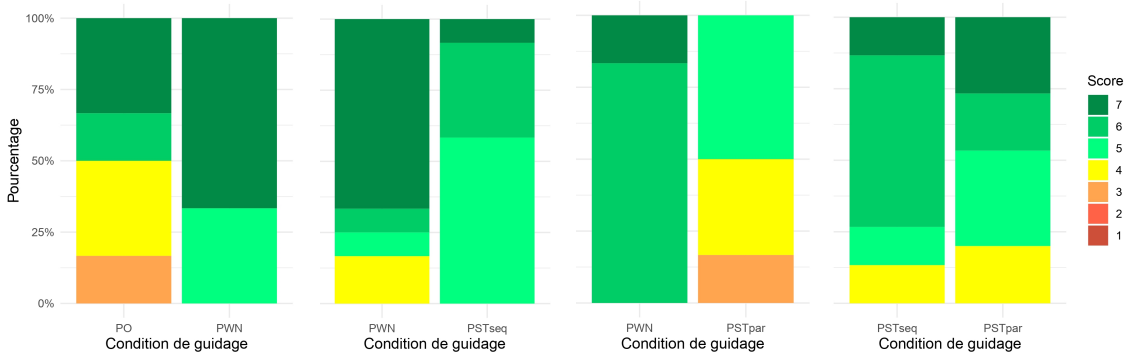
(d) PSTseq - PSTpar

Figure 5.10 – Précision des tirs.
Distance au centre de la cible, en mètre, séparé par comparaison de sonification, par groupe de participant et par session. Chaque boîte représente la médiane (ligne horizontale), l'intervalle interquartile (boîte) et les valeurs aberrantes (moustaches). La distance moyenne est affichée au dessus de chaque boîte. Les points représentent la précision de chaque participant.



(a) PO - PWN **(b)** PWN - PSTseq **(c)** PWN - PSTpar **(d)** PSTseq - PSTpar

Figure 5.11 – Pourcentage de participants normo-voyants ayant donné chaque score à chaque sonification, pour les quatre comparaisons sur la tâche de tir en temps limité.
Les trois critères sont pris en compte (efficacité, confort, facilité). Le gradient de couleur représente les niveaux de notation, le rouge correspondant à 1 et le vert foncé à 7, de sorte que les zones plus vertes correspondent à des jugements plus positifs.



(a) PO - PWN **(b)** PWN - PSTseq **(c)** PWN - PSTpar **(d)** PSTseq - PSTpar

Figure 5.12 – Pourcentage de participants malvoyants ayant donné chaque score à chaque sonification, pour les quatre comparaisons sur la tâche de tir en temps limité.
Les trois critères sont pris en compte (efficacité, confort, facilité). Le gradient de couleur représente les niveaux de notation, le rouge correspondant à 1 et le vert foncé à 7, de sorte que les zones plus vertes correspondent à des jugements plus positifs.

— **PWN - PSTpar** : L'analyse de régression ordinale montre un effet significatif de la sonification sur les scores donnés par les participants normo-voyants ($z = 2{,}04$; $p = 0{,}042$) qui ont donné de meilleurs scores à PSTpar, et par les participants malvoyants ($z = 2{,}28$; $p = 0{,}023$) qui ont donné de meilleurs scores à PWN.

— **PSTseq - PSTpar** : L'analyse de régression ordinale montre un effet significatif de la sonification sur les scores donnés par les participants normo-voyants ($z = 2{,}30$; $p = 0{,}022$), qui ont donné de meilleurs scores à PSTseq. Il n'y a pas de différence de scores d'ergonomie entre PSTseq et PSTpar pour les participants malvoyants. Il n'y a pas de différence de scores d'ergonomie entre la session 1 et la session 2 chez les participants normo-voyants ni chez les participants malvoyants.

5.3.5 Discussion

Dans cette étude, nous avions quatre hypothèses.

Notre première hypothèse était que l'ajout d'un bruit blanc indiquant le centre de la cible (PWN) permettrait de meilleures performances et une ergonomie que la condition contrôle (PO). Cette hypothèse est vérifiée. En effet, les participants normo-voyants ont donné de meilleurs scores d'ergonomie à PWN qu'à PO. De plus, les participants réussissaient plus de tir en phase de tir temps limité à 50 secondes avec la sonification PWN qu'avec la sonification PO. Ce résultat ne se retrouve cependant pas chez les participants déficients visuels. Il est à noter que le nombre de participants déficients visuel est faible pour effectuer des analyses (deux personnes), et que l'absence d'effet significatif, pour ces analyses et pour les suivantes, peut s'expliquer par la taille réduite de l'échantillon. Dans les questionnaires remplis en fin de session expérimentale, les participants déficients visuels ont exprimé une préférence pour le mode avec indication du centre de la cible (PWN). Cependant cette préférence ne s'est pas traduite dans les mesures objectives. Si nous observons le temps mis par les participants pour atteindre les 5 tirs réussis en phase de tir en temps illimité (voir figure 5.6), nous voyons que les participants déficients visuels étaient plus rapides avec PWN (474 s en moyenne) qu'avec PO (820 s en moyenne) au premier bloc, mais qu'il n'y a aucune différence entre PO et PWN au deuxième bloc (340 s et 332 s respectivement). La grande difficulté dans la condition PO était d'identifier et retenir quelle fréquence était la plus aiguë et indiquait le centre de la cible. Comme nous l'avons vu dans l'état de l'art, les personnes déficientes visuelles auraient des capacités de discrimination des fréquences supérieures à celles des personnes normo-voyantes (Arnaud, Gracco et Ménard 2018). Ces capacités de perception auditives supérieures pourraient expliquer l'absence de différence de performance entre les conditions PO et PWN chez les personnes déficientes visuelles.

Notre deuxième hypothèse était que la dissociation des dimensions verticales et horizontales en deux métaphores sonores distinctes (PSTseq et PSTpar) permettrait de meilleures performances et une meilleure ergonomie qu'une sonification ne dissociant pas les dimensions (PWN). Cette hypothèse est en partie vérifiée chez nos participants normo-voyants, pour qui la sonification PSTseq offrait un meilleur compromis vitesse-précision (scores composites) que la soni-

fication PWN, et permettait d'atteindre plus de cible en 50 secondes. De plus, les participants normo-voyants ont donné de meilleurs scores d'ergonomie à PSTseq et à PSTpar qu'à PWN. Cependant, pour les participants déficients visuels, on ne retrouve pas de différence dans les mesures de performance objectives entre PSTseq et PWN, les participants ont été plus rapides pour atteindre cinq fois la cible avec PWN qu'avec PSTpar et ils ont donné de meilleurs scores d'ergonomie à PWN qu'à PSTseq ou PSTpar.

Alors que pour les participants normo-voyants on retrouve un avantage clair à dissocier les dimensions, tant en terme de performance qu'en terme de préférence, cet avantage ne ressort pas clairement chez les personnes déficientes visuelles. Bien que ces dernières aient obtenues de meilleures performances avec PSTseq qu'avec PSTpar, elles ont unanimement préféré la sonification PWN à PSTpar et à PSTseq, une préférence qui s'est reflétée à la fois dans les scores d'ergonomie et dans les discussions avec l'expérimentatrice. Ce résultat pourrait encore une fois s'expliquer par les capacités de discrimination des fréquences accrues des personnes déficientes visuelles, comparé aux personnes normo-voyantes. En effet, les personnes déficientes visuelles ont pu avoir plus de facilité à interpréter les changements de fréquence que les personnes normo-voyantes. De plus, la multiplication des informations avec la dissociation des dimensions pour PSTseq et PSTpar a été vécu comme un surplus, une surcharge d'informations. Avec PSTseq, bien qu'il n'y ait qu'une seule information à interpréter à la fois, il faut retenir plus d'informations, et le passage d'une métaphore sonore à l'autre peut être déroutant. Les personnes déficientes visuelles n'ont pas trouvé cette dissociation utile. Il est intéressant de noter que la moitié des participants déficients visuels et environ 20 % des participants normo-voyants ont rapporté chercher la cible sur une dimension après l'autre même avec PWN, en cherchant le son le plus aigu sur un axe avant de chercher sur l'autre. Cette technique utilisée avec PWN demande de plus grandes capacités de discrimination des fréquences qu'avec PSTseq ou PSTpar, où la stéréo indique la direction où se diriger, et le passage dans la bonne direction. Une proportion plus faible de participants effectuait des petits cercles pour trouver une augmentation de fréquence. Chez les participants normo-voyants, une grande majorité ne rapportait pas de stratégie particulière.

Notre troisième hypothèse était qu'une sonification présentant les dimensions de manière séquentielle (PSTseq) permettrait de meilleures performances et une meilleure ergonomie qu'une sonification présentant les dimensions en parallèle (PSTpar). Cette hypothèse est validée pour les participants normo-voyants, pour qui la sonification PSTseq offrait un meilleur compromis vitesse-précision que PSTpar et permettait d'atteindre la cible un plus grand nombre de fois en 50 secondes. De plus, les participants normo-voyants ont donné de meilleurs scores d'ergonomie à PSTseq qu'à PSTpar. Le meilleur compromis vitesse-précision pour PSTseq est retrouvé également chez les participants déficients-visuels, cependant ceux-ci n'ont pas donné de meilleurs scores d'ergonomie à une condition par rapport à l'autre. L'avantage de la présentation des dimensions l'une après l'autre ressort donc à la fois chez les participants normo-voyants et déficients visuels sur les mesures objectives, même si les participants déficients visuels n'avaient pas de préférence entre les deux. Ces résultats sont très intéressants du point de vue de l'application de ces méthodes de guidage à des conditions de sport réel. En effet, au laser-run, les participants

doivent toucher la cible un certain nombre de fois en un temps limité. La vitesse et la précision sont donc tous deux essentiels. Contrairement à l'étude C1 ou nous n'avons pas pu départager PSTseq (plus rapide) de PSTpar (plus précis), les résultats de l'étude présente montrent un net avantage de PSTseq sur PSTpar sur le compromis vitesse-précision. Cet avantage ressort également d'un point de vue ergonomique pour les participants normo-voyants, qui ont préféré PSTseq. Dans le questionnaire de fin d'expérience, 7 participants sur les 10 normo-voyants, et les 2 participants déficients visuels ont indiqué que pour la sonification PSTpar, ils adoptaient une stratégie similaire à celle imposée par PSTseq, c'est-à-dire chercher à réduire l'écart au centre de la cible sur un axe à la fois. Cela peut vouloir dire que même sans contrainte, c'est la stratégie naturellement adoptée par les participants, ou bien que les participants ont trouvé le mode séquentiel efficace et s'en sont inspiré.

Notre quatrième hypothèse était que l'avantage de la sonification PSTseq par rapport à la sonification PSTpar diminuerait avec l'entraînement. Cette hypothèse n'est pas vérifiée. Si les performances des deux sonifications augmente avec l'entraînement, la différence de performance entre PSTseq et PSTpar n'a pas évolué entre la première et la deuxième session expérimentale.

5.3.6 Conclusion

Dans l'ensemble, nous pouvons dire que le simple ajout d'un signal binaire indiquant le centre de la cible permet d'améliorer la sonification de référence utilisée au tir à la carabine par les athlètes déficients visuels, c'est-à-dire la conversion de la distance à la cible en fréquence. Contrairement à l'étude précédente, nous avons montré un réel avantage à dissocier les axes verticaux et horizontaux en deux métaphores sonores plutôt que d'utiliser un seul paramètre sonore pour coder un écart global. De plus, nous avons montré un avantage à présenter ces métaphores sonores codant les deux axes de manière séquentielle, plutôt que de toutes les fournir en même temps. Cette dissociation temporelle permet aux utilisateurs d'avoir une seule information à traiter à la fois, et s'adapte à la stratégie majoritairement utilisée par les participants qui est de chercher la cible sur une dimension à la fois, ce qui allège la charge cognitive. La sonification séquentielle est celle qui présentait le meilleur compromis vitesse-précision, pour les deux groupes de participants. C'est donc celle qui serait la plus pertinente pour un sport comme le laser-run, qui nécessite de la rapidité et de la précision.

La dissociation des dimensions en métaphores distinctes n'a cependant pas été jugée particulièrement pertinente par les participants déficients visuels, qui n'ont pas apprécié le surplus d'information, qui ne leur a pas permis d'obtenir de meilleures performances. Notre étude a pour limite un échantillon de participants déficients visuels trop faible par rapport à l'échantillon de personnes normo-voyantes. De plus, les deux échantillons présentaient des différences en termes d'âge (les participants déficients visuels étaient plus âgés) et de genre (les participants normo-voyants étaient majoritairement des femmes, les déficients visuels des hommes). Les différences de préférence pourraient donc également être dues à l'âge plus avancé des participants déficients visuels, qui ont une mémoire à court terme moins bonne et ont donc pu avoir plus de mal avec les

encodages dissociés qui demandent de retenir plus de "règles" que les encodages non dissociés.

Dans l'objectif concret de permettre l'accessibilité du laser-run aux personnes déficientes visuelles, ces sonifications devraient être testées dans des conditions plus proches des conditions réelles, en incorporant une phase de course ou de marche rapide avant le tir. De plus, elles devront être testées par des tireurs plus expérimentés. Si l'avantage de la dissociation des dimensions semble robuste à travers les études de cette thèse, pour l'atteinte de cible 3D avec la main et le tir 2D, à la fois du point de vue de la performance et de la charge cognitive, nous pourrons encore améliorer les sonifications, en utilisant des sons plus plaisants.

5.3.7 Résumé de l'étude C2

Cette étude explore la sonification pour améliorer le guidage de la visée dans le tir à l'aveugle, qui est encore très peu explorée dans la littérature et est en plein développement. Contrairement à l'étude précédente, nous avons introduit dans la tâche de tir sur laquelle sont testées les sonifications, une contrainte de temps et de précision. Les quatre modes de sonification testés sont les mêmes sur ceux de l'étude C1. A la méthode traditionnelle qui donne simplement l'écart angulaire en fréquence sonore, nous avons rajouté l'indication de bonne direction visée par un bruit blanc. Nous avons de plus dissocié les dimensions verticales et horizontales des écarts, en délivrant l'information soit simultanément (en parallèle), soit séquentiellement, en tenant compte de l'écart le plus critique.

Comme attendu, l'ajout d'un bruit blanc indiquant le centre de la cible a permis de meilleures performances et une meilleure ergonomie que la condition contrôle. Il y a un avantage à dissocier les dimensions en deux métaphores sonores en terme de performance, et en terme d'ergonomie pour les participants normo-voyants. En revanche, les participants déficients visuels ont préféré la sonification avec un bruit blanc de bonne direction ne dissociant pas les dimension, bien que leurs performances n'aient pas été meilleures dans cette condition. Enfin, nous avons montré l'avantage de présenter les écarts sur chaque dimension séquentiellement plutôt que parallèlement, à la fois pour les participants normo-voyants et déficients visuels. Cette dissociation des dimensions a permis le meilleur compromis vitesse-précision, et permettrait donc les meilleures performances dans un contexte sportif réel.

Synthèse : tir au pistolet guidé par du son

Cette thèse vise à répondre à la question suivante : comment concevoir et évaluer des métaphores sonores pour guider une personne vers une cible sans la vue ? Les deux chapitres précédents portaient sur la conception et l'évaluation des métaphores sonores pour une tâche générique, l'atteinte de cible avec la main.

Ce chapitre étend l'application de ces métaphores sonores à un contexte sportif et ludique, le tir à l'aveugle. Nous voulons donc répondre ici à notre troisième sous-question : Les métaphores optimisées pour l'atteinte de cible 3D avec la main sont-elles transposables à la visée au pistolet, où la précision et la rapidité sont critiques ?

Comme pour l'atteinte de cible avec la main, nous prenons en compte les limites que nous avons relevées dans la littérature :

— Le manque de validation empirique des sonifications (nous voulons valider empiriquement sur une tâche de tir 2D les sonifications testées sur une tâche d'atteinte de cible avec la main 3D dans les séries expérimentales A et B).
— La surcharge cognitive engendrée par des systèmes trop complexes (les sonifications testées ici sont directement tirées des séries A et B, et comme pour les deux autres séries, sont conçues pour être naturelles et intuitives).

Contrairement à l'atteinte de cible avec la main, les études sur la visée du tir au pistolet guidée par le son sont quasi inexistantes (Apavou, Bouchara et Bourdot 2024). Les deux études de ce chapitre s'inscrivent dans la lignée des recommandations de l'état de l'art sur l'utilisation de métaphores sonores intuitives et sur la simplification du système.

Nous testons ici quatre modes de sonifications. PO est la condition contrôle, le mode de sonification utilisé au biathlon au tir à la carabine par les athlètes déficients visuels. PO convertit la distance au centre de la cible en fréquence sonore. PWN ajoute aux variations de fréquence un bruit blanc indiquant le centre de la cible. PSTseq et PSTpar dissocient les dimensions verticales et horizontale : la fréquence et une stéréo discrètes codent la dimension horizontale, des voyelles binaires codent la dimension verticale, et le bruit blanc indique le centre. PSTpar sépare les dimensions en différents paramètres sonores mais présentés simultanément. Avec PSTseq, les dimensions sont présentées séquentiellement, l'une après l'autre.

Les sonifications sont testées par paire sur une tâche de visée et de tir sur une cible fixe dans un espace de capture de mouvement comme les études précédentes.

Dans l'étude C1, le protocole est similaire à celui des séries expérimentales précédentes, avec un nombre d'essais (tirs) fixe. Nous mesurons le temps mis pour tirer et la précision du tir (distance au centre de la cible). Les résultats montrent que toutes nos sonifications

optimisées améliorent la précision et l'ergonomie par rapport à la méthode traditionnelle (PO). L'ajout d'un bruit blanc (PWN) augmente considérablement la précision et l'ergonomie. L'amélioration des performances par la séparation des dimensions (PSTseq et PSTpar) est moins prononcée. La présentation séquentielle des dimensions (PSTseq) permet des tirs plus rapide, mais leur présentation parallèle (PSTpar) permet une meilleure précision.

Dans l'étude C2, le protocole se rapproche des conditions de tir au laser-run, avec l'introduction de contraintes de temps et de précision. Les sonifications sont testées sur des participants normo-voyants, et sur un échantillon de participants déficients visuel plus important que dans les études précédentes. Avec le calcul d'un score composite prenant en compte à la fois le temps mis pour tirer et la distance entre le point de visée et le centre de la cible, nous pouvons déterminer quelle sonification permet le meilleur compromis vitesse-précision. Les résultats montrent un avantage à l'ajout du bruit blanc indiquant le centre de la cible (PWN) par rapport à un simple écart convertit en fréquence (PO), un avantage à dissocier les deux dimensions en métaphores sonores distinctes (PSTseq et PSTpar), et un avantage à présenter les deux métaphores sonores séquentiellement (PSTseq). Ce dernier mode de sonification est celui qui permet le meilleur compromis vitesse-précision, à la fois pour les participants normo-voyants et déficients visuels, bien qu'en terme d'ergonomie, les participants déficients visuels aient indiqué une préférence pour PWN.

Les principes de sonification développés dans la série A sur une tâche générique de marche vers une cible puis d'atteinte avec la main, validés dans la série B sur une tâche simplifiée à la phase d'atteinte avec la main, ont confirmé leurs avantages dans cette série C sur une tâche d'atteinte 2D, du tir au pistolet laser. Nos meilleures sonifications se sont montrées avantageuses en terme de vitesse et de précision de tir, par rapport à la sonification traditionnelle du biathlon qui indique par la fréquence sonore l'écart angulaire globale de la direction visée par rapport à celle de la cible.

6. Discussion générale

Sommaire

6.1 Rappel du contexte et des principaux résultats 154

6.2 Mise en perspective par rapport à l'état de l'art 155

6.2.1 Dissociation des dimensions 155

6.2.2 Présentation séquentielle des dimensions 156

6.2.3 Simplification des métaphores sonores 158

6.2.4 Évaluation des métaphores 159

6.3 Recrutement de participants déficients visuels 160

6.4 Perspectives 161

6.4.1 Perspectives théoriques 161

6.4.2 Perspectives pratiques 162

6.5 Conclusion 164

6.1 Rappel du contexte et des principaux résultats

Dans cette thèse, nous développons des méthodes de sonification permettant d'atteindre une cible avec la main dans l'espace 3D ou sur un plan 2D avec un pistolet laser. Dans ce cadre, nous posons la question suivante : Comment concevoir et évaluer des métaphores sonores pour atteindre une cible sans la vue ? Et plus précisément : quelles métaphores sonores sont les plus efficaces et intuitives pour coder les trois dimensions spatiales sans dépasser les limites cognitives des utilisateurs ? Comment évaluer rigoureusement ces systèmes de guidage ? Les méthodes optimisées pour l'atteinte de cible 3D avec la main sont-elles transposables à la visée au pistolet ?

Ce développement de sonifications va dans le sens d'une simplification des systèmes de substitution sensorielle en tenant compte de la charge cognitive, et d'une validation empirique par comparaison systématique des méthodes. Nous comparons l'efficacité et l'ergonomie de différentes solutions sur des mesures quantitatives (temps d'atteinte de cible, précision) et qualitatives (scores d'ergonomie).

Trois séries expérimentales ont été menées sur un principe commun de jeu virtuel, sur des participants normo-voyants avec les yeux fermés, et sur des participants déficients visuels. L'utilisateur, dont la position est détectée par un système de capture de mouvement, cherchait à atteindre des cibles invisibles, à la main ou en visant au pistolet. A chaque essai, un retour sonore lui indiquait en temps réel les écarts à la cible.

Dans la première série, l'utilisateur se déplaçait pour mettre la main sur des sphères virtuelles positionnées dans l'espace 3D. Après chaque comparaison de sonifications par paire, la meilleure méthode était testée contre une nouvelle. L'analyse des mesures quantitatives (temps d'atteinte de cible) et qualitatives (scores d'ergonomies) a révélé un intérêt à dissocier les trois dimensions spatiales. En effet, la cible est atteinte plus rapidement et est jugée plus ergonomique si des paramètres sonores distincts (fréquence, stéréo, tempo, timbre) codent séparément les écarts dans chaque dimension (azimut, profondeur, hauteur). Les performances et l'ergonomie augmentent encore davantage quand ces métaphores sonores sont délivrées de manière séquentielle plutôt que parallèle, c'est-à-dire une seule dimension donnée à la fois, remplacée par une autre dimension quand l'écart sur la première est passe sous un certain seuil. Ce mode de sonification séquentiel permet de réduire les temps d'atteinte de cible de moitié par rapport à un simple signal binaire donnant uniquement la bonne direction, et est également plus rapide et aussi ergonomique que le guidage vocal d'un guide humain dans les mêmes conditions.

Dans la seconde série expérimentale, la tâche d'atteinte était simplifiée. La position de l'utilisateur était fixe et les cibles dans l'espace péripersonnel du participant. Nous avons fait varier la difficulté de la tâche en manipulant la taille des cibles et leur distance au point de départ de l'utilisateur, afin de pouvoir appliquer la loi de Fitts. Celle-ci prédit que le temps d'atteinte augmente quand la cible est plus petite et plus distante. Les modes de sonification développés dans la première série ont été testés sur cette nouvelle configuration. L'avantage de la dissociation des dimensions et de leur séquentialisation a été répliqué grâce aux métriques classiques (temps

d'atteinte de cible et scores d'ergonomie), mais également grâce au *throughput* (TP), une mesure de performance tirée de la loi de Fitts. Etant indépendant de la configuration de la tâche, cette mesure devrait permettre la comparaison à d'autres études utilisant des configurations différentes (différentes positions et tailles de cible).

Enfin dans la troisième série expérimentale, nous avons transposé les sonifications testées en 3D sur une tâche 2D de tir au pistolet. Nous les avons comparées au système actuellement utilisé au tir à la carabine au biathlon, une simple conversion de la distance globale du point de visée au centre de la cible en fréquence sonore. Nos meilleures sonifications se sont montrées avantageuses en termes de vitesse et précision, et également en termes d'ergonomie, par rapport à cette sonification traditionnelle. Une nouvelle fois, nous avons montré l'avantage de dissocier les dimensions en métaphores sonores distinctes, et à les présenter séquentiellement. Ce résultat, tranché chez les participants normo-voyants yeux fermés, s'est révélé moins saillant chez les participants déficients visuels, qui ont préféré la plus simple de nos sonifications, celle convertissant la distance en fréquence, enrichie d'un bruit blanc indiquant le centre de la cible. Malgré cette préférence, l'avantage de la dissociation des dimensions est ressorti dans les performances objectives (compromis vitesse-précision).

Les trois séries expérimentales révèlent un principe commun dans la conception d'un dispositif de guidage : l'intérêt de dissocier les trois dimensions spatiales en utilisant des métaphores sonores distinctes pour chaque dimension. L'avantage de cette dissociation augmente d'autant plus quand ces métaphores sonores sont présentées séquentiellement (l'une après l'autre) plutôt que parallèlement (toutes en mêmes temps). Les métaphores sonores codant chaque dimension dans ce mode de sonification séquentiel ont été testées au préalable par comparaisons sur des tâches en une dimension. Ces tests ont mis en avant l'importance de simplifier le codage des dimensions (codage discret plutôt que continu) et le rendre intuitif en utilisant les connaissances incarnées des utilisateurs (stéréo, sons verbaux, correspondances cross-modales).

6.2 Mise en perspective par rapport à l'état de l'art

6.2.1 Dissociation des dimensions

Comme nous l'avons souligné dans notre problématique, peu d'études ont mené des comparaisons systématiques de sonifications pour l'atteinte de cible. Beaucoup d'études proposent des systèmes ad hoc et évaluent s'ils sont efficaces ou non plutôt que de tester si un certain système de guidage est plus efficace et ergonomique qu'un autre. Certaines études comparent des modes de guidage utilisant des modalités différentes comme l'audio vs le tactile, mais peu effectuent des comparaisons empiriques entre métaphores sonores.

Quelques études ont comparé une sonification spatialisée seule à une sonification spatialisée à laquelle étaient ajoutés des sons non spatialisés indiquant l'élévation de la cible et/ou la distance à la cible, sur une tâche de pointage de cible ou d'atteinte de cible avec la main. Ces

études ont montré l'avantage de l'ajout de cette sonification non spatialisée au son spatialisé, en termes de performance et de charge cognitive (Gao et al. 2022 ; Lokki et Grohn 2005 ; May et al. 2019). Les résultats de ces études pourraient être interprétés comme un avantage à dissocier les informations sur chaque dimension. Cependant ces études ne comparent pas une sonification ne dissociant pas les dimensions à une sonification dissociant les dimensions, mais une sonification spatialisée à une sonification spatialisée associée à une sonification non spatialisée. Ces études démontrent davantage les limites des capacités du système auditif pour localiser une source sonore, en particulier pour la précision de la localisation verticale. Elles soulignent donc l'avantage de la sonification non spatialisée par rapport à la sonification spatialisée seule et la pertinence du choix que nous avons fait d'utiliser de la sonification non spatialisée pour les tâches d'atteinte de cible, qui nécessitent de la précision

Sur une tâche de visée au pistolet, l'étude de Apavou, Bouchara et Bourdot 2024 ont comparé quatre modes de sonification, dont la sonification classiquement utilisée au biathlon à laquelle était ajouté un son de référence sur la cible. Les participants ont préféré à celle-ci deux sonifications dissociant horizontal et vertical en deux paramètres sonores (tempo pour l'horizontale et fréquence binaire ou continue pour la verticale). Cependant ces préférences ne se sont pas traduites par de meilleures performances de précision ou de rapidité. La quatrième sonification testée dissociait également vertical et horizontal mais utilisait trois paramètres sonores pour coder les deux dimensions : une illusion de Shepard pour coder l'horizontale (montante d'un côté de la cible, descendante de l'autre) avec accélération des changements de fréquence quand la distance à la cible diminue et inversement, et des variations de rugosité (au-dessous de la cible) et des battements (au-dessus de la cible) pour coder la verticale. C'était donc la condition la plus précise, mais c'est celle qui a causé la charge cognitive la plus élevée aux participants, ainsi que les performances les moins bonnes, avec un temps mis pour tirer plus élevé. L'absence d'amélioration des performances malgré la dissociation des dimensions pour les trois sonifications pourrait donc être dû aux métaphores codant chaque dimension. Leurs résultats confirment nos conclusions sur les éléments à prendre en compte dans la conception d'un système de guidage par sonification : dissocier les axes pour permettre la recherche sur un axe à la fois et augmenter la précision, mais utiliser des paramètres simples pour coder ces axes, pour limiter la charge cognitive et faciliter l'apprentissage du système.

6.2.2 Présentation séquentielle des dimensions

Nous sommes allés plus loin que la seule dissociation des axes en métaphores sonores distinctes, en les dissociant également temporellement, ce qui a amélioré les performances et l'ergonomie du guidage. Comme l'ont souligné Apavou, Bouchara et Bourdot 2024, sur la tâche de tir, la dissociation des axes permet une stratégie de recherche axe par axe, que ne permet pas (ou difficilement) la sonification utilisée classiquement au biathlon, qui convertit l'écart global en fréquence sonore.

Nous l'avons observé dans nos études, où beaucoup de participants cherchaient spontanément

la cible sur une dimension à la fois avec un codage global de la direction de la cible en fréquence : ils cherchaient le son le plus aigu sur un axe, puis sur l'autre (tout en avançant vers la cible sur la tâche en 3D). Le fait de dissocier explicitement les dimensions en métaphores sonores distinctes s'adapte à la stratégie naturelle des utilisateurs et soulage les utilisateurs de la charge cognitive de faire eux-mêmes ce travail de dissociation. Nous l'avons observé également en comparant notre mode de sonification séquentiel à un guidage vocal spontané par un guide humain : les instructions données par le guide étaient dans leur quasi-totalité séquentielles, c'est-à-dire que le guide indiquait spontanément au participant des mouvements d'approche sur une dimension à la fois, plutôt que sur une trajectoire en diagonal vers la cible.

Les utilisateurs peuvent également adopter cette stratégie dimension par dimension quand celles-ci sont données par des métaphores sonores distinctes mais fournies en parallèle. Cependant, comme l'ont montré nos résultats, cela peut conduire à un surplus d'information difficile à interpréter, et de moins bonnes performances. Certains utilisateurs ont comparé ça à une « gymnastique cognitive » pour parvenir à interpréter tous les sons en même temps. On peut faire le lien ici avec les résultats obtenus par Apavou, Bouchara et Bourdot 2024 avec leur sonification inspirée de celle utilisée par Ziemer et Schultheis 2018 conçue pour la chirurgie, qui conseillaient de fournir tous les paramètres audios de manière simultanée. Bien que cette sonification dissocie les deux axes, ses performances étaient inférieures à celle de la sonification ne séparant pas les axes et la charge cognitive accrue, car les différentes métaphores sonores étaient trop compliquées à interpréter. En séparant également les dimensions dans le temps, nous gardons la précision du codage de chaque dimension, en supprimant le surplus d'information causé par le codage des autres dimensions dont l'utilisateur n'a pas besoin à cet instant. On s'adapte ainsi d'autant plus à la stratégie intuitive des utilisateurs de chercher la cible sur une dimension à la fois.

La dissociation et la séquentialisation des dimensions ont donc des conséquences bénéfiques d'un point de vue psychologique, notamment en termes de charge cognitive et, par conséquent, d'ergonomie du système. Le fait que cette dissociation semble plus incarnée que l'utilisation d'une seule métaphore sonore codant l'écart global à la cible pourrait s'expliquer par certains mécanismes perceptifs et sensorimoteurs.

En premier lieu, la proprioception, c'est-à-dire la perception de la position et du mouvement du corps, repose sur des récepteurs situés dans les muscles, les tendons et les articulations. Ces récepteurs fournissent des informations distinctes sur le mouvement, comme la position, la vitesse ou l'effort, selon les degrés de liberté articulaires. Cette organisation favorise un traitement séparé des différentes dimensions du mouvement dès les premiers niveaux du système sensoriel. Au niveau cortical, cette séparation est prolongée par l'organisation somatotopique du cortex somatosensoriel, ainsi que par l'existence de neurones sensibles à des directions de mouvement spécifiques. L'intégration de ces informations se fait progressivement dans les aires pariétales et prémotrices, ce qui suggère que les dimensions du mouvement sont d'abord traitées séparément avant d'être combinées en une perception unifiée.

La proprioception est complétée par le système vestibulaire, localisé dans l'oreille interne, qui contribue au sens de l'équilibre et à la perception de l'orientation de la tête par rapport à la gravite. Ce système comprend notamment des récepteurs sensibles aux accélérations linéaires selon des axes distincts (saccule et utricule) ainsi que trois canaux semi-circulaires sensibles aux rotations de la tête selon les trois plans de l'espace (Graf et Klam 2006 ; Lamy 2006).

Enfin, la localisation d'une source sonore repose elle aussi sur des mécanismes distincts selon le plan considéré. La la localisation horizontale repose principalement sur des indices binauraux (ITD et ILD, voir section 2.2.2) alors que la localisation verticale repose sur des informations spectrales plus subtiles qui viennent de la façon dont les sons sont occlus par la tête, les oreilles et les épaules.

L'existence de ces mécanismes perceptifs différenciés selon les dimensions spatiales peut contribuer à expliquer pourquoi la décomposition des dimensions dans la sonification apparaît comme plus ergonomique pour l'utilisateur et permet de meilleures performances.

6.2.3 Simplification des métaphores sonores

Pour concevoir un système intuitif, la simplicité des métaphores sonores choisies pour coder chaque dimension joue un aussi grand rôle que la dissociation de ces dimensions. Pour choisir les paramètres sonores que nous avons utilisés dans nos modes de sonification dissociant les dimensions (mode séquentiel et mode parallèle), nous avons mené deux études en 1D comparant systématiquement différentes métaphores sonores. Celles retenues comme les plus efficaces sont aussi les plus simples. La stéréo joue sur les capacités de localisation de source sonore du système auditif et est donc intuitive, mais est simplifiée par sa présentation discrète qui passe outre les imprécisions de notre système auditif pour localiser une cible. L'imprécision de nos capacités de localisation est la raison des études de Gao et al. 2022 ; Lokki et Grohn 2005 ; May et al. 2019, qui utilisent des métaphores sonores pour pallier ces lacunes et améliorer la précision du guidage. Associée à des variations de fréquence continue, cette stéréo permettait à la fois de savoir de quel côté se diriger, mais également de connaître précisément l'écart à la cible grâce à nos capacités de discrimination fines des fréquences sonores (Scharine, Cave et Letowski 2009 ; Wier, Jesteadt et Green 1977 ; Zwicker et Fastl 2013). Enfin, l'utilisation de voyelles s'appuie sur des connaissances langagières pour faciliter l'interprétation du signal : un [o] va être facilement interprété comme une consigne d'aller vers le haut, et un [a] comme une consigne d'aller vers le bas.

La simplification des métaphores sonores va dans le sens de la distinction entre deux approches décrites par Simmons et al. 2024. Ici nous avons choisi une approche incarnée, qui s'appuie sur des connaissances contextuelles socialement acquises pour faciliter l'interprétation. Cela ne signifie cependant pas nécessairement de rejeter l'approche traditionnelle reposant sur des paramètres classiques, comme la fréquence et l'intensité. Effectivement, les paramètres les plus utilisés le sont souvent pour leur intuitivité, et on sait qu'il existe des correspondances cross-

modales entre la hauteur sonore et la hauteur spatiale, et entre l'intensité sonore ou le tempo et la proximité à la cible. Il faut cependant s'assurer par des tests empiriques de l'efficacité et de l'intuitivité de ces métaphores sonores dans le contexte qui nous intéresse et sur la population cible (Preziosi et Coane 2017 ; Walker et Kramer 2005), car les associations cross-modales dépendent à la fois du contexte et des expériences individuelles (Wilson 2002).

En montrant l'avantage de la dissociation des dimensions et leur présentation séquentielle, et de la simplification des métaphores sonores sur les performances et sur l'ergonomie du guidage, nous nous alignons sur plusieurs théories de la littérature : l'approche incarnée de Simmons et al. 2024 que nous venons d'évoquer ; la compréhension des sonifications reposant sur des connaissances incarnées (Roddy et Bridges 2020) ; l'idée qu'une sonification trop complexe peut saturer la perception et rendre l'interprétation difficile (John H Flowers 2005 ; Wirfs-Brock et al. 2021), et la nécessité d'intégrer des conventions culturelles pour rendre la sonification des données intuitive (Gresham-Lancaster 2012). Nous nous éloignons des premiers travaux sur les dispositifs de substitution visuo-tactile et visuo-auditive, qui cherchaient à remplacer la vue. Nos découvertes vont dans le sens d'une simplification des systèmes en tenant compte de la charge cognitive, et de l'adaptation des systèmes à une tâche spécifique.

6.2.4 Évaluation des métaphores

Dans nos trois séries expérimentales, nous avons comparé nos modes de sonification sur des mesures objectives (temps d'atteinte de cible, précision, compromis vitesse-précision) et subjectives. Dans notre deuxième série expérimentale, nous avons utilisé un autre mode d'évaluation, la loi de Fitts. Elle prédit le temps mis pour atteindre une cible selon la distance à la cible et la taille de la cible. Plus la tâche est difficile (Indice de Difficulté (ID) = log(2A/W), c'est-à-dire plus la taille de la cible (W) est petite et la distance à la cible (A) grande, plus le temps d'atteinte de cible (MT) est élevé (MT = a + b * ID). Nous avons utilisé une métrique issue de cette loi, le *throughput* (TP = 1/b) pour comparer nos sonification. Plus le TP est élevé, moins le temps d'atteinte augmente avec la difficulté de la tâche. Nous avons retrouvé des résultats similaires en comparant nos sonifications avec cette métrique qu'en les comparant avec les métriques traditionnelles, démontrant la pertinence de cet outil de mesure.

Cette métrique, en théorie indépendante des paramètres expérimentaux (taille des cibles, distance à la cible), devrait permettre la comparaison à d'autres études qui testeraient des systèmes de guidage sur d'autres configurations. En effet, la comparaison sur la seule base du temps d'atteinte entre les études est difficile, car les designs expérimentaux varient beaucoup d'une étude à l'autre. En pratique, très peu d'études ont utilisé cette métrique pour évaluer leur système de guidage sonore, et celles qui le font n'utilisent pas nécessairement le même calcul du TP (Hu et al. 2022) ou de l'ID (Marentakis et S. A. Brewster 2006), et beaucoup se contentent de fournir le coefficient de corrélation entre ID et MT (Ahmaniemi et Lantz 2009 ; Charoenchaimonkon et Janecek 2015 ; J. C. Lock et al. 2020) ce qui ne permet pas la comparaison avec nos travaux.

L'utilisation de la loi de Fitts pour évaluer les systèmes de guidage par le son est encore peu répandue, et la comparaison avec d'autres études de la littérature reste compliquée par la diversité des versions de la loi employées, et les diverses version du TP. Pour pallier ce problème, dans nos études utilisant la loi de Fitts, nous avons fourni les équations complètes obtenues pour chaque condition de guidage, ainsi que le détail des tailles de cible, distances à la cible, ID calculés et MT obtenus pour chaque condition de difficulté. Ainsi, si nous n'avons pas pu pour l'instant comparer nos travaux avec d'autres avec la loi de Fitts, d'autres pourront le faire.

Si la comparaison entre études reste compliquée, les métriques traditionnelles comme le temps d'atteinte de cible et la précision restent pertinentes pour comparer différentes modalités de guidage sur un protocole expérimental identique.

6.3 Recrutement de participants déficients visuels

Pour toutes nos séries expérimentales, nous avons d'abord testé nos différentes sonifications sur des participants normo-voyants qui fermaient les yeux (afin que tous les participants effectuent les tests dans les mêmes conditions, sans repère visuel), puis sur des personnes déficientes visuelles. Nous avons en effet tenu à nous rapprocher de l'usager final à chaque étape de notre recherche. Pour nos deux premières séries expérimentales qui testaient des sonifications sur une tâche d'atteinte de cible avec la main, notre échantillon de participants déficients visuel se limitait à un ou deux participants par études, mais nous trouvions à chaque fois des patterns de résultats similaires à ceux des participants normo-voyants.

Nous avons recruté un plus grand nombre de participants pour notre étude finale sur la sonification du tir au pistolet, avec une douzaine de volontaires déficients visuels. Pour cela nous avons contacté plusieurs associations de handisport et de personnes déficientes visuelles. C'est par ces associations que nos participants ont pris connaissance de nos recherches, dont l'aspect ludique a suscité leur enthousiasme. Cela nous encourage d'autant plus dans le développement d'un système pour l'assistance sur une tâche ludique, que les personnes se verront réellement utiliser, plutôt que des systèmes visant à remplacer la vue qui ne tiendront pas leurs promesses. Plusieurs participants déficients visuels nous ont en effet rapporté qu'ils utiliseraient ce système de tir s'il était disponible par exemple en club ou au sein de l'association dont ils font partie. L'enthousiasme des personnes déficientes visuelles pour le tir valide notre approche centrée sur le jeu qui est un terrain d'étude intermédiaire entre le laboratoire et le quotidien.

Contrairement aux participants normo-voyants que nous recevions seulement pour le temps de l'expérience, nous recevions les participants déficients visuels après les tests pour recueillir leur témoignage, discuter de leurs besoins et leurs centres d'intérêt. La moyenne d'âge de ce groupe était supérieure à celle de des participants normo-voyants (58 ans en moyenne contre 25 ans), et les profils plus variés. Par exemple, Monsieur M, 35 ans, déficient visuel depuis sa naissance et éducateur, autonome dans sa vie quotidienne et ses déplacements ; Madame E, 74 ans, complètement aveugle et ayant perdu progressivement la vue entre l'âge de 40 ans et 50

ans, qui s'est faite à son handicap mais dépend des services de transport en minibus dédiés aux personnes à mobilité réduite pour ses déplacements ; Monsieur J, 72 ans, tout récemment déficient visuel et commençant tout juste à s'adapter à son handicap. Grâce à ces discussions, nous avons pu ajuster nos objectifs pour le tournoi de laser-run qui doit conclure le projet SAM-Guide ayant financé cette thèse. La phase de course sera probablement changée en marche pour s'adapter aux usagers.

Avec cet échantillon de participants déficients visuels, nous avons pu mener plusieurs comparaisons de sonifications pour le tir au pistolet. Sur cette tâche, bien que les mesures de performance révèlent un avantage à dissocier les dimensions, les scores d'ergonomie et les discussions avec les participants ont révélé des différences en termes de préférences entre nos participants normo-voyants et nos participants déficients visuels. En effet, ces derniers ont préféré la sonification codant l'écart global au centre de la cible en fréquence, enrichie d'un bruit blanc indiquant le centre la cible, plutôt que nos deux sonifications séparant les deux dimensions en métaphores sonores distinctes (l'une en mode séquentiel, l'autre en mode parallèle). Ces préférences sont à l'opposé de celles de notre groupe de participants normo-voyants. Ce résultat inattendu pourrait être expliqué par le facteur de la déficience visuelle. En effet, on sait que les personnes malvoyantes ont de meilleures capacités d'audition que les personnes normo-voyantes, notamment de meilleures capacités de discrimination des fréquences. Les participants déficients visuels ont ainsi pu se contenter de l'écart global converti en fréquence, qu'ils ont eu moins de mal à interpréter que les participants normo-voyants, ou bien l'augmentation de la quantité d'informations a pu surcharger leurs capacités cognitives. Ce résultat confirme la pertinence de mener des tests sur l'usager final, malgré les difficultés à recruter des participants déficients visuels, moins accessibles que les personnes normo-voyantes.

6.4 Perspectives

6.4.1 Perspectives théoriques

A travers nos trois séries expérimentales, nous avons dégagé un grand principe de la conception d'une sonification pour le guidage vers une cible : l'avantage de la dissociation des dimensions. Cependant, si les métaphores sonores que nous avons utilisées pour chaque dimension ont été testées au préalable et jugées comme étant efficaces et intuitives, il existe des disparités entre les utilisateurs qu'il convient de prendre en compte. En effet, nous avons remarqué que certaines métaphores n'étaient pas interprétées de la même manière selon les personnes. Pour le codage de la hauteur de la cible, nous avons utilisé deux voyelles : un [ɑ] pour indiquer qu'il faut aller vers le bas, et un [o] pour indiquer qu'il faut aller vers le haut. Certaines personnes ont très bien intégré ce principe. D'autres ont mis plus de temps et devaient réfléchir quand elles entendaient la voyelle, à quelle direction elles devaient prendre. Certains participants nous ont indiqué que spontanément, ils interprétaient les voyelles dans l'ordre inverse : le [ɑ] voulait alors dire "tu es trop bas", et le [o] "tu es trop haut". Un phénomène similaire s'est produit pour la stéréo chez

un petit nombre de participants : au lieu d'interpréter un son à droite ou à gauche comme un appel pour aller dans cette direction, certains l'ont interprété dans l'autre sens. Comme l'a noté Wilson 2002, les associations cross-modales dépendent à la fois du contexte et des expériences individuelles, ce qui complique la création de métaphores universellement efficaces.

Dans notre première étude sur le tir, nous avions utilisé les deux voyelles [o] et [ɑ]. Un certain nombre de participants ont indiqué avoir eu du mal à distinguer ces deux sons synthétisés par ordinateur. C'est pourquoi dans notre deuxième étude sur le tir, nous avons changé le [o] en [i] pour améliorer la distinction entre les deux, bien qu'on perde ici l'association des voyelles aux mots « haut » et « bas ». Les deux modes de sonification séquentiel et parallèle utilisant ces voyelles ont permis de meilleures performances et une meilleure ergonomie que des modes de sonification ne dissociant pas les dimensions, mais d'après les commentaires des participants, le codage de la dimension verticale pourrait être encore amélioré. Nous avons évoqué en effet le fait que la dissociation des dimensions ne suffit pas si les métaphores sonores utilisées sont trop complexes, comme c'est le cas dans l'étude de Apavou, Bouchara et Bourdot 2024, dont la sonification inspirée des travaux de Ziemer et Schultheis 2018 a obtenu les moins bonnes performances et la charge cognitive la plus élevée. Le fait que l'effet bénéfique de la dissociation des dimensions ait été moins marqué chez nos participants déficients visuels pourrait être dû au fait qu'ils ont trouvé les métaphores sonores codant chaque dimension trop complexes ou pas assez intuitives.

Cela nous encourage à améliorer nos métaphores en menant davantage de comparaisons et en prenant en compte les remarques des participants. Dans le futur nous pourrons mener d'autres tests pour améliorer nos métaphores. Par exemple, nous pourrons tester de manière rigoureuse s'il vaut mieux fournir les informations de manière binaire (qui indique la direction du mouvement à effectuer) ou de manière continue (qui indique l'amplitude du mouvement à effectuer) ou les deux, pour chaque dimension codée dans la situation testée. De plus, ici, nous avons utilisé uniquement la modalité auditive pour transmettre les informations spatiales. Dans le futur, nous pourrons également tester des métaphores tactiles ou multimodales (auditives et tactiles). En effet, au regard de notre objectif de simplification des métaphores, le multimodal présente plusieurs possibilités : remplacer certaines informations sonores en tactile ce qui permettrait d'alléger la quantité d'informations sonores à interpréter, ou ajouter du tactile en redondance avec le sonore pour plus d'efficacité.

6.4.2 Perspectives pratiques

Nos travaux par comparaisons systématiques montrent l'efficacité et l'ergonomie de sonifications qui dissocient les dimensions, pour guider la main d'une personne privée de vision vers un objet à saisir. Ces résultats nous rapprochent de notre objectif d'une assistance pour certaines tâches du quotidien. En effet, avec les progrès des technologies de capture et de localisation, nos approches optimisées pourraient être intégrées à des applications mobiles. Nos sonifications pourraient ainsi aider les personnes privées de vision dans certaines situations du quotidien,

comme trouver rapidement un article dans un rayon, à retrouver chez eux un objet perdu, à trouver le bon bouton d'interphone devant la porte du médecin, à trouver l'entrée d'un magasin quand le GPS n'est pas assez précis, etc.

Nos travaux ont également montré l'efficacité de nos sonifications optimisées sur une tâche de tir au pistolet laser, comparées au système de tir traditionnellement utilisé en compétition au biathlon par exemple, pour le tir à la carabine. Au biathlon, le système utilisé est très réglementé, et le seul système approuvé pour la compétition est le "EcoAims E00739 E-BSS2006 Biathlon". Comparé au para-biathlon, où les spécifications du système de tir utilisé sont fixes, le laser-run est un sport nouveau, et les spécifications pour l'adapter aux personnes déficientes visuelles sont encore à déterminer. Nous avons vu que la sonification traditionnelle peut être améliorée par la dissociation des dimensions et surtout par l'ajout d'un son indiquant le centre de la cible. Nous espérons que les découvertes théoriques de nos travaux pourront être prises en compte lors du développement de futurs systèmes de tir pour les athlètes paralympiques.

Dans le cadre du projet SAM-Guide qui a financé cette thèse, les travaux que nous avons menés sur le tir pourront être combinés aux travaux des laboratoires partenaires. Ceux-ci portent sur le guidage par une ceinture tactile avec des moteurs vibrants (Faugloire et al. 2022 ; Morice, Mantel et Faugloire 2022), le guidage par du son spatialisé (Ferrand 2021 ; Ferrand, Francois Alouges et Matthieu Aussal 2018 ; Ferrand, François Alouges et Matthieu Aussal 2017, 2019, 2020 ; Ferrand, Aussal et Alouges 2018), la conception d'un système de tir utilisable en extérieur dans des conditions réelles, et sur la détection de lignes d'une piste d'athlétisme (Luo, Boudjit et al. 2025 ; Luo, Ennaji et al. 2025). L'association de ces différents travaux permettra dans un futur proche la création d'un système complet permettant la course (ou la marche) d'une personne déficiente visuelle sur une piste, l'atteinte d'un pas de tir, et le tir sur une cible fixe au pistolet laser. La finalité du projet sera donc de créer un système complet ludique permettant l'accessibilité du laser-run aux personnes déficientes visuelles, tenant compte des découvertes que nous avons faites sur les principes à prendre en compte dans la conception d'un système de guidage par sonification.

Dans un domaine à part, nos travaux ouvrent également des perspectives sur le contrôle d'un robot par la voix. En effet, dans notre première série expérimentale, nous avons comparé notre sonification séquentielle à un guidage vocal spontané sur une tâche d'atteinte de cible avec la main. Cette étude a été menée en collaboration avec un doctorant, Allan Henry, dont les travaux visent à découvrir et explorer les principes fondamentaux de guidage par la voix pour concevoir des dispositifs de contrôle d'un robot par la voix. Cette étude lui a permis d'explorer en premier lieu comment les gens interagissent spontanément pour guider une personne vers un objet 3-D invisible, et de découvrir que les personnes guident spontanément de manière séquentielle, en donnant une direction après l'autre (Fons, Henry et al. 2025 ; Henry et al. 2025).

6.5 Conclusion

Cette thèse a mis en évidence des principes importants pour la conception de métaphores sonores pour atteindre une cible sans la vue : la dissociation des dimensions spatiales (utilisation de métaphores sonores distinctes pour coder les écarts sur chaque dimension) et leur présentation séquentielle (l'une après l'autre), et l'utilisation de métaphores sonore intuitives et incarnées, testées empiriquement par comparaisons. Ces résultats ont été répliqués sur trois tâches : marcher et atteindre une cible avec la main, toucher une cible de la main et tirer sur une cible au pistolet laser. L'application de ces principes permet de réduire la charge cognitive et améliorer les performances (vitesse et précision) sur ces tâches. Il apparaît nécessaire d'optimiser l'information transmise plutôt que de maximiser la quantité d'information. Les résultats vont dans le sens d'une simplification des systèmes en tenant compte de la charge cognitive.

Un apport important de ces travaux est la participation de personnes déficientes visuelles à chaque étape de la recherche. Les différences de préférence des participants déficients visuels comparés aux participants normo-voyants appuient la nécessité d'adapter les métaphores sonores aux capacités et aux préférences des usagers. Au-delà des simples mesures quantitatives, le recueil de leur témoignage a permis de mieux comprendre leurs besoins et ainsi d'ajuster nos objectifs.

Ce travail invite à poursuivre la simplification des systèmes de guidage et leur adaptation à la tâche et au profil des utilisateurs, en tenant compte des préférences individuelles. Il ouvre de nouvelles pistes pour améliorer le guidage des personnes, que ce soit dans un contexte d'assistance au quotidien, ou un contexte plus ludique de sport adapté.

Bibliographie

Abboud, Sami et al. (2014). « EyeMusic : Introducing a "visual" colorful experience for the blind using auditory sensory substitution ». Dans : *Restorative neurology and neuroscience* 32.2, p. 247-257.

Aggius-Vella, Elena, Monica Gori et al. (2022). « Auditory distance perception in front and rear space ». Dans : *Hearing Research* 417, p. 108468.

Aggius-Vella, Elena, Andrew J Kolarik et al. (2020). « Comparison of auditory spatial bisection and minimum audible angle in front, lateral, and back space ». Dans : *Scientific reports* 10.1, p. 6279.

Ahmaniemi, Teemu Tuomas et Vuokko Tuulikki Lantz (2009). « Augmented reality target finding based on tactile cues ». Dans : *Proceedings of the 2009 international conference on Multimodal interfaces*, p. 335-342.

Ahmetovic, Dragan et al. (2020). « Recog : Supporting blind people in recognizing personal objects ». Dans : *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, p. 1-12.

Apavou, Florian, Tifanie Bouchara Bouchara et Patrick Bourdot (2024). « Accessibility of shooting task for blind and visually impaired : A sonification method comparison ». Dans : *29th International Conference on Auditory Display (ICAD 2024)*.

Arfib, Daniel et al. (2002). « Strategies of mapping between gesture data and synthesis model parameters using perceptual spaces ». Dans : *Organised sound* 7.2, p. 127-144.

Arnaud, Laureline, Vincent Gracco et Lucie Ménard (2018). « Enhanced perception of pitch changes in speech and music in early blind adults ». Dans : *Neuropsychologia* 117, p. 261-270.

Auvray, Malika et al. (2005). « There is something out there : distal attribution in sensory substitution, twenty years later ». Dans : *Journal of integrative neuroscience* 4.04, p. 505-521.

Bach-y-Rita, P (1989). « Physiological considerations in sensory enhancement and substitution ». Dans : *Europa Medicophysica* 25, p. 107-128.

Bach-y-Rita, Paul (1967). « Sensory plasticity : Applications to a vision substitution system ». Dans : *Acta Neurologica Scandinavica* 43.4, p. 417-426.

— (1972). « Brain mechanisms in sensory substitution ». Dans : *(No Title)*.

— (1987). « Brain plasticity as a basis of sensory substitution ». Dans : *Journal of Neurologic Rehabilitation* 1.2, p. 67-71.

— (1995). *Nonsynaptic diffusion neurotransmission and late brain reorganization*. Demos Medical Publishing.

— (2003). « Theoretical basis for brain plasticity after a TBI ». Dans : *Brain Injury* 17.8, p. 643-651.

Bach-y-Rita, Paul (2004). « Tactile sensory substitution studies ». Dans : *Annals of the New York Academy of Sciences* 1013.1, p. 83-91.

Bach-y-Rita, Paul, Carter C Collins et al. (1969). « Vision substitution by tactile image projection ». Dans : *Nature* 221.5184, p. 963-964.

Bach-y-Rita, Paul et Stephen W Kercel (2003). « Sensory substitution and the human–machine interface ». Dans : *Trends in cognitive sciences* 7.12, p. 541-546.

Barrass, Stephen (2005). « A perceptual framework for the auditory display of scientific data ». Dans : *ACM Transactions on Applied Perception (TAP)* 2.4, p. 389-402.

Barrera Machuca, Mayra Donaji et Wolfgang Stuerzlinger (2019). « The effect of stereo display deficiencies on virtual hand pointing ». Dans : *Proceedings of the 2019 CHI conference on human factors in computing systems*, p. 1-14. doi : 10.1145/3290605.3300437.

Batmaz, Anil Ufuk et Wolfgang Stuerzlinger (2021). « The effect of pitch in auditory error feedback for fitts' tasks in virtual reality training systems ». Dans : *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. IEEE, p. 85-94.

Bauman, Adrian E et al. (2012). « Correlates of physical activity : why are some people physically active and others not ? » Dans : *The lancet* 380.9838, p. 258-271.

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 ». Dans : *Journal of the audio Engineering Society* 49.10, p. 904-916.

Bernstein, Ira H et Barry A Edelstein (1971). « Effects of some variations in auditory input upon visual choice reaction time. » Dans : *Journal of experimental psychology* 87.2, p. 241.

Best, Virginia et al. (2020). « Sound externalization : A review of recent research ». Dans : *Trends in hearing* 24, p. 2331216520948390.

Blauert, Jens (1997). *Spatial hearing : the psychophysics of human sound localization*. MIT press.

Bonebright, Terri et al. (2010). « Sonification report : status of the field and research agenda ». Dans : *Sonification report : Status of the field and research agenda*.

Brimijoin, W Owen et Michael A Akeroyd (2014). « The moving minimum audible angle is smaller during self motion than during source motion ». Dans : *Frontiers in neuroscience* 8, p. 273.

Brown, David J, Andrew JR Simpson et Michael J Proulx (2014). « Visual objects in the auditory system in sensory substitution : how much information do we need ? » Dans : *Multisensory Research* 27.5-6, p. 337-357.

Brughera, Andrew, Larisa Dunai et William M Hartmann (2013). « Human interaural time difference thresholds for sine tones : The high-frequency limit ». Dans : *The Journal of the Acoustical Society of America* 133.5, p. 2839-2855.

Burton, Matthew J et al. (2021). « The lancet global health commission on global eye health : vision beyond 2020 ». Dans : *The Lancet Global Health* 9.4, e489-e551.

Butorova, Anastasia et Aleksandr Sergeev (2025). « From traditional algorithms to artificial intelligence : a review of the development of sensory substitution sonification methods ». Dans : *The European Physical Journal Special Topics*, p. 1-23.

Capelle, Christian et al. (1998). « A real-time experimental prototype for enhancement of vision rehabilitation using auditory substitution ». Dans : *IEEE Transactions on Biomedical Engineering* 45.10, p. 1279-1293.

Carlini, Alessandro, Camille Bordeau et Maxime Ambard (2024). « Auditory localization : A comprehensive practical review ». Dans : *Frontiers in Psychology* 15, p. 1408073.

Cha, Yeonjoo et Rohae Myung (2013). « Extended Fitts' law for 3D pointing tasks using 3D target arrangements ». Dans : *International Journal of Industrial Ergonomics* 43.4, p. 350-355. doi : 10.1016/j.ergon.2013.05.005.

Charoenchaimonkon, Eakachai et Paul Janecek (2015). « Characterizing non-visual target acquisition tasks with the aid of a tactile display : investigating factors beyond the classical Fitts' theorem ». Dans : *Universal Access in the Information Society* 14, p. 459-475.

Charoenchaimonkon, Eakachai, Paul Janecek et al. (2010). « A comparison of audio and tactile displays for non-visual target selection tasks ». Dans : *2010 International Conference on User Science and Engineering (i-USEr)*. IEEE, p. 238-243.

Choutri, Kheireddine et al. (2022). « A Multi-Lingual Speech Recognition-Based Framework to Human-Drone Interaction ». Dans : *Electronics* 11.12, p. 1829.

Clark, Logan D, Aakash B Bhagat et Sara L Riggs (2020). « Extending Fitts' law in three-dimensional virtual environments with current low-cost virtual reality technology ». Dans : *International journal of human-computer studies* 139, p. 102413. doi : 10.1016/j.ijhcs.2020.102413.

Collins, Carter C et Paul Bach-y-Rita (1973). « Transmission of pictorial information through the skin. » Dans : *Adv. Biol. Med. Phys* 14, p. 285-315.

Contreras, Ruben, Angel Ayala et Francisco Cruz (2020). « Unmanned Aerial Vehicle Control through Domain-Based Automatic Speech Recognition ». en. Dans : *Computers* 9.3, p. 75.

Darcy, Simon, Daniel Lock et Tracy Taylor (2017). « Enabling inclusive sport participation : Effects of disability and support needs on constraints to sport participation ». Dans : *Leisure Sciences* 39.1, p. 20-41.

Deroy, Ophelia et al. (2016). « Differentiated audio-tactile correspondences in sighted and blind individuals. » Dans : *Journal of experimental psychology : human perception and performance* 42.8, p. 1204.

Diaz, Robert et al. (2019). « Impact of adaptive sports participation on quality of life ». Dans : *Sports medicine and arthroscopy review* 27.2, p. 73-82.

Doucet, M-E et al. (2005). « Blind subjects process auditory spectral cues more efficiently than sighted individuals ». Dans : *Experimental brain research* 160.2, p. 194-202.

Drewes, Heiko (2010). « Only One Fitts' Law Formula Please ! » Dans : *CHI '10 Extended Abstracts on Human Factors in Computing Systems*. CHI EA '10. Atlanta, Georgia, USA : Association for Computing Machinery, p. 2813-2822. isbn : 9781605589305. doi : 10.1145/1753846.1753867. url : https://doi.org/10.1145/1753846.1753867.

— (2023). « The Fitts' Law Filter Bubble ». Dans : *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. CHI EA '23. Hamburg, Germany : Association for

Computing Machinery. ISBN : 9781450394222. DOI : 10.1145/3544549.3582739. URL : https://doi.org/10.1145/3544549.3582739.

Dubus, Gaël et Roberto Bresin (2013). « A systematic review of mapping strategies for the sonification of physical quantities ». Dans : *PloS one* 8.12, e82491. DOI : 10.1371/journal.pone.0082491.

Edström, Viking et Fredrik Hallberg (2014). *Human Interaction in 3D Manipulations : Can sonification improve the performance of the interaction ?*

Elipe-Lorenzo, Pablo et al. (2025). « Barriers faced by people with disabilities in mainstream sports : a systematic review ». Dans : *Frontiers in Sports and Active Living* 7, p. 1520962.

Evans, Karla K et Anne Treisman (2010). « Natural cross-modal mappings between visual and auditory features ». Dans : *Journal of vision* 10.1, p. 6-6. DOI : 10.1167/10.1.6.

Faugloire, Elise et al. (2022). « Spatiotemporal influences on the recognition of two-dimensional vibrotactile patterns on the abdomen. » Dans : *Journal of Experimental Psychology : Applied* 28.3, p. 606.

Faul, Franz et al. (2007). « G* Power 3 : A flexible statistical power analysis program for the social, behavioral, and biomedical sciences ». Dans : *Behavior research methods* 39.2, p. 175-191.

Fayjie, Abdur Razzaq et al. (2017). « Voice enabled smart drone control ». Dans : *2017 Ninth International Conference on Ubiquitous and Future Networks (ICUFN)*. Milan : IEEE, p. 119-121.

Ferguson, Jamie et Stephen Brewster (2019). « Modulating personal audio to convey information ». Dans : *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, p. 1-6.

Ferguson, Jamie et Stephen A Brewster (2018). « Investigating perceptual congruence between data and display dimensions in sonification ». Dans : *Proceedings of the 2018 CHI conference on human factors in computing systems*, p. 1-9.

Ferguson, Sam et al. (2006). « Using psychoacoustical models for information sonification ». Dans : *International Conference on Auditory Display*. ICAD.

Ferrand, Sylvain (2021). « Techniques de spatialisation binaurale pour le guidage de sportifs non-voyants ». Thèse de doct. Institut Polytechnique de Paris.

Ferrand, Sylvain, Francois Alouges et Matthieu Aussal (2018). « An augmented reality audio device helping blind people navigation ». Dans : *International Conference on Computers Helping People with Special Needs*. Springer, p. 28-35.

Ferrand, Sylvain, François Alouges et Matthieu Aussal (2017). « Binaural spatialization methods for indoor navigation ». Dans : *Audio Engineering Society Convention 142*.

— (2019). « HRTF and panning evaluations for binaural audio guidance ». Dans : *23rd Internationnal Congress on Acoustics*.

— (2020). « An electronic travel aid device to help blind people playing sport ». Dans : *IEEE Instrumentation & Measurement Magazine* 23.4, p. 14-21.

Ferrand, Sylvain, M Aussal et F Alouges (2018). « Guidage de personnes déficientes visuelles par audio spatialisé ». Dans : *14ème Congrès Francais d'acoustique*.

Fitts, Paul M (1954). « The information capacity of the human motor system in controlling the amplitude of movement. » Dans : *Journal of experimental psychology* 47.6, p. 381. DOI : 10.1037/h0055392.

Fitts, Paul M et James R Peterson (1964). « Information capacity of discrete motor responses. » Dans : *Journal of experimental psychology* 67.2, p. 103. DOI : 10.1037/h0045689.

Fitts, Paul M et Barbara K Radford (1966). « Information capacity of discrete motor responses under different cognitive sets. » Dans : *Journal of Experimental psychology* 71.4, p. 475. DOI : 10.1037/h0022970.

Flowers, John H (2005). « Thirteen years of reflection on auditory graphing : Promises, pitfalls, and potential new directions ». Dans : *International Conference on Auditory Display*.

Fons, Coline, Allan Henry et al. (juill. 2025). « Guiding sightless people one dimension at a time : how the strategy used in spontaneous voice instructions benefits sonification-based guidance ». working paper or preprint. DOI : 10.5281/zenodo.15797531. URL : https://hal.science/hal-05165934.

Fons, Coline, Sylvain Huet, Denis Pellerin, Silvain Gerber et al. (2023). « Moving Towards and Reaching a 3-D Target by Embodied Guidance : Parsimonious Vs Explicit Sound Metaphors ». Dans : *International Conference on Human-Computer Interaction*. Springer, p. 229-243. DOI : 10.1007/978-3-031-35681-0_15.

Fons, Coline, Sylvain Huet, Denis Pellerin et Christian Graff (2024). « Using Fitts' Law to Compare Sonification Guidance Methods for Target Reaching Without Vision ». Dans : *19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications-HUCAPP*. SCITEPRESS-Science et Technology Publications, p. 444-454. DOI : 10.5220/0012346400003660.

— (2025a). « Dissociate Dimensions to Improve Sound Metaphors Guiding the Hand Towards 3D Non-visible Targets ». Dans : *SN Computer Science* 6.4, p. 387. DOI : https://doi.org/10.1007/s42979-025-03877-7.

— (2025b). « Refining sonification methods to improve shooting performance and ergonomics for the visually impaired ». Dans : *The 30th International Conference on Auditory Display (ICAD 2025)*.

Fristot, Vincent et al. (2012). « Depth—Melody substitution ». Dans : *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*. IEEE, p. 1990-1994.

Fryer, Louise, Jonathan Freeman et Linda Pring (2014). « Touching words is not enough : How visual experience influences haptic–auditory associations in the "Bouba–Kiki" effect ». Dans : *Cognition* 132.2, p. 164-173.

Gallace, Alberto et Charles Spence (2006). « Multisensory synesthetic interactions in the speeded classification of visual size ». Dans : *Perception & psychophysics* 68.7, p. 1191-1203.

Gao, Zihan et al. (2022). « Exploring Sonification Mapping Strategies for Spatial Auditory Guidance in Immersive Virtual Environments ». Dans : *ACM Transactions on Applied Perceptions (TAP)*. DOI : 10.1145/3528171.

Garcia, Sara E et al. (2017). « Auditory localisation biases increase with sensory uncertainty ». Dans : *Scientific Reports* 7.1, p. 40567.

Gori, Julien et al. (2018). « The perils of confounding factors : How Fitts' law experiments can lead to false conclusions ». Dans : *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, p. 1-10. DOI : 10.1145/3173574.3173770.

Gougoux, Frédéric et al. (2004). « Pitch discrimination in the early blind : people blinded in infancy have sharper listening skills than those who lost their sight later. » Dans : *Nature*.

Graf, Werner et François Klam (2006). « Le système vestibulaire : anatomie fonctionnelle et comparée, évolution et développement ». Dans : *Comptes Rendus Palevol* 5.3-4, p. 637-655. DOI : 10.1016/j.crpv.2005.12.009.

Grassi, Massimo (2005). « Do we hear size or sound ? Balls dropped on plates ». Dans : *Perception & psychophysics* 67.2, p. 274-284.

Gresham-Lancaster, Scot (2012). « Relationships of sonification to music and sound art ». Dans : *AI & society* 27.2, p. 207-212.

Guarese, Renan et al. (2024). « Exploring audio interfaces for vertical guidance in augmented reality via hand-based feedback ». Dans : *IEEE Transactions on Visualization and Computer Graphics*.

Guiard, Yves et Halla Olafsdottir (oct. 2011). « On the Measurement of Movement Difficulty in the Standard Approach to Fitts' Law ». Dans : *PloS one* 6, e24389. DOI : 10.1371/journal.pone.0024389.

Hamilton-Fletcher, Giles, Michele Mengucci et Francisco Medeiros (2016). « Synaestheatre : sonification of coloured objects in space ». Dans : *Proceedings of the 2016 International Conference on Live Interfaces*, p. 252-256.

Hamilton-Fletcher, Giles, Marianna Obrist et al. (2016). « " I Always Wanted to See the Night Sky" Blind User Preferences for Sensory Substitution Devices ». Dans : *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, p. 2162-2174.

Hamilton-Fletcher, Giles, Michal Pieniak et al. (2020). « Visual Experience influences associations between Pitch and Distance, but not Pitch and Height ». Dans : *Journal of Vision* 20.11, p. 1316-1316. DOI : 10.1167/jov.20.11.1316.

Hanneton, Sylvain, Malika Auvray et Barthélemy Durette (2010). « The Vibe : a versatile vision-to-audition sensory substitution device ». Dans : *Applied Bionics and Biomechanics* 7.4, p. 269-276.

Heimler, Benedetta et al. (2019). « The topo-speech algorithm : an intuitive sensory substitution for spatial information ». Dans : *2019 International Conference on Virtual Rehabilitation (ICVR)*. IEEE, p. 1-2.

Henry, Allan et al. (2025). « Voice Commands for Guidance to a 3D Position : To Collect Spontaneous Data ». Dans : *Proceedings of the 18th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, p. 410-411.

Hild, Michael et Fei Cheng (2014). « Grasping guidance for visually impaired persons based on computed visual-auditory feedback ». Dans : *2014 International Conference on Computer Vision Theory and Applications (VISAPP)*. T. 3. IEEE, p. 75-82. DOI : 10.5220/0004653200750082.

Hillan, Olivia et al. (2023). « Barriers to and facilitators of physical activity : a qualitative study from the perspective of individuals living with sight loss in Cambridgeshire ». Dans : *Vision* 7.4, p. 70.

Hoffmann, Errol R (2013). « Which version/variation of Fitts' law ? A critique of information-theory models ». Dans : *Journal of motor behavior* 45.3, p. 205-215. doi : 10.1080/00222895.2013.778815.

Hu, Xuhui et al. (2022). « StereoPilot : A wearable target location system for blind and visually impaired using spatial audio rendering ». Dans : *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 30, p. 1621-1630. doi : 10.1109/tnsre.2022.3182661.

Jesteadt, Walt, Craig C Wier et David M Green (1977). « Intensity discrimination as a function of frequency and sensation level ». Dans : *The Journal of the acoustical society of America* 61.1, p. 169-177.

Kantan, Prithvi Ravi, Erika G Spaich et Sofia Dahl (2021). « A metaphor-based technical framework for musical sonification in movement rehabilitation ». Dans : *The 26th International conference on auditory display (ICAD 2021)*.

Katz, Brian FG et al. (2012). « NAVIG : augmented reality guidance system for the visually impaired ». Dans : *Virtual Reality* 16.4, p. 253-269. doi : 10.1007/s10055-012-0213-6.

Kilian, Jakob et al. (2022). « The unfolding space glove : A wearable spatio-visual to haptic sensory substitution device for blind people ». Dans : *Sensors* 22.5, p. 1859.

Kiuppis, Florian (2018). « Inclusion in sport : Disability and participation ». Dans : *Sport and disability*. Routledge, p. 4-21.

Kolarik, Andrew J, Silvia Cirstea et Shahina Pardhan (2013). « Evidence for enhanced discrimination of virtual auditory distance among blind listeners using level and direct-to-reverberant cues ». Dans : *Experimental brain research* 224.4, p. 623-633.

Kristjánsson, Árni et al. (2016). « Designing sensory-substitution devices : Principles, pitfalls and potential ». Dans : *Restorative neurology and neuroscience* 34.5, p. 769-787.

Kurakata, Kenji, Tazu Mizunami et Kazuma Matsushita (2013). « Sensory unpleasantness of high-frequency sounds ». Dans : *Acoustical Science and Technology* 34.1, p. 26-33. doi : https://doi.org/10.1250/ast.34.26.

Lamy, Jean-Charles (2006). « Bases neurophysiologiques de la proprioception ». Dans : *Kinésithérapie scientifique* 472, p. 15-23.

Lenay, Charles et al. (2000). « La substitution sensorielle : limites et perspectives ». Dans : *Toucher pour Connaître. Psychologie Cognitive de la Perception Tactile Manuelle. Paris, PUF*, p. 287-306.

Lessard, Nadia et al. (1998). « Early-blind human subjects localize sound sources better than sighted subjects ». Dans : *Nature* 395.6699, p. 278-280.

Letué, Frédérique et al. (2018). « Statistical Methodology for the Analysis of Repeated Duration Data in Behavioral Studies ». Dans : *Journal of Speech, Language, and Hearing Research* 61.3, p. 561-582.

Li, Song et Jürgen Peissig (2020). « Measurement of head-related transfer functions : A review ». Dans : *Applied Sciences* 10.14, p. 5014.

Liu, Yang, Noelle RB Stiles et Markus Meister (2018). « Augmented reality powers a cognitive assistant for the blind ». Dans : *ELife* 7, e37841. doi : 10.7554/eLife.37841.001.

Lock, Jacobus C et al. (2020). « Experimental analysis of a spatialised audio interface for people with visual impairments ». Dans : *ACM Transactions on Accessible Computing (TACCESS)* 13.4, p. 1-21. doi : 10.1145/3412325.

Lokki, Tapio et Matti Grohn (2005). « Navigation with auditory cues in a virtual environment ». Dans : *IEEE MultiMedia* 12.2, p. 80-86. doi : 10.1109/MMUL.2005.33.

Luo, Wenqi, Saadi Boudjit et al. (2025). « Towards Vehicular Cloud Computing for Assistive Navigation : A Wireless Collaborative Vision Prototype ». Dans : *2025 14th IFIP/IEEE International Conference on Performance Evaluation and Modeling in Wired and Wireless Networks (PEMWN)*. IEEE, p. 1-8.

Luo, Wenqi, Abdellatif Ennaji et al. (2025). « Comparison study of different algorithms for Lane line detection ». Dans : *Pattern Recognition and Artificial Intelligence : Selected papers from the 6th Mediterranean Conference on Pattern Recognition and Artificial Intelligence (MedPRAI24)*. Springer, p. 977-988.

MacKenzie, I Scott (1989). « A note on the information-theoretic basis for Fitts' law ». Dans : *Journal of motor behavior* 21.3, p. 323-330. doi : 10.1080/00222895.1989.10735486.

— (1992). « Fitts' law as a research and design tool in human-computer interaction ». Dans : *Human-computer interaction* 7.1, p. 91-139. doi : 10.1207/s15327051hci0701_3.

Maeda, Fumiko, Ryota Kanai et Shinsuke Shimojo (2004). « Changing pitch induced visual motion illusion ». Dans : *Current biology* 14.23, R990-R991.

Majdak, Piotr, Robert Baumgartner et Claudia Jenny (2020). « Formation of three-dimensional auditory space ». Dans : *The technology of binaural understanding*. Springer, p. 115-149.

Makous, James C et John C Middlebrooks (1990). « Two-dimensional sound localization by human listeners ». Dans : *The journal of the Acoustical Society of America* 87.5, p. 2188-2200.

Manduchi, Roberto et James M Coughlan (2014). « The last meter : blind visual guidance to a target ». Dans : *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, p. 3113-3122. doi : 10.1145/2556288.2557328.

Marentakis, Georgios N et Stephen A Brewster (2006). « Effects of feedback, mobility and index of difficulty on deictic spatial audio target acquisition in the horizontal plane ». Dans : *Proceedings of the SIGCHI conference on Human Factors in computing systems*, p. 359-368. doi : 10.1145/1124772.1124826.

Marston, James R et al. (2007). « Nonvisual route following with guidance from a simple haptic or auditory display ». Dans : *Journal of Visual Impairment & Blindness* 101.4, p. 203-211. doi : 10.1177/0145482X0710100403.

Martínez, Brian David Cano et al. (2011). « Visual perception substitution by the auditory sense ». Dans : *International Conference on Computational Science and Its Applications*. Springer, p. 522-533.

Martino, Gail et Lawrence E Marks (1999). « Perceptual and linguistic interactions in speeded classification : Tests of the semantic coding hypothesis ». Dans : *Perception* 28.7, p. 903-923.

May, Keenan R et al. (2019). « Auditory displays to facilitate object targeting in 3D space ». Dans : *The 25th International Conference on Auditory Display (ICAD 2019)*. Georgia Institute of Technology, p. 155-162. DOI : 10.21785/icad2019.008.

Meijer, Peter BL (1992). « An experimental system for auditory image representations ». Dans : *IEEE transactions on biomedical engineering* 39.2, p. 112-121. DOI : 10.1109/10.121642.

Melara, Robert D et Thomas P O'Brien (1987). « Interaction between synesthetically corresponding dimensions. » Dans : *Journal of Experimental Psychology : General* 116.4, p. 323.

Middlebrooks, John C (2015). « Sound localization ». Dans : *Handbook of clinical neurology* 129, p. 99-116. DOI : 10.1016/B978-0-444-62630-1.00006-8.

Mills, Allen William (1958). « On the minimum audible angle ». Dans : *The Journal of the Acoustical Society of America* 30.4, p. 237-246.

Morice, Antoine HP, Bruno Mantel et Elise Faugloire (2022). « Assistance vibrotactile à l'interception locomotrice de cibles mobiles : étude de faisabilité ». Dans : *56e congrès de la Société d'Ergonomie de Langue Française (SELF 2022) : Vulnérabilités et risques émergents : penser et agir ensemble pour transformer durablement*, p. 348-353.

Murata, Atsuo et Hirokazu Iwase (2001). « Extending Fitts' law to a three-dimensional pointing task ». Dans : *Human movement science* 20.6, p. 791-805. DOI : 10.1016/S0167-9457(01)00058-6.

Netzer, Ophir et al. (2021). « Backward spatial perception can be augmented through a novel visual-to-auditory sensory substitution algorithm ». Dans : *Scientific Reports* 11.1, p. 11944.

Olafsdottir, Halla H et al. (2012). « A new test of throughput invariance in Fitts' law : Role of the intercept and of Jensen's inequality ». Dans : *26th BCS Conference on Human Computer Interaction (HCI 2012)*. DOI : 10.14236/ewic/hci2012.17.

Parise, Cesare Valerio et Charles Spence (2009). « 'When birds of a feather flock together' : Synesthetic correspondences modulate audiovisual integration in non-synesthetes ». Dans : *PloS one* 4.5, e5664.

Parseihian, Gaëtan et al. (2016). « Comparison and evaluation of sonification strategies for guidance tasks ». Dans : *IEEE Transactions on Multimedia* 18.4, p. 674-686. DOI : 10.1109/TMM.2016.2531978.

Pedley, Philip E et Robert S Harper (1959). « Pitch and the vertical localization of sound ». Dans : *The American Journal of Psychology* 72.3, p. 447-449. DOI : 10.2307/1420051.

Pedrielli, Francesca, Eleonora Carletti et Camilla Casazza (2008). « Just noticeable differences of loudness and sharpness for earth moving machines ». Dans : *Journal of the Acoustical Society of America* 123.5, p. 3164-3164.

Poncela, Alberto et Leticia Gallardo-Estrella (2015). « Command-Based Voice Teleoperation of a Mobile Robot via a Human-Robot Interface ». Dans : *Robotica* 33.1, p. 1-18.

Pratt, Carroll C (1930). « The spatial character of high and low tones. » Dans : *Journal of Experimental Psychology* 13.3, p. 278. DOI : 10.1037/h0072651.

Preziosi, Melissa A et Jennifer H Coane (2017). « Remembering that big things sound big : Sound symbolism and associative memory ». Dans : *Cognitive Research : Principles and Implications* 2.1, p. 10.

Radford, Alec et al. (2023). « Robust Speech Recognition via Large-Scale Weak Supervision ». Dans : *Preprint*.

Ramachandran, Vilayanur S et Edward M Hubbard (2001). « Synaesthesia–a window into perception, thought and language ». Dans : *Journal of consciousness studies* 8.12, p. 3-34.

— (2003). « Hearing colors, tasting shapes ». Dans : *Scientific American* 288.5, p. 52-59.

Richardson, Michael et al. (2019). « How much spatial information is lost in the sensory substitution process ? Comparing visual, tactile, and auditory approaches ». Dans : *Perception* 48.11, p. 1079-1103.

Roddy, Stephen et Brian Bridges (2020). « Mapping for meaning : the embodied sonification listening model and its implications for the mapping problem in sonic information design ». Dans : *Journal on Multimodal User Interfaces* 14.2, p. 143-151.

Rokem, Ariel et Merav Ahissar (2009). « Interactions of cognitive and auditory abilities in congenitally blind individuals ». Dans : *Neuropsychologia* 47.3, p. 843-848.

Rusconi, Elena et al. (2006). « Spatial representation of pitch height : the SMARC effect ». Dans : *Cognition* 99.2, p. 113-129. doi : 10.1016/j.cognition.2005.01.004.

Ruvolo, Paul (2021). « Considering Spatial Cognition of Blind Travelers in Utilizing Augmented Reality for Navigation ». Dans : *2021 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*. IEEE, p. 99-104. doi : 10.1109/PerComWorkshops51409.2021.9430997.

Sapir, Edward (1929). « A study in phonetic symbolism. » Dans : *Journal of experimental psychology* 12.3, p. 225.

Scharine, A, K Cave et T Letowski (2009). « Auditory perception and cognitive performance ». Dans : *Helmet-mounted displays : Sensation, perception, and cognition issues*, p. 391-489.

Shaw, Edgar AG (1974). « Transformation of sound pressure level from the free field to the eardrum in the horizontal plane ». Dans : *The Journal of the Acoustical Society of America* 56.6, p. 1848-1861.

Shih, Meng-Li et al. (2018). « Dlwv2 : A deep learning-based wearable vision-system with vibrotactile-feedback for visually impaired people to reach objects ». Dans : *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, p. 1-9. doi : 10.1109/IROS.2018.8593711.

Simmons, Joe et al. (2024). « The ballad of the bots : sonification using cognitive metaphor to support immersed teleoperation of robot teams ». Dans : *Frontiers in Virtual Reality* 5, p. 1404865.

Smith, Linda B et Maria D Sera (1992). « A developmental analysis of the polar structure of dimensions ». Dans : *Cognitive psychology* 24.1, p. 99-142.

Song, Qi et al. (2022). « Combining spatial and non-spatial sonification for sound localization in virtual environment : a subjective study ». Dans : *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, p. 1-7.

Soukoreff, R William et I Scott MacKenzie (2004). « Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI ». Dans : *International journal of human-computer studies* 61.6, p. 751-789. doi : 10.1016/j.ijhcs.2004.09.001.

Spence, Charles (2011). « Crossmodal correspondences : A tutorial review ». Dans : *Attention, Perception, & Psychophysics* 73.4, p. 971-995.

Stevens, SS (1957). « On the physiological law ». Dans : *Psychol. Rev* 64, p. 153-181.

Stevens, Stanley Smith et Edwin Broomell Newman (1936). « The localization of actual sources of sound ». Dans : *The American journal of psychology* 48.2, p. 297-306.

Stoll, Chloé et al. (2015). « Navigating from a depth image converted into sound ». Dans : *Applied bionics and biomechanics* 2015.1, p. 543492.

Tabry, Vanessa, Robert J Zatorre et Patrice Voss (2013). « The influence of vision on sound localization abilities in both the horizontal and vertical planes ». Dans : *Frontiers in psychology* 4, p. 932.

Teather, Robert J et al. (2009). « Effects of tracking technology, latency, and spatial jitter on object movement ». Dans : *2009 IEEE symposium on 3D user interfaces*. IEEE, p. 43-50. doi : 10.1109/3dui.2009.4811204.

Thakoor, Kaveri et al. (2014). « A system for assisting the visually impaired in localization and grasp of desired objects ». Dans : *European Conference on Computer Vision*. Springer, p. 643-657. doi : 10.1007/978-3-319-16199-0_45.

Troncoso Aldas, Nelson Daniel et al. (2020). « AIGuide : An Augmented Reality Hand Guidance Application for People with Visual Impairments ». Dans : *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility*, p. 1-13. doi : 10.1145/3373625.3417028.

Tyler, Mitchell, Yuri Danilov et Paul Bach-y-Rita (2003). « Closing an open-loop control system : vestibular substitution through the tongue ». Dans : *Journal of integrative neuroscience* 2.02, p. 159-164.

Vickers, Paul (2016). « Sonification and music, music and sonification ». Dans : *The Routledge companion to sounding art*. Routledge, p. 135-144.

Vickers, Paul et al. (2014). « Sonification aesthetics and listening for network situational awareness ». Dans : *arXiv preprint arXiv :1409.5282*.

Voong, Tray Minh et Michael Oehler (2019). « Auditory spatial perception using bone conduction headphones along with fitted head related transfer functions ». Dans : *2019 IEEE conference on virtual reality and 3D user interfaces (VR)*. IEEE, p. 1211-1212.

Voss, Patrice, Vanessa Tabry et Robert J Zatorre (2015). « Trade-off in the sound localization abilities of early blind individuals between the horizontal and vertical planes ». Dans : *Journal of Neuroscience* 35.15, p. 6051-6056.

Walker, Bruce N et Gregory Kramer (2005). « Mappings and metaphors in auditory displays : An experimental assessment ». Dans : *ACM Transactions on Applied Perception (TAP)* 2.4, p. 407-412.

Wan, Catherine Y et al. (2010). « Early but not late-blindness leads to enhanced auditory perception ». Dans : *Neuropsychologia* 48.1, p. 344-348.

Ward, Jamie et Peter Meijer (2010). « Visual experiences in the blind induced by an auditory sensory substitution device ». Dans : *Consciousness and cognition* 19.1, p. 492-500.

Welford, Alan Traviss (1970). « Fundamentals of skill. » Dans : *The American journal of Psychology* 83. doi : 10.2307/1421343.

Wenzel, Elizabeth M et al. (1993). « Localization using nonindividualized head-related transfer functions ». Dans : *The Journal of the Acoustical Society of America* 94.1, p. 111-123.

White, Benjamin W et al. (1970). « Seeing with the skin ». Dans : *Perception & Psychophysics* 7, p. 23-27.

Wier, Craig C, Walt Jesteadt et David M Green (1977). « Frequency discrimination as a function of frequency and sensation level ». Dans : *The Journal of the Acoustical Society of America* 61.1, p. 178-184.

Wightman, Frederic L et Doris J Kistler (1989). « Headphone simulation of free-field listening. II : Psychophysical validation ». Dans : *The Journal of the Acoustical Society of America* 85.2, p. 868-878.

— (1997). « Factors affecting the relative salience of sound localization cues ». Dans : *Binaural and spatial hearing in real and virtual environments* 1, p. 1-23.

Wilson, Margaret (2002). « Six views of embodied cognition ». Dans : *Psychonomic bulletin & review* 9.4, p. 625-636.

Wirfs-Brock, Jordan et al. (2021). « Examining narrative sonification : using first-person retrospection methods to translate radio production to interaction design ». Dans : *ACM Transactions on Computer-Human Interaction (TOCHI)* 28.6, p. 1-34.

Yost, William A (2016). « Sound source localization identification accuracy : Level and duration dependencies ». Dans : *The Journal of the Acoustical Society of America* 140.1, EL14-EL19.

Zahorik, Pavel (2002). « Assessing auditory distance perception using virtual acoustics ». Dans : *The Journal of the Acoustical Society of America* 111.4, p. 1832-1846.

Zhai, Shumin (2004). « Characterizing computer input with Fitts' law parameters—the information and non-information aspects of pointing ». Dans : *International Journal of Human-Computer Studies* 61.6, p. 791-809. doi : 10.1016/j.ijhcs.2004.09.006.

Ziemer, Tim, Nuttawut Nuchprayoon et Holger Schultheis (2019). « Psychoacoustic sonification as user interface for human-machine interaction ». Dans : *arXiv preprint arXiv :1912.08609*. doi : 10.48550/arXiv.1912.08609.

Ziemer, Tim et Holger Schultheis (2018). « A psychoacoustic auditory display for navigation ». Dans : *The 24th International Conference on Auditory Display (ICAD 2018)*. Georgia Institute of Technology, p. 136-144. doi : 10.21785/icad2018.007.

— (2019a). « Psychoacoustical signal processing for three-dimensional sonification ». Dans : *The 25th International Conference on Auditory Display (ICAD 2019)*. Georgia Institute of Technology, p. 277-284. doi : 10.21785/icad2019.018.

— (2019b). « Three orthogonal dimensions for psychoacoustic sonification ». Dans : *arXiv preprint arXiv :1912.00766*. doi : 10.48550/arXiv.1912.00766.

Zientara, Peter A et al. (2017). « Third eye : A shopping assistant for the visually impaired ». Dans : *Computer* 50.2, p. 16-24. doi : 10.1109/mc.2017.36.

Zwicker, Eberhard et Hugo Fastl (2013). *Psychoacoustics : Facts and models*. T. 22. Springer Science & Business Media.

7. Annexes

7.1 Annexe 1 : Questionnaire sur l'ergonomie

A remplir par l'expérimentateur Thème :
Numéro participant :
Date : Accueil à : __h__ Fin à : __h__ Lieu :
Gratification :

1. Participant : Age : … … Sexe : …..
Main dominante : main droite / main gauche ?
Participation précédente à une expérience semblable (oui ou non) :
Troubles sensori-moteurs ou autres en lien avec l'expérience :

2. Difficultés particulières rencontrées au cours de l'expérience (techniques ou liées au participant) :

2. Commentaires généraux : quel est votre ressenti sur l'expérience ? Était-elle amusante, oppressante, pénible ? Les explications étaient-elles claires ? Précisez les points forts ou faibles motivant ces impressions.

4. Jouez-vous régulièrement aux jeux vidéos ? OUI / NON
5. Utilisez-vous régulièrement un casque de réalité virtuelle ? OUI / NON
6. Notez chacune des conditions (de façon absolue et comparée) sur une échelle allant de 1 (très mauvais) à 7 (très bon) :

Parallèle	Efficacité apparente : pas du tout efficace 1 2 3 4 5 6 7 très efficace Confort ressenti : pas du tout confortable 1 2 3 4 5 6 7 très confortable Facilité à employer le dispositif : pas du tout facile 1 2 3 4 5 6 7 très facile *Stratégies employées :* *Autres commentaires sur votre ressenti :*

A remplir par l'expérimentateur Thème :
Numéro participant :
Date : Accueil à : __h__ Fin à : __h__ Lieu :
Gratification :

Séquentiel	Efficacité apparente : pas du tout efficace **1 2 3 4 5 6 7** très efficace Confort ressenti : pas du tout confortable **1 2 3 4 5 6 7** très confortable Facilité à employer le dispositif : pas du tout facile **1 2 3 4 5 6 7** très facile ***Stratégies employées pour chercher la cible :*** ***Autres commentaires sur votre ressenti :***

7. Suggestions, propositions pour améliorer le dispositif :

7.2 Annexe 2 : Liste des tableaux

Liste des tableaux

3.1 Métaphores sonores utilisées pour la sonification BI 42

3.2 Métaphores sonores utilisées pour la sonification AP 43

3.3 Métaphores sonores utilisées pour la sonification DVH 44

3.4 Métaphores utilisées pour les deux modes de sonification séquentiel et parallèle 72

4.1 Conditions de difficulté de la tâche de Fitts. A représente la distance à la cible (en centimètres), W le diamètre de la cible (en centimètres), et ID l'indice de difficulté (en bits). MT représente le temps (durée) de mouvement ("Movement Time") ou temps d'atteinte de cible (en secondes) obtenu pour chaque condition de difficulté. 94

4.2 Métaphores sonores utilisées pour la sonification AP sur la tâche de Fitts 95

4.3 Métaphores sonores utilisées pour la sonification DVH sur la tâche de Fitts . . . 96

4.4 Conditions de difficulté de la tâche de Fitts verticale et temps d'atteinte de cible. A représente la distance à la cible (en centimètres), W l'épaisseur de la cible (en centimètres), et ID l'indice de difficulté (en bits). MT représente le temps de mouvement ou temps d'atteinte de cible (en secondes) obtenu pour chaque condition de difficulté, pour chaque métaphore. 105

4.5 Conditions de difficulté de la tâche de Fitts horizontale et temps d'atteinte de cible. A représente la distance à la cible (en degrés), W la largeur de la cible (en degrés), et ID l'indice de difficulté (en bits). MT représente le temps de mouvement ou temps d'atteinte de cible (en secondes) obtenu pour chaque condition de difficulté, pour chaque métaphore. 105

4.6 Equations de régression et résultats du test de régression entre MT et ID pour chaque métaphore pour les deux tâches verticale et horizontale. Le TP est présenté, en bits/s. 105

4.7 Métaphores sonores utilisées pour les deux modes de sonification séquentiel et parallèle sur la tâche de Fitts. 107

4.8 Temps d'atteinte de cible moyen pour chaque condition de difficulté des sonifications parallèle et séquentielle. 109

4.9 Résultats de l'analyse de régression entre ID et MT pour les deux conditions de sonification séquentiel et parallèle. 109

5.1 Métaphores sonores utilisées pour la sonification PO 122

5.2 Métaphores sonores utilisées pour la sonification PWN 122

5.3 Métaphores sonores utilisées pour les deux modes de sonification séquentiel et parallèle de la tâche de tir 122

7.3 Annexe 3 : Liste des figures

Table des figures

1.1 Dispositif expérimental utilisé dans toutes les études présentées dans cette thèse. 4

1.2 Illustration des trois tâches d'atteinte de cible guidée par le son correspondant aux trois séries expérimentales. Les cibles sont représentées en rouge. 7

2.1 **Loi de Fitts et ses paramètres.** 29

3.1 **Représentation de la tâche d'atteinte de cibles 3D.** La figure montre les objets avec réflecteurs utilisés pour capturer la position des membres du participant, la position des sphères virtuelles (la sphère rouge étant la cible à atteindre) et les métriques spatiales utilisées pour les modes de sonification BI (Binary), AP (Angle-to-Pitch) et DVH (Dissociated Vertical and Horizontal). La sphère cible est située à l'une des 27 positions possibles. Les points O (coude) et P (extrémité de l'index) déterminent la direction indiquée par le participant. Avec BI, un bruit blanc est déclenché lorsque la ligne OP coupe la cible (θ nul). Avec AP, la hauteur du son varie en fonction de θ. Avec DVH la hauteur du son varie en fonction de θ_h (le projeté de l'angle θ dans le plan horizontal) et un bruit blanc est déclenché lorsque P est à la même hauteur que la cible. 40

3.2 Scène virtuelle représentant une séquence possible (un bloc) de 6 cibles parmi les 27 positions de sphères. La sphère verte représente la première cible, la sphère rouge la dernière cible, les sphères grises les cibles intermédiaires. Il y a une distance de 1,24 m entre deux colonnes de sphères en x, de 1,35 m en y, et trois hauteurs possibles en z : 70 cm, 110 cm et 150 cm. 41

3.3 **Représentation d'un pointage naturel (a) vs. des coordonnées polaires classiques (b)** Pointage naturel : l'angle θ a pour origine O alors que la distance d a pour origine P. Coordonnées polaires classiques : l'angle θ et la distance d ont la même origine O. 42

3.4 **Temps d'atteinte de cibles (échelle log10) pour chaque condition de sonification pour les comparaisons BI vs AP et AP vs DVH.** BI (Binary) donne la bonne direction par un bruit blanc binaire. AP rajoute à BI l'écart de direction global exprimé en fréquence. DVH dissocie les écarts sur les dimensions horizontale (fréquence) et verticale (bruit blanc binaire). Chaque boîte représente la médiane (ligne horizontale), l'intervalle interquartile (boîte) et les valeurs extrêmes aberrantes (moustaches). Les croix rouges représentent les temps médians des participants malvoyants. 46

3.5 **Pourcentage de participants ayant donné chaque score dans les deux conditions de sonification BI et AP.** BI (Binary) donne la bonne direction par un bruit blanc binaire. AP rajoute à BI l'écart de direction global exprimé en fréquence.Les trois critères sont pris en compte (efficacité, confort, facilité). Le gradient de couleur représente les niveaux de notation, le rouge correspondant à 0 et le vert foncé à 4, de sorte que les zones plus vertes correspondent à des jugements plus positifs. Les croix noires indiquent le score attribué par le participant malvoyant. 46

3.6 **Pourcentage de participants ayant donné chaque score dans les deux conditions de sonification AP et DVH.** AP rajoute au bruit blanc de bonne direction l'écart de direction global exprimé en fréquence. DVH dissocie les écarts sur les dimensions horizontale (fréquence) et verticale (bruit blanc binaire).Les trois critères sont pris en compte (efficacité, confort, facilité). Le gradient de couleur représente les niveaux de notation, le rouge correspondant à 1 et le vert foncé à 7, de sorte que les zones plus vertes correspondent à des jugements plus positifs. Les croix noires indiquent le score attribué par le participant malvoyant. 47

3.7 **Illustration des métriques spatiales utilisées pour les tâches d'atteinte de cible verticale et horizontale**. θ_h est la projection de l'angle formé par les demi-droites OP et OT sur un plan horizontal. ΔZ est la projection de la distance [PT] sur l'axe vertical Z. 55

3.8 **Temps d'atteinte de cible (échelle log10) pour les quatre métaphores testées dans la tâche verticale.** Chaque boîte représente la médiane (ligne horizontale), l'intervalle interquartile (boîte) et les valeurs extrêmes aberrantes (moustaches). 59

3.9 **Temps d'atteinte de cible (échelle log10) pour les trois métaphores testées dans la tâche horizontale.** Chaque boîte représente la médiane (ligne horizontale), l'intervalle interquartile (boîte) et les valeurs extrêmes aberrantes (moustaches). 60

3.10 Pourcentage de participants ayant donné chaque score pour les quatre métaphores de la tâche verticale. Les trois critères sont pris en compte (efficacité, confort, facilité). Le gradient de couleur représente les niveaux de notation, le rouge correspondant à 1 et le vert foncé à 7, de sorte que les zones plus vertes correspondent à des jugements plus positifs. 60

3.11 Pourcentage de participants ayant donné chaque score pour les trois métaphores de la tâche horizontale. Les trois critères sont pris en compte (efficacité, confort, facilité). Le gradient de couleur représente les niveaux de notation, le rouge correspondant à 1 et le vert foncé à 7, de sorte que les zones plus vertes correspondent à des jugements plus positifs. 61

3.12 **Écran de visualisation de l'environnement virtuel contenant la cible.** Le guide visualise la scène du point de vue du doigt du participant. La sphère orange représente la cible. 70

3.13 **Représentation des positions des cibles (sphères) et des mesures spatiales converties en sons.** Ici, la cible à atteindre (en rouge) est positionnée à la hauteur la plus basse et au milieu d'un petit côté extérieur de la grille. θ_h est la projection de l'angle formé par les droites OP et OT sur le plan horizontal parallèle au sol. d est la projection de la distance [PT] sur le plan horizontal parallèle au sol. ΔZ est la projection de la distance [PT] sur l'axe vertical Z. 71

3.14 T**emps d'atteinte de la cible (échelle log10) pour les deux modalités de sonification séquentiel et parallèle.** Chaque boîte représente la médiane (ligne horizontale), l'intervalle interquartile (boîte) et les valeurs extrêmes (moustaches). Les croix rouges représentent les temps médians du participant malvoyant. . . . 76

3.15 **Temps d'atteinte de cible (échelle log10) pour les deux conditions de guidage : sonification séquentielle et guidage vocal.** Chaque boîte représente la médiane (ligne horizontale), l'intervalle interquartile (boîte) et les valeurs extrêmes (moustaches). 77

3.16 **Pourcentage de participants ayant donné chaque score pour les deux conditions de sonification, séquentiel et parallèle.** Les trois critères sont pris en compte (efficacité, confort, facilité). Le gradient de couleur représente les niveaux de notation, le rouge correspondant à 1 et le vert foncé à 7, de sorte que les zones plus vertes correspondent à des jugements plus positifs. Les croix noires indiquent le score attribué par le participant malvoyant. 78

3.17 **Pourcentage de participants ayant donné chaque score pour les deux conditions : sonification séquentielle et guidage vocal.** Les trois critères sont pris en compte (efficacité, confort, facilité). Le gradient de couleur représente les niveaux de notation, le rouge correspondant à 1 et le vert foncé à 7, de sorte que les zones plus vertes correspondent à des jugements plus positifs. 78

3.18 **Pourcentage d'énoncés dans chaque catégorie de direction.** 1470 énoncés ont été catégorisés. La première catégorie correspond aux déclarations ne comportant aucune catégorie de direction (par exemple, "stop", "encore", etc.). La deuxième correspond aux énoncés ne contenant qu'une seule des six catégories de direction (par exemple « avance », « un peu plus à gauche », etc.). La troisième correspond aux énoncés contenant deux catégories de directions (par exemple « va en haut à droite », etc.). La dernière catégorie représente les erreurs de segmentation de Whisper. 79

3.19 **Distribution cumulative des énoncés répétés en fonction de leur occurrence, étiquetés par catégorie de mouvement.** Chaque point représente un énoncé, sa position sur l'axe des y correspondant au nombre de fois qu'il est répété dans l'ensemble de données (100 sont des énoncés dits une seule fois). L'axe des x représente l'ordre des énoncés, du plus fréquent au moins fréquent (100 est l'énoncé répété le plus souvent). La couleur de chaque point indique la catégorie de direction pré-assignée à l'énoncé. L'axe antéropostérieur (en rouge) contient les catégories de direction « avant » et « arrière ». L'axe médio-latéral (en bleu) contient les catégories de direction « gauche » et « droite ». L'axe vertical (en vert) contient les catégories de direction « haut » et « bas ». La catégorie « autre » (en gris) contient divers énoncés sans direction comme « stop », des répétitions comme « encore », des adverbes de degré comme « un peu plus », etc. Les étiquettes contiennent l'énoncé et le nombre de fois qu'il apparaît. 80

4.1 **Illustration du dispositif expérimental utilisé pour la tâche d'atteinte de cibles de type Fitts.** Le vecteur $\overrightarrow{OP}$ définit la direction de pointage du participant. La distance à la cible est calculée de P au centre de la cible T. Les cibles peuvent être placées à 12 cm (sphères rouges), 18 cm (sphères vertes) ou 27 cm (sphères bleues) du point de départ (« start trial button »). Le diamètre des cibles varie entre les essais : 5,33 cm ; 8 cm ; 12 cm. 93

4.2 **Métriques spatiales utilisées pour les sonifications AP et DVH testées sur une tâche de Fitts.** θ_h est la projection de l'angle formé par les droites OP et OT sur le plan horizontal parallèle au sol. d est la projection de la distance [PT] sur le plan horizontal parallèle au sol. ΔZ est la projection de la distance [PT] sur l'axe vertical Z. 95

4.3 **Droites de régression linéaire pour les temps de mouvement moyens (MT) en fonction de l'indice de difficulté (ID) pour chaque participant normovoyants et malvoyant.** 98

4.4 **Droites de régression linéaire pour les temps de mouvement moyens (MT) en fonction de l'indice de difficulté (ID) pour chaque participant, pour les deux conditions de sonification, AP et DVH.** 99

4.5 Pour les deux tâches, S représente le bouton sur lequel les participants doivent appuyer pour commencer un essai. A représente l'écart entre la cible et S, et W la largeur de la cible. ΔZ est l'écart en distance converti en son pour la tâche verticale, et θ_h est l'écart angulaire converti en son pour la tâche horizontale. . . 104

5.1 **Objets équipés de réflecteurs sphériques et localisés par le système de capture de mouvement.** 121

5.2 **Tâche de tir au pistolet**. Le participant vise la cible avec le pistolet factice. La cible physique utilisée pour la phase de familiarisation est située à un mètre de la table derrière laquelle le participant est positionné pour tirer. La cible virtuelle que le participant vise pendant le test (visible sur l'écran de contrôle à gauche) est située à 5 mètres du stand de tir et à 1,4 mètre au-dessus du sol. Sur l'ordinateur pilote (à gauche), l'expérimentateur visualise la visée en temps réel et les impacts des tirs virtuels. 123

5.3 **Distance du point de visée par rapport au centre de la cible au moment du tir (échelle log10)**. Distance globale (ΔD), distance horizontale (ΔDx), distance verticale (ΔDz), en mètres, pour chacune des quatre conditions de sonification. Chaque boîte représente la médiane (ligne horizontale), l'intervalle interquartile (boîte) et les valeurs extrêmes aberrantes (moustaches). Les croix rouges représentent la précision médiane du participant malvoyant, qui a testé les modes PO et PWN. 125

5.4 **Temps mis pour tirer, en secondes (échelle log10), pour chacune des quatre conditions de sonification.** Chaque boîte représente la médiane (ligne horizontale), l'intervalle interquartile (boîte) et les valeurs extrêmes aberrantes (moustaches). Les croix rouges représentent le temps médian de tir du participant malvoyant, qui a testé les modes PO et PWN. 126

5.5 **Pourcentage de participants ayant donné chaque score pour les quatre conditions de la tâche de tir.** Les trois critères sont pris en compte (efficacité, confort, facilité). Le gradient de couleur représente les niveaux de notation, le rouge correspondant à 1 et le vert foncé à 7, de sorte que les zones plus vertes correspondent à des jugements plus positifs. Les croix noires indiquent le score attribué par le participant malvoyant, pour chaque critère des deux conditions de sonification qu'il a testées, PO et PWN. 126

5.6 **Temps mis pour atteindre la cible 5 fois** Temps mis pour atteindre la cible 5 fois durant la phase de tir en temps illimité, séparé par comparaison de sonification, par groupe de participant et par session. Chaque boîte représente la médiane (ligne horizontale), l'intervalle interquartile (boîte) et les valeurs aberrantes (moustaches). Le temps moyen est affiché au dessus de chaque boîte. Les points représentent les temps de chaque participant. 138

5.7 **Scores composites vitesse-précision (partie 1)** Score prenant en compte la vitesse et la précision du tir, séparé par comparaison de sonification, par groupe de participant, par phase (temps illimité ; 50 secondes) par bloc et par session. Les scores sont calculés en additionnant la note z du temps et la note z de la distance à la cible. Les barres d'erreur représentent l'écart-type. Le score moyen est affiché au dessus de chaque barre. Les points représentent les scores de chaque participant. 140

5.7 **Scores composites vitesse-précision (partie 2)** Score prenant en compte la vitesse et la précision du tir, séparé par comparaison de sonification, par groupe de participant, par phase (temps illimité ; 50 secondes) par bloc et par session. Les barres d'erreur représentent l'écart-type. Le score moyen est affiché au dessus de chaque barre. Les points représentent les scores de chaque participant. 141

5.8 **Nombre de tirs réussis en 50 secondes** Nombre de tirs réussis en 50 secondes, séparé par comparaison de sonification, par groupe de participant, par bloc et par session. Les barres d'erreur représentent l'écart-type. Le nombre de tirs réussis moyen est affiché au dessus de chaque barre. Les points représentent le nombre de tir réussi de chaque participant. 142

5.9 **Temps mis pour tirer** Temps mis pour tirer, en secondes, séparé par comparaison de sonification, par groupe de participant et par session. Chaque boîte représente la médiane (ligne horizontale), l'intervalle interquartile (boîte) et les valeurs aberrantes (moustaches). Le temps moyen est affiché au dessus de chaque boîte. Les points représentent les temps de chaque participant. 144

5.10 **Précision des tirs.** Distance au centre de la cible, en mètre, séparé par comparaison de sonification, par groupe de participant et par session. Chaque boîte représente la médiane (ligne horizontale), l'intervalle interquartile (boîte) et les valeurs aberrantes (moustaches). La distance moyenne est affichée au dessus de chaque boîte. Les points représentent la précision de chaque participant. 145

5.11 **Pourcentage de participants normo-voyants ayant donné chaque score à chaque sonification, pour les quatre comparaisons sur la tâche de tir en temps limité.** Les trois critères sont pris en compte (efficacité, confort, facilité). Le gradient de couleur représente les niveaux de notation, le rouge correspondant à 1 et le vert foncé à 7, de sorte que les zones plus vertes correspondent à des jugements plus positifs. 146

5.12 **Pourcentage de participants malvoyants ayant donné chaque score à chaque sonification, pour les quatre comparaisons sur la tâche de tir en temps limité.** Les trois critères sont pris en compte (efficacité, confort, facilité). Le gradient de couleur représente les niveaux de notation, le rouge correspondant à 1 et le vert foncé à 7, de sorte que les zones plus vertes correspondent à des jugements plus positifs. 146

7.4 Annexe 4 : Liste des abréviations

AP : Angle-to-Pitch

AVSS : Audio-Vision Substitution System

BI : Binary

DSS : Dispositifs de Substitution Sensorielle

DVH : Dissociated Vertical-Horizontal

DVHpar : Dissociated Vertical-Horizontal, parallel mode

DVHseq : Dissociated Vertical-Horizontal, sequential mode

HRTF : Head Related Transfer Functions

ID : Indice de Difficulté

ILD : Interaural Level Difference

ITD : Interaural Time Difference

MT : Temps de Mouvement

OSC : Open Sound Control

OSG : OpenScene Graph

PO : Pitch Only

PSTpar : Pitch, Stereo and Timbre en mode séquentiel

PSTseq : Pitch, Stereo and Timbre en mode séquentiel

PWN : Pitch and White Noise

TP : Throughput

TVSS : Tactile-Vision Substitution System

VRPN : Virtual-Reality Peripheral Network

7.5 Annexe 5 : Publications

7.5.1 Conférences

Fons, C., Huet, S., Pellerin, D., Gerber, S., & Graff, C. (2023, July). Moving towards and reaching a 3-d target by embodied guidance : parsimonious vs explicit sound metaphors. In International Conference on Human-Computer Interaction (pp. 229-243). Cham : Springer Nature Switzerland.

Fons, C., Huet, S., Pellerin, D., & Graff, C. (2024, February). Using Fitts' Law to Compare Sonification Guidance Methods for Target Reaching Without Vision. In VISIGRAPP 2024-19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications-HUCAPP (Vol. 1, pp. 444-454). SCITEPRESS-Science and Technology Publications.

Fons, C., Huet, S., Pellerin, D., & Graff, C. (2025, June). Refining sonification methods to improve shooting performance and ergonomics for the visually impaired. In The 30th International Conference on Auditory Display (ICAD 2025).

7.5.2 Revue

Fons, C., Huet, S., Pellerin, D., & Graff, C. (2025). Dissociate Dimensions to Improve Sound Metaphors Guiding the Hand Towards 3D Non-visible Targets. SN Computer Science, 6(4), 387.

7.5.3 Soumis à PLOS One

Fons, C., Henry, A., Huet, S., Pellerin, D., Rossato, S., & Graff, C. (2025). Guiding sightless people one dimension at a time : how the strategy used in spontaneous voice instructions benefits sonification-based guidance.